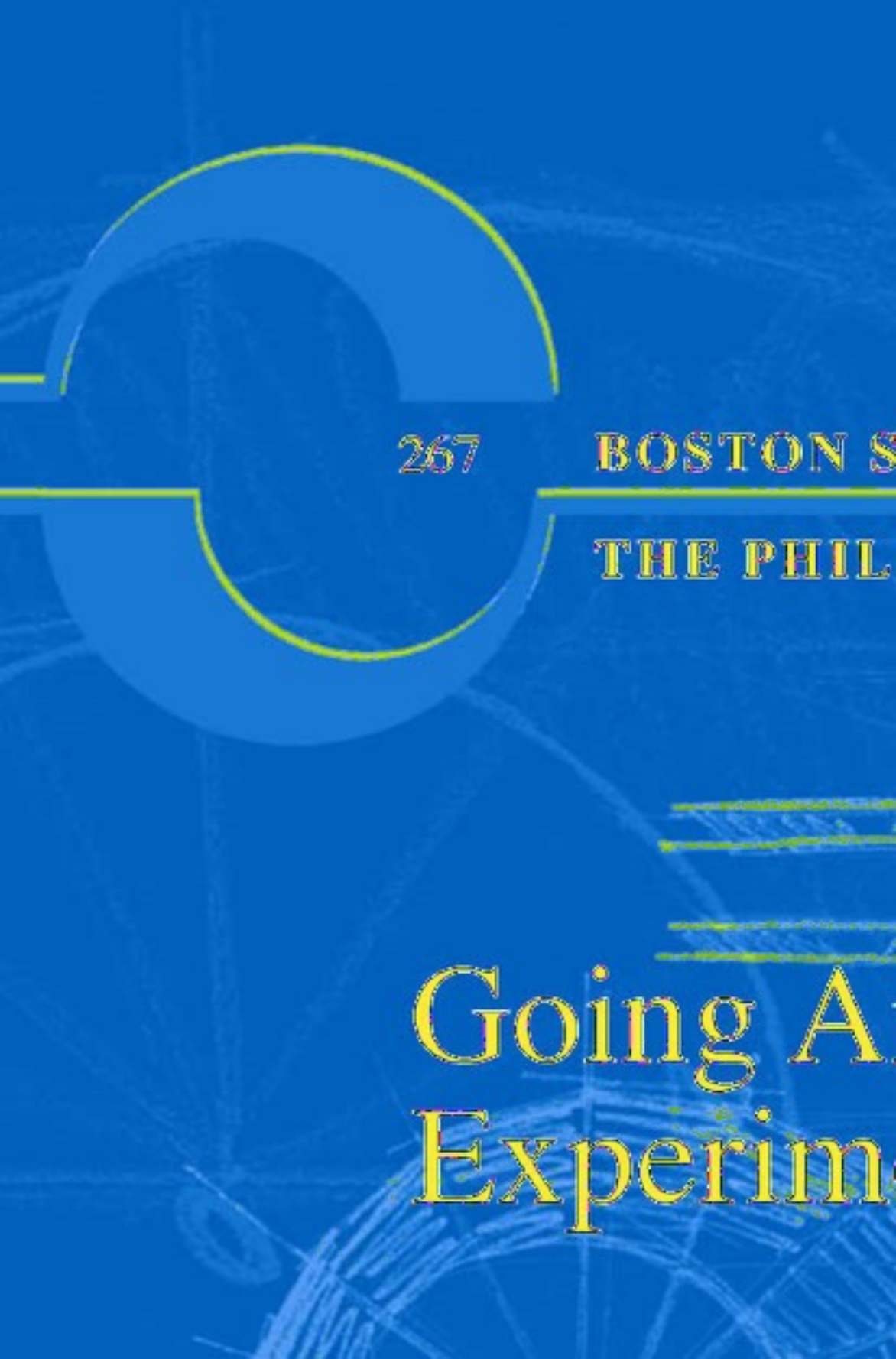

267
BOSTON S
THE PHIL
Going A
Experim

GOING AMISS IN EXPERIMENTAL RESEARCH

BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

VOLUME 267

GOING AMISS IN EXPERIMENTAL RESEARCH

by

GIORA HON
University of Haifa, Israel

JUTTA SCHICKORE
Indiana University, USA

FRIEDRICH STEINLE
Wuppertal University, Germany

Editors
Giora Hon
University of Haifa
Department of Philosophy
31905 Haifa
Israel
hon@research.haifa.ac.il

Jutta Schickore
Indiana University
Department of History
and Philosophy of Science
Bloomington, IN 47405
USA
jschicko@indiana.edu

Friedrich Steinle
University of Wuppertal
Fachbereich A, Wissenschafts- und
Technikgeschichte
Gaussstr. 20
42119 Wuppertal
Germany
steinle@uni-wuppertal.de

ISBN: 978-1-4020-8892-6 e-ISBN: 978-1-4020-8893-3

DOI 10.1007/978-1-4020-8893-3

Library of Congress Control Number: 2008937805

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Acknowledgments

Most of the essays comprising this volume grew out of presentations at the conference, *Going Wrong and Making it Right: Error as a Crucial Feature of Concept Adjustments in Experimental Contexts*, which we, the editors of this volume, organized in Aegina, Greece, in Spring 2003. The three of us had earlier met at the Max-Planck Institute for the History of Science, Berlin, where we had ample opportunity to discuss our common interest in concept formation and the pivotal role of error in experimental practice. We are most grateful to Hans-Jörg Rheinberger, one of the directors of the MPI, for constant encouragement and generous support of our research projects. Our conference proposal received the financial backing of the Dibner Institute whose then acting director, George Smith, responded enthusiastically to our suggestion to have an international meeting on the theme of going amiss. The conference had been originally planned to take place in Israel, but circumstances were such that it had to be moved to a different location. The meeting proved successful due in large measure to the invaluable support of Kostas Gavroglu, who graciously offered to act as the local organizer for the event. We are most grateful to him for hosting our conference as well as to Dina Dalouka for her assistance. We thank the conference participants for lively and fruitful discussions, and we are particularly grateful to our authors for their contributions and for their patience throughout the volume's long gestation period. The bulk of the editorial work on the volume was carried out while the three of us were members of the German-Israeli Foundation (GIF) project: *Generating Experimental Knowledge: Experimental Systems, Concept Formation and the Pivotal Role of Error*—a cooperation between the Philosophy Department at the University of Haifa (Giora Hon), the Max Planck Institute for the History of Science, Berlin (Hans-Jörg Rheinberger) and the Department of History at the University of Wuppertal (Friedrich Steinle). The project was funded by the Foundation Grant No. G-767–217.4/2002. We thank the foundation for its generous support.

Giora Hon *University of Haifa, Israel*
Jutta Schickore *Indiana University, USA*
Friedrich Steinle *University of Wuppertal, Germany*

Contents

Contributors

Erez Braun Network Biology Research Laboratories, and Department of Physics, (E.B.) and Medicine (S.M.), Technion, Haifa 32000, Israel; erez@physics.technion.ac.il

Allan Franklin Department of Physics, UCB 390, University of Colorado, Boulder, CO 80309-0390, USA; Allan.Franklin@colorado.edu

Jan Frercks Leuphana College, University of Lüneburg, Scharnhorststr. 1, 21335 Lüneburg; Germany; jan.frercks@t-online.de

Kostas Gavroglu Department of History and Philosophy of Science, University of Athens, Panepistimioupolis, 157 71 Ilisia; Greece; kgavro@phs.uoa.gr

Gerd Graßhoff History and Philosophy of Science, Institute of Philosophy, Länggassstr. 49a, University of Berne, 3012 Berne, Switzerland; gerd.grasshoff@philo.unibe.ch

Christoph Hoffmann Max Planck Institute for the History of Science, Boltzmannstr. 22, D-14195 Berlin, Germany; hoffmann@mpiwg-berlin.mpg.de

Giora Hon Department of Philosophy, University of Haifa, Haifa 31905, Israel; hon@research.haifa.ac.il

Ursula Klein Max Planck Institute for the History of Science, Boltzmannstr. 22, D-14195 Berlin, Germany; klein@mpiwg-berlin.mpg.de

Shimon Marom Network Biology Research Laboratories, and Departments of Physics (E.B.) and Medicine (S.M.), Technion, Haifa 32000, Israel; marom@technion.ac.il

Kärin Nickelsen History and Philosophy of Science, Institute of Philosophy, Sidlerstr. 5, University of Berne, 3012 Berne, Switzerland; nickelsen@philo.unibe.ch

Wendy Parker Department of Philosophy, Ellis Hall 202, Ohio University, Athens, OH 45701, USA; parkerw@ohio.edu

Hans-Jörg Rheinberger Max Planck Institute for the History of Science, Boltzmannstr. 22, D-14195 Berlin, Germany; rheinbg@mpiwg-berlin.mpg.de

Jutta Schickore Department of History and Philosophy of Science, Indiana University, Bloomington, IN 47405, USA; jschicko@indiana.edu

Friedrich Steinle Fachbereich A, Wissenschafts- und Technikge-Schichte, Gaussstr. 20, 42119 Wuppertal, University of Wuppertal, Germany; steinle@uni-wuppertal.de

Introduction: Mapping "Going Amiss"

Giora Hon, Jutta Schickore and Friedrich Steinle

It is a common human trait to wish to disown one's errors. While it is a truism that one can learn from one's failures, no one wants to be remembered for them, best to forget one's faults, left buried in layers of history. Philosophers are concerned with warranted knowledge—error is simply everything that is excluded from the domain of accepted claims to knowledge. It is the historians' task to uncover the past, but they too prefer to leave failures hidden away. Their worries, however, are more concrete. Historians fear that the study of past errors is intrinsically Whiggish and inadvertently produces anachronistic historical accounts. We take these worries seriously and transform them productively. We are convinced that it is fruitful to uncover forgotten and lost failures, subject them to analysis and learn from their moral. The central tenet of this volume is that failures count; they are quarries for knowledge. To be sure, failures should not be considered knowledge. Strictly speaking, they have proven to be false claims to knowledge, or, alternatively, the ground for a claim to be formulated could not be provided. We argue, however, that the study of failures, errors, pitfalls and mistakes shed light on the way knowledge is pursued and indeed generated, and we substantiate this position with historical accounts and philosophical analyses.

Science is a field of inquiry in which failures assume specific characteristics. If there is a method to scientific pursuits, their principles and features determine the scope and nature of the failures. We propose to examine the failures of scientific claims like an engineer who studies the breakdown of a certain technological system. However, unlike the engineer who knows well the expected performance of the technological system he or she has helped design, the historian and the philosopher of science are not privy to the original design; hence the inherent vagueness in the determination of characteristics of scientific failures. This is reflected in the title of this volume, *Going Amiss in Experimental Research*. "Going amiss" comprises two related themes: first the experimental results that proved wrong, and secondly the challenges that practitioners are facing in their everyday endeavors to generate experimental knowledge. The notion

G. Hon (✉)
Department of Philosophy, University of Haifa, Israel
e-mail: hon@research.haifa.ac.il

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_1,

"going amiss" reminds us of the fact that even in those cases where everything turns out right, numerous pitfalls and confusions had to be overcome. We are dealing not only with errors but also with misguided conceptions, dead ends, and reorientations.

Approaches to Error and Going Amiss

One specific aspect of our enterprise has already received some attention in recent years— the problem of error. Philosophers of science have begun exploiting error as a probe into scientific practice. Deborah Mayo and others have provided detailed analyses of the conceptual tools of error statistics (Mayo 1996). According to Giora Hon, an experimental setting comprises an ensemble of materials, instruments, measuring devices and of course experimenters. The setting might be devised according to explicit clear-cut questions and goals and thus employ specific background assumptions and auxiliary theories. The instances of error are shaped then by the procedures that the setting stipulates and by the underlying methodological assumptions. Bringing to light those elements that were most prone to error or failure and characterizing the sources of these problems could elucidate the structure of the experiment at stake (Hon 1989). Jutta Schickore has presented an analytical map of the field while arguing further that error can play epistemologically productive roles. She shows that "arguing from error" took center stage in the early nineteenth century, and that the scientists' encounters with the possibility and diversity of error gave rise to epistemological optimism (Schickore 2005). Here we see how the theme of error provides a fertile ground for dovetailing philosophy of science with history of science.

Like philosophers, historians too begin to see the riches of failures. Jed Z. Buchwald and Allan Franklin state their claim in the very title of their book: *Wrong for the Right Reasons*. They remark that "there is more to scientific error than merely noting that *x*... working in a particular locale, just believed that *y* was wrong. Of course, one might argue that it is hardly of any compelling interest to know that *x* was right and *y* wrong, at least for matters long gone. So what?" And they respond vigorously by claiming that "it's not a question of handing out report cards on the past"; rather, the aim is to understand "thoroughly" what was done. This, they contend, "can be achieved only by means of a mastery of contemporary technique ... that uncovers apparent *lacunae* and problems" (Buchwald and Franklin 2005, 1). As historians of science, Buchwald and Franklin seek understanding of past practices and claims to knowledge through the examination not only of successful but also of failed attempts at gaining knowledge. The determination and identification of error in past scientific studies thus become a historiographical means of conveying coherently not only the "state of the art", but also its shortcomings. Buchwald and Franklin imply that more general lessons can be drawn from this endeavor. They maintain that "meaningful statements can be made about mistakes and errors in science, and that these statements reach beyond the momentary and the local" (Buchwald and Franklin 2005, 3). By

examining how past scientists mastered the problem of error, they hope to recover enduring standards of scientific practice.

However, scientists must grapple with many kinds of pitfalls and not all of them are best described as errors and mistakes. Confusion, for example, indicates that not everything is going right. But this does not necessarily mean that an error has occurred. In this volume we highlight the diverse ways in which practicing scientists may go amiss. We revisit the historiographical worries arising from the study of past failures and suggest ways of alleviating them. We develop a nomenclature that is designed to capture different ways of going amiss. It provides a framework for the analysis of the specific epistemic roles of going amiss in scientific practices. The proposed nomenclature reflects our conviction that while there are countless ways of going amiss in experimental research, a viable taxonomy, if not a systematic analysis, of this intriguing phenomenon may be given. Finally, we demonstrate the productivity of going amiss.

Historiographical Challenges

One of the major worries that historians have about the study of past failures is that applying "failure" or related notions to past scientific endeavors may encourage anachronistic history. This is because such an approach often assesses whether past scientists went amiss according to current theories and today's standards, or explains with hindsight the reason for going amiss, knowledge to which the contemporary scientist had no access.

But we are not forced to adopt this perspective. We have a variety of options. We may for example seek to adopt the *perspective of the past practitioners*: How did *they* identify failures and other pitfalls in their work? What kinds of terms did *they* use? How did *they* cope with the problem? What standards of evaluation and assessment did *they* apply? We can trace this notion of *failure over time* and follow the *changing discourse on error*. What factors shaped the course of these debates? How did technological, socio-cultural, institutional, metaphysical, and other conditions influence the scientists' discourses and practices surrounding the phenomenon of going amiss? We may also utilize this analysis of past failures to shed light on the *metaphysical and epistemological commitments* of the practitioners (Hon 2004). In this perspective, the examination of things that went amiss can serve as a tool for clarifying modes of obtaining experimental knowledge. Alternatively, we may seek to establish whether the scientists went amiss according to the *standards of their own time*. This is indeed the perspective that Buchwald and Franklin offer (see especially Buchwald's contribution). Finally, to assess the range and limits of past tools and evaluate the practitioners' judgments about them, we may also investigate the *performance of extant historical instruments* with modern means. None of these enterprises is intrinsically Whiggish, but all of them are extremely fruitful for understanding the role of failure in the generation of experimental knowledge.

The Nomenclature of Going Amiss in Experimental Research

While the problem of failure and error is pertinent to every research activity, even in social science and humanities, we restrict our analysis to domains that rely on experiment and observation. We group terms in the semantic field of "going amiss" with a view to providing an analytic tool for distinguishing different kinds of failure. We seek to capture the richness and diversity of scientists' use of such terms, which varies widely over periods and fields of study. At the same time we also attend to the analyst's demand for tools of inquiry into failed experimental enterprises.

To organize the multitude of terms, we distinguish analytically between three main aspects of experimentation: (1) the agent's reasoning, perceptions and actions; (2) the tools of research, and (3) the object of investigation. In each of these respects, things can "go amiss".

First, what can go amiss with the agent's reasoning, perception, and action? The most straightforward case is the failure to apply an acknowledged convention, that is, an accepted rule or standard; this results in a **mistake**. Once a mistake is detected, rectification is immediate: applying properly the known rule or standard, and using rigorous checking procedure. Common examples of mistakes are typos and miscalculations. In the case of mistakes the standard is already established (Hon 1995). The agent may also be misled with respect to a matter of fact and we call this phenomenon **deception**. An example for a deceiving phenomenon is atmospheric refraction in astronomical observations. Optical illusions belong to this category because in such physical circumstances the apparent position or shape of the object is different from its true location and form.

A fundamentally more complex situation arises when knowledge of the standard is vague or missing altogether. The practitioners' language, concepts, perceptual skills, and basic theory may fail to provide an appropriate frame for analysis because they yield only **ill-formed questions** and **categories,** or **misguided expectations** that may result in **misinterpretation**. Similarly, the process of investigation may go amiss because of absent, overabundant, or vague, ideas of how to proceed. The agents find themselves in a situation where they do not know or are uncertain what to look for. A plethora of possibilities may present itself and nothing seems to give a lead.

There are cases where the mismatch is between experimental findings and theoretical expectation. **Anomalies** occur when experimental outcomes do not fit the relevant conceptual framework. **Discordance** is a disagreement between theoretically derived and observed findings in which the difference exceeds the error bounds on the sides of both theory and observation. And in these cases it is more often a matter of pragmatic concerns rather than of epistemology whether to look for the failure on the part of the theory or the experiment. In contrast, there may be clearly identifiably **inconsistencies** within the theory itself.

It is important to keep in mind that these categories are usually applied retrospectively, i.e. after the failure has been located or even removed. The experimenters'

actual situation when faced with a failure typically looks less clear. The practitioners speak of **surprise**, **puzzlement**, **confusion**, and the like. They try to determine the epistemic status of their claims and to track down the precise location of the failure, but they are not always successful. There may be situations in which it is impossible to decide whether the devil is in the agent, the tools, or the object. Sometimes experimenters give up and acknowledge the fact that they are stuck in a **dead end**—a situation that often leads to abandoning the original problem and taking up different questions, and at times to completely reorganizing the research field.

Turning to the tools we identify **malfunctions**. These failures result in divergence from the desired performance of the instruments as well as the setup. Epistemologically, the case is not too troubling, since the standard against which the instrument can be calibrated is (at least in principle) known. By contrast, much deeper problems arise with the central and omnipresent problem of **instrumental artifacts**, the unintended, and often unrecognized, (side-)effects that are generated by instruments and setups. The possible occurrence of systematic error is a veritable challenge: in experimental analysis one can never be absolutely sure that no previously unknown element interferes with the functioning of the instrument. Such artifacts cannot be avoided or eliminated altogether. Similarly, **noise** and **random errors** are intrinsic features of the experimental procedure. We call **instrumental discrepancies** the specific, identifiable divergence between two or more sets of generated output.

Thirdly, we consider the object under study. **Faulty experimental objects** may be due to unintended modifications. **Artifacts of preparation** may be produced through the scientists' inadvertent intervention. And again, the consequences of this intervention may remain unrecognized. Alternatively, the object may have certain unrecognized features that make it inappropriate for the intended study. A well-known example for such an object is the evening primrose in mutation experiments. Early theories of large-scale mutations were based on investigations of an organism whose chromosomes behave in a very peculiar way, and this unusual behavior of the plant misled the researcher.

Of course, in proposing such a set of terms, the historiographical perspectives outlined above are of critical importance. The difference between modern, analytical categories and those used by the historical actors may be highlighted with the use of this scheme of terms. Furthermore, it is important to realize that several of the terms of the conceptual framework have a historical dimension in the sense that they were introduced at one point and in a specific investigative context. These concepts may change their meaning over time. For example, the term "discordance" has a very precise meaning in celestial mechanics. Another case is the deviation from what is considered the normal, namely, the pathological of which one instance is **monstrosity**. Monsters were once conceived as **errors of nature**. From the 18th century onward, however, hardly any natural or experimental philosopher would have acknowledged that nature may err.

When failures are successfully identified and eventually removed, categories like those that we have listed in this section typically come into play. The set of terms that we have proposed serves both as an analytic tool and as a means of capturing the experimenters' own terminology. We hope that our nomenclature can help trace

the scientists' attempts at recognizing what exactly it is that is amiss. Moreover, as our Epilogue shows, this glossary prepares the grounds for discussions about the epistemic roles of going amiss.

The contributions to this volume indicate that failures, errors, confusions, and related epistemic phenomena such as dead ends are more than mere obstacles to scientific advancement and inconsequential curios for history and philosophy of science. In fact, this volume demonstrates that accounting for going amiss is productive both in the scientific domain and in reflecting on scientific practices and the knowledge thus gained. For the scientist the recognition of going amiss is an enticement for further work that may result in putting it aright or even in producing novel experimental knowledge. Similarly, for the historian and philosopher of science the historical account of going amiss and its philosophical analysis present a rich source for reflection on the nature of scientific practice and the knowledge that it generates.

References

Buchwald, J. Z. and A. Franklin (2005). "Introduction: beyond disunity and historicism." In: Buchwald, J. Z. and A. Franklin (eds.), *Wrong for the Right Reasons*. Dordrecht: Springer, pp. 1–16.

Hon, G. (1989). "Towards a typology of experimental errors: an epistemological view." *Studies in History and Philosophy of Science* 20: 469–504.

Hon, G. (1995). "Going wrong: to make a mistake, to fall into an error." *Review of Metaphysics* 49: 3–20.

Hon, G. (2004). "Putting error to (historical) work: error as a tell-tale in the studies of Kepler and Galileo." *Centaurus* 46: 58–81.

Mayo, D. (1996). *Error and the Growth of Experimental Knowledge*. Chicago: Chicago University Press.

Schickore, J. (2005). "'Through thousands of errors we reach the truth'—but how? On the epistemic roles of error in scientific practice." *Studies in History and Philosophy of Science* 36: 539–556.

Further Reading

Allchin, D. (1997). "A twentieth-century phlogiston: constructing error and differentiating domains." *Perspectives on Science* 5: 81–127.

Allchin, D. (2001). "Error types." *Perspectives on Science* 9: 38–58.

Bates, D. (1996). "The epistemology of error in late enlightenment France." *Eighteenth-Century Studies* 29: 307–327.

Cadeddu, A. (2000). "The heuristic function of 'error' in the scientific methodology of Louis Pasteur: the case of the silkworm diseases." *History and Philosophy of the Life Sciences* 22: 3–28.

Canguilhem, G. (1991). "A new concept in pathology: error." *The Normal and the Pathological*. New York: Zone Books, pp. 275–287.

Coen, D. R. (2002). "Scientists' errors, nature's fluctuations, and the law of radioactive decay, 1899–1926." *Historical Studies in the Physical and Biological Sciences* 32: 179–205.

Elliot, K. (2004). "Error as means to discovery." *Philosophy of Science* 71: 1–24.
Gurba, K. and E. Zarnecka-Bialy (eds.) (1998). *Philosophy and Error*. Krakow: Jagiellonian University Press.
Hon, G. (1998). "'If this be error': probing experiment with error." In: Heidelberger, M. and F. Steinle (eds.), *Experimental Essays—Versuche zum Experiment*, Baden-Baden: Nomos, pp. 227–248.
Holton, G. (guest ed.) (2005). Errors: consequences of big mistakes in the natural and social sciences, *Social Research* 72 (special issue).
Janis, A. and T. Horowitz (Eds.) (1994). *Scientific Failure*. Lanham: Rowman and Littlefields.
Kyburg, H. E. (1992). "Measuring errors of measurement." In: Savage, C. W. and P. Ehrlich (eds.), *Philosophical and Foundational Issues in Measurement Theory*. Hillsdale, NJ: Erlbaum, pp. 75–91.
Matthen, M. and E. Levy (1984). "Teleology, error, and the human immune system." *Journal of Philosophy* 81: 351–371.
Mellor, D. H. (1965). "Experimental error and deducibility." *Philosophy of Science* 32: 105–122.
Mittelstraß, J. (1999). "Vom Nutzen des Irrtums in der Wissenschaft." In: Carl, W. and L. Daston (eds.), *Wahrheit und Geschichte: Ein Kolloquium zu Ehren des 60. Geburtstages von Lorenz Krüger*. Göttingen: Vandenhoeck & Ruprecht, pp. 101–114.
Olesko, K. M. (1995). "The meaning of precision. The exact sensibility in early nineteenth-century Germany." In: Wise, M. N. (ed.), *The Values of Precision*. Princeton: Princeton University Press, pp. 103–134.
Petroski, H. (1994). *Design paradigms. Case Histories of Error and Judgment in Engineering*. Cambridge: Cambridge University Press.
Rasmussen, N. (1993). "Facts, artifacts, and mesosomes: practicing epistemology with the electron microscope." *Studies in History and Philosophy of Science* 24: 227–265.
Rescher, N. (2007). *Error: On Our Predicament When Things Go Wrong*. Pittsburgh: University of Pittsburgh Press.
Rudge, D. W. (2001). "Kettlewell from an error statistician's point of view." *Perspectives on Science* 9: 59–77.
Schickore, J. (2002). "(Ab)Using the past for present purposes: exposing contextual and transcontextual features of error." *Perspectives on Science* 10: 433–456.
Staley, K. W. (1996). "Novelty, severity, and history in the testing of hypotheses: the case of the top quark." *Philosophy of Science 63* (Supplement): S248–S255.
Zik, Y. (1999). "Galileo and the telescope: the status of theoretical and practical knowledge and techniques of measurement and experimentation in the development of the instrument." *Nuncius* 2: 31–67.
Zik, Y. (2001). "Science and instruments: the telescope as a scientific instrument at the beginning of the seventeenth century." *Perspectives on Science* 9: 259–284.

Part I
Error as an Object of Study

Error: The Long Neglect, the One-Sided View, and a Typology

Giora Hon

> To avoid errors… one must seek to disclose and to explain their source, illusion. Very few philosophers have done that, however. They have only sought to refute the errors themselves, without indicating the illusion from which they arise. This disclosure and breaking up of illusion is a far greater service to truth, however, than the direct refutation of errors, whereby one does not block their source and cannot guard against the same illusion misleading one into errors again in other cases because one is not acquainted with it.
>
> Kant (1800/1992, 562)

The long neglect

"The essays and lectures of which this book is composed are variations upon one very simple theme—the thesis that *we can learn from our mistakes.*" Thus prefaced in 1963 Sir Karl Popper (1902–1994) his book, *Conjectures and Refutations: The Growth of Scientific Knowledge* (Popper 1963/1974, vii, italics in the original). Popper characterizes his philosophical work as an attempt at developing a theory of knowledge and its growth. Both reason and experience function in this theory as means of exposing errors from which one may learn and thereby advance knowledge. A few years after the publication of this volume, when the second edition was ready for publication, Popper appeared to have had some additional thoughts concerning his "one sentence" summation of his thesis. He now assures the reader that "*all* our knowledge grows *only* through the correcting of our mistakes." He states that the general method of learning from mistakes is the method of trial and error and he further clarifies that in order to apply this method "we must already have some aim: we err if we stray from this aim" (Popper 1963/1974, ix, italics in the original). This appears to be an afterthought that is indicative of the situation:

G. Hon (✉)
Department of Philosophy, University of Haifa, Israel
e-mail: hon@research.haifa.ac.il

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_2,

Popper did not develop a sustained and full fledged theory of error. We may surmise that Popper exploits the issue of error as a rhetorical means for promoting and enhancing his philosophical position which is indeed well captured by the title of the book, *Conjectures and Refutation.*

It should not therefore surprise us that in Popper's earlier seminal work, *The Logic of Scientific Discovery*, no theory of error is developed and error is discussed incidentally and briefly, only with respect to the problem of measurement and issues concerning precision and accuracy (Popper 1959/1980). Notwithstanding, "error elimination" is a key feature in Popper's philosophy. It is the means by which his notion of critical rationalism functions in advancing knowledge.

> The proper answer to my question "How can we hope to detect and eliminate error?" is, I believe, "By *criticizing* the theories or guesses of others and—if we can train ourselves to do so—by *criticizing* our own theories and guesses (Popper 1974, 26, italics in the original).

Thus his view of the scientific method is summed up in four steps:

1. We select *some problem*—perhaps by stumbling over it.
2. We try to *solve* it by proposing a *theory* as a tentative solution.
3. Through the *critical discussion of our theories* our knowledge grows by the elimination of some of our errors, and in this way we learn to understand our problems, and our theories, and the need for new solutions.
4. The critical discussion of even our best theories always reveals new problems (Popper 1997, 159, italics in the original).

Popper singles out the third step as the most characteristic of the scientific practice, the error-elimination through criticism. For Popper the *objectivity of science,* and indeed its *rationality* are aspects of this stage, namely, the *critical discussion* of scientific theories (Popper 1997, 159). Here lies the difference, according to Popper, between an amoeba and a great scientist like Newton or Einstein: "the distinctive feature of science is conscious application of the *critical method;* in Stage 3 of our model, the stage of error elimination, we act in a consciously critical manner" (Popper 1999, 7, italics in the original).

It is in fact from the domain of biology that Popper takes the view of the crucial role which error correction may play in a theory of knowledge. He claims that "in biological evolution,… [error correction] appears to be the only means of progress" (Popper 1997, 100). The issues whether evolution is progressing or not and whether the concept of error is meaningful at all in this context should not detain us here. The point is however clear, the occurrence of error and the search for its elimination is central to Popper's philosophy. Yet Popper did not develop any theory of error, nor did he discuss this crucial epistemic phenomenon at length.

Popper is of course aware of philosophical theories of error. He refers to the "epistemological optimists"—Plato, Bacon and Descartes for whom truth is manifest—as being responsive to the fact that we sometimes "mistake error for truth", and he points out that, "in order to save the doctrine of manifest truth they were forced to explain the occurrence of error" (Popper 1997, 203). Having sketched

very briefly and in a most schematic way these theories of error, he concludes, "Bacon's theory of error is, in spite of its desirable consequences, untenable" (Popper 1997, 203). What would be a tenable philosophical theory of error, Popper does not say, let alone develop. This is rather disappointing. The reader of Popper's philosophical writings would be hard pressed to find discussions of theories of error, conditions of error, kinds of error, and in general the nature of error, for the simple reason that the theme of error is not developed in this philosophy. We may conclude that Popper presents the problem for purely rhetorical purposes. His claim then that,

> nobody is exempt from making mistakes. The great thing is to learn from them. And this is done by criticism, and by the discovery of new *problems* brought forth by criticism (Popper 1997, 144),

is more a slogan than a thorough, well argued philosophical position.[1]

Turning our attention to another contemporary influential philosopher, we may record that Willard van Orman Quine (1908–2000) opens his book, *From Stimulus to Science*, with the very problem of error.

> We and other animals notice what goes on around us. This helps us by suggesting what we might expect and even how to prevent it, and thus fosters survival. However, the expedient works only imperfectly. There are surprises, and they are unsettling. How can we tell when we are right? We are faced with the problem of error. These are worries about our knowledge of the external world. To deal with them we have had to run inward and seek knowledge *of* knowledge (Quine 1995, 1, italics in the original).

The reader soon realizes, however, that the problem of error is being used here as a literary device. Quine does not consider error a serious philosophical issue worthy of further consideration; he does not discuss the problem of error in the remaining parts of the book. One may therefore infer that by developing a theory of knowledge one implicitly takes care also of the problem error.

This implied view that the problem of error is some sort of a mirror image of the problem of knowledge is misleading if not mistaken. To be sure, error is an epistemological phenomenon that in the final analysis has to be analyzed with the tools of a theory of knowledge. However, nothing in such a theory reflects directly the phenomenon of error and it is clear that a special inquiry has to be undertaken.

I now turn from philosophy and philosophy of science to history of science where the problem of error is acute, pertaining to scientific knowledge and its historiography. It quickly transpires that here too the problem has been neglected and attention to its occurrence has been either superficial or uneven, stressing its importance in the historical context but performing no analysis. I proceed then to another example in which it appears that the author, though much appreciative of the problem, makes a rhetorical use of error and does not attempt a proper analysis.

Alexandre Koyré (1892–1964) is amongst the historians of science worth considering with respect to the problem of error. He expresses explicit interest in the problem and appears to be convinced of its importance. In Koyré's view, as Biagioli writes,

"mistakes played fundamental role in the production of scientific knowledge." Koyré's fascination with error, Biagioli remarks further, went so far as "to attribute an almost providential role to error." Thus, according to Biagioli, for Koyré "the fact that the mind can develop wrong scientific theories was not a symptom of its biases, idols, or conceptual shortcomings, but rather a sign of its freedom" (Biagioli 1987, 169). Whether the mind is indeed free and the occurrence of error is an expression of this freedom or, conversely, that error has an objective, ontological status and the mind stumbles upon it, are in fact issues for a theory of error to address and analyze. Whatever is the case, the road to truth—*itinerarium mentis in veritatem*—as Koyré writes, is not straight:

> The road to truth is full of traps, strewn with errors, and the failures there are more frequent than the successes.... But we would be wrong to neglect the study of errors (*l'étude des erreurs*)—it is by way of them that the mind progresses towards truth (Koyré 1966, 361; quoted by Jardine 2000b, 363).

This appears to be a promising line of inquiry. It is an expression of a view that seems to take a keen interest in the concept of error. Here I am in complete agreement with Koyré's claim that the study of errors can teach us about the working of the mind as much as the study of truth. Knowledge and error, as Ernst Mach (1838–1916) observes, flow after all from the same mental sources, and it is only success that can tell the one from the other. Thus, "a clearly recognized error, by way of corrective, can benefit knowledge just as a positive piece of knowledge can" (Mach 1905/1976, 84).

In the Introduction to his study of the law of falling bodies, Koyré expresses his fascination with the fact that both Galileo and Descartes made similar errors. The passage is worth quoting at length:

> For the historian of scientific thought, at least for the historian-philosopher, failure and error, especially the error of a Galileo or a Descartes, can sometimes be just as valuable as their successes. They can, perhaps, be even more so. They are, in fact, very instructive. They sometimes enable us to grasp and to understand the hidden processes of their thinking.
>
> No doubt the objection could be made that one should not look for a rational explanation of error. Error is a consequence of the weakness and limitations of the human mind, a function of its psychological, and even biological, conditioning. Everyone is capable of falling into error. Anyone can make mistakes. Nobody is an exception to this. It is enough to explain error by a lack of attention, by distraction, or by "inadvertence". We cannot accept this objection, or at least not entirely. No doubt any mistaken reasoning is inadvertent. And when Galileo and Descartes made mistakes they were guilty of this. But that this duplicated inadvertence (this duplication being in itself already an extremely curious fact) should lead them to exactly the same error, it is this... that cannot have been the result of pure chance.... Nevertheless it is far from plausible. There must be some reason for similarity in error (Koyré 1966/1978, 66).

Yet Koyré's exuberance of the reveries of error in history of science (here specifically the case of Galileo's and Descartes's duplicate error) does not hold good, for in this study of the law of falling bodies, and indeed in any other study of Koyré, no detailed analysis of the problem of error is forthcoming and no general theory of error is developed. Koyré considers error an unproblematic, primitive concept,

basic and well understood, namely, what is not the case vis-à-vis current theories and hindsight knowledge.

> Descartes himself… though he was a mathematician of genius, was never able either to recognise the mistake he had made, nor even, when he came across the correct formula in Galileo, to recognise that it was different from the one that he had put forward earlier himself. From which we can see once again just how difficult it was to isolate and grasp those simple and clear ideas with which we are so familiar from classical physics and from Cartesian philosophy. Even for Galileo. Even for Descartes (Koyré 1966/1978, 86).

I note that at issue is not a historiography that involves both judgmental evaluations and elements of anachronism (see Jardine 2000a). Rather, I stress the unproblematic meaning of the concept of error to which Koyré resorts. It is indeed based on the success criterion of which Mach has spoken. Koyré does not accompany his appeal to the concept of error with an appropriate analysis of the concept and therefore one is left with rather picturesque metaphors of errors as obstacles on the *itinerarium mentis in veritatem.* In sum, no insight into the epistemic phenomenon of error is offered in Koyré's historical studies.

Consider from another perspective Koyré's return to Plato; the study, *Discovering Plato* (1945), which he published after his inquiry into the law of falling bodies and in the aftermath of the Second World War. While detailing the arguments concerning the definition of knowledge that Socrates develops in the *Theaetetus*, at the juncture where Socrates argues that knowledge cannot be just true opinion because it will be then impossible to err, Koyré states that, "the problem of error is one of philosophy's very serious and crucial problems." Indeed, the problem of error together with its dour consequences was realized right at the inception of philosophy. Plato acknowledges in the *Theaetetus* that his theory of knowledge must account for the occurrence of error lest it would collapse altogether. "If [false opinion]… is found not to exist, we shall be forced to admit many absurdities," Socrates cautions Theaetetus (Plato, *Theaetetus*, 190E; Fowler 1977, 183). A theory of knowledge which cannot explain error and its occurrence, cannot discriminate it from truth, and so cannot explain that either. Indeed, the mark of a false proposition is that it is indistinguishable in form from a true one.

Attempting an explanation, Plato assumes an object of error as well as an object of knowledge and studies the consequences of this assumption. "It is, then,… possible for the mind to regard one thing as another and not as what it is" (Plato, *Theaetetus*, 189D–E; Fowler 1977, 179). Error consists then in taking the one object for the other. However, to know this we must know both, and knowing the object of error as such is not an error. Plato reaches thus an *impasse.* Error involves the contradiction that we must simultaneously both know and not know in the same cognitive reference, "to know what one does not know; not to know what one does know", to use Koyré's formulation (Koyré 1945, 50)—a consequence that may threaten the coherence of any theory of knowledge that takes up seriously the challenge of error (Schiller 1908).

Koyré's assessment of the status of the problem of error is thus most appropriate. Notwithstanding, he states this important observation in a footnote (Koyré 1945, 40 n. 9). The observation epitomizes the state of the problem of error; the problem is

indeed "very serious and crucial", yet the considerations it has received have generally been scanty and peripheral, that is, metaphorically they amount to a footnote not only to philosophy but to science as well.

It appears however that things have changed. Historians and philosophers of science are increasingly paying attention to the vast and varied problem of error, both as a probabilistic epistemic phenomenon and as an inherent difficulty in the context of observation and experimentation. This growing concern with the concept of error is connected to the shift in attention which the discipline of history and philosophy of science is currently undergoing. The actual nitty-gritty practices of scientific research, the bras tacks of research in both the natural and the social sciences, including the social context within which knowledge is generated, have become legitimate objects of historical and philosophical studies as much as the conceptual content itself and its cognitive aspect (e.g., Galison 1987; Shapin 1994). In parallel to this development the literature of philosophy and history and philosophy of science has seen a growing interest in the problem of error, its historical background in philosophy (Evans 1998) and its presence in philosophy and discussions in the cultural-literary realm (Affentranger 2000; Almeder 1999; Bates 1996, 2002; Crocker 1953; Kenaan 1999); as a probabilistic phenomenon (Krüger et àl. 1989, Section III: Uncertainty; Mirowski 1994, 1995; Sheynin 1983; Stigler 1986), its occurrence in mathematics (Sherry 1997) and in science (Schlich 1993), its focus in engineering (Pool 1997) and generally as an acknowledged theory (Swijtink 2000, for an annotated bibliography of "Error Theory"). Gone are the days when one could flip casually the remark that "once [errors of measurement and other forms of experimental error]… have been discounted, our attention can turn to the logico-mathematical structure" (Sellars 1961, 73; cf. Mellor 1965, 106). The occurrence of errors, especially in observation and experimentation, constitutes a permanent feature which deserves proper attention. The problem of error is not incidental to the pursuits of science, it deserves the attention of philosophers and historians of science. As Mellor pointed out, error should not be treated as "a tiresome but trivial excrescence on the neat deductive structure of science" (Mellor 1967, 6).

The One-Sided View

In this vein, Deborah Mayo (1996) has made an attempt in her book, *Error and the Growth of Experimental Knowledge*, to bring the problem of error to the center of discussion. The ambitious title takes its cue from the subtitle of Popper's book, *Conjectures and Refutations: The Growth of Scientific Knowledge*, which as we have seen evolves around the seemingly simple theme that we can learn from our mistakes. Mayo's reference to experimental, rather than to scientific, knowledge has to do with her claim that experimental knowledge is knowledge grounded on argument from error (1996, 7). Clearly, we expect here to find the missing Popperian insights into how to learn from mistakes.

Mayo is vigorously critical of the orthodoxies of Popper and Kuhn and vehemently rejects the popular Bayesian approach in philosophy of science. Building on a classical statistical tradition (1996, x, 10, 337; see however Mirowski 1995, 547 n. 10), she develops a sophisticated technical machinery with which she underpins a non-Bayesian philosophy of science. Mayo sides with Peirce, Neyman and Pearson and stands against their common opponents—the Bayesians. She christens her position "error-statistical philosophy of experiment" (1996, 410, 442, 457, 464), because the chief feature that her approach retains from the Neyman-Pearson statistical methods is the centrality of error probabilities (1996, x–xi). The demand that it is necessary to take into account the error probabilities of experimental procedure in order to determine what inferences are licensed by data, is the principal element that fundamentally distinguishes, according to Mayo, her approach from others (1996, 442).

The statistical methods which Mayo offers are designed not only to stabilize experimental knowledge and explain its growth but also to grapple with other issues pertinent to philosophy of science, such as the Duhem problem (1996, 103, 106–109). Indeed, in referring to an error-statistical philosophy of science, she has in mind the various ways in which statistical methods based on error probabilities may be used in philosophy of science generally. Mayo seeks to convince the reader that the error-statistical philosophy of science that she has developed has a structure and a logic, so that its parts hang together to provide a full-bodied philosophy. In her view, this philosophy presents a viable alternative to the Bayesian Way (1996, 442–444). She thus concludes that,

> the ability to make successful inductions, our success in obtaining experimental knowledge, is explained by the error-statistical properties of our methods. We make progress in experimental knowledge—experimental knowledge grows—because we have methods that are manifestly adequate for learning from errors (1996, 464).

On this account, errors and the statistical methods for treating them have become the tools for building the body of knowledge we call science. Has error then gained in this rejuvenated classical statistical approach the attention it deserves?

Error covers multiple sins. It is a multifarious epistemological phenomenon of great breadth and depth. To be sure, error-statistical analysis is a powerful tool which is much needed in the technical domain of the reduction of data. As such, it can undoubtedly throw light on methodological issues, but it cannot do philosophical justice to such complex concepts as error and experimental knowledge. Mayo's book is not about error but about error probabilities, and the notion of experimental knowledge it develops is rather the knowledge of the probabilities of specified outcomes in some series of experiments (1996, 12).

The central problem which Mayo addresses is how to link experimental data to primary theoretical hypotheses. It is commonly known that data gathered from experiment are corrupted by various kinds of error introduced by intermediary processes of observation and measurement. Moreover, data are finite and discrete, while primary hypotheses may refer to an infinite number of cases and involve continuous quantities such as weight and temperature (1996, 132). Nevertheless, since the mandatory linkage between data and hypothesis is the only game in town

which deserves the appellation scientific, one must use experimental data to assess the values of theoretical quantities. How is this done? How are we to proceed?

Mayo perceptively points out that the study of the relation between evidence and hypotheses solely in terms of logical relationships ignores completely all the deliberate and active intervention in which the experimenter is engaged (1996, 212). In focusing too exclusively on the appraisal of global theories, philosophers have overlooked how positive grounds are provided for local hypotheses, namely, whenever evidence counts as having severely tested them. By attempting to talk about data and hypotheses in some general way, apart from the specific context in which the data and hypothesis are generated, modeled, and analyzed to answer specific questions, philosophers have missed the power of such a piecemeal strategy, and underdetermination arguments have flourished (1996, 213). In sum, Mayo instructs not to follow the Bayesian Way, but rather the path of the classical statistician and to search, as Pearson put it,

> *for a way of expressing in mathematical terms what appeared… to be the requirements of the scientist in applying statistical tests to his data* (quoted by Mayo 1996, 381, italics added by Mayo).

The application of statistical tests is the key idea, and as Pearson reported,

> from the start we [Neyman and Pearons] shared Professor Fisher's view that in scientific enquiry, a statistical test is "a means of learning" (quoted by Mayo 1996, 382).

On this account one learns in science not from how much the evidence confirms the hypothesis tested, but rather from *how* discordant evidence shows a given model to be in a specified respect. Learning from experiments requires not some update of the probability assignment that one starts out with, but deliberate and often devious methods of testing with which one builds, corrects, and fills out a model (1996, 212, 433).

The theme of learning from error thus plays a central role in the experimental program which Mayo develops. She demonstrates how the famous Popperian thesis of "conjecture and refutation" does not stand up to criticism. According to Mayo, Popper's account falls far short of showing how reliable knowledge is obtained from experiment or how that knowledge grows. Mayo finds in the Peircean error-correcting justification of induction, the very justification she needs for her error-statistical methods of science.

> The justification for these methods lies in their ability to control error probabilities, hence sustain learning from error, hence provide for the growth of experimental knowledge (1996, 413).

Mayo's central thesis is that the argument from error, that is, learning from error, may be described in terms of a test of a hypothesis, H, that a given error is absent. The evidence indicates the correctness of hypothesis H, when H passes a severe test—one with a high probability of failing H, if H is false. An analogous argument is used to infer the presence of an error (1996, 64; cf., Giere 1997, S183–S184;

Howson 1997; Mayo 1997; Chalmers 2000, 198ff.). This then is the framework of Mayo's error-statistical philosophy of experiment.

Mayo's philosophy of experiment relies neither on scientific theories nor on a theory of experiment; rather, it relies on methods—statistical methods—for producing experimental effects (1996, 15). This observation is crucial. It explains the limited view of experiment exhibited in this study. In spite of the fact that Mayo speaks voluminously about the need to address the actual practice of experimentation, she focuses her attention solely on statistical calculations. As Mirowski, criticizing both Bayesians and classical statisticians, aptly puts it: "the empirical inquirer cranks through the formulas, assigns the error probabilities and reports an outcome—all as a hermetically self-contained procedure" (Mirowski 1995, 542). This is not what one would expect of, say, a Faraday, a Helmholtz, a Hertz, a Rutherford, or a Kapitza. Consider Peirce's observations on experimental style:

> Of all men of the century Faraday had the greatest power of drawing ideas straight out of his experiments and making his physical apparatus do his thinking, so that experimentation and inference were not two proceedings, but one. To understand what this means, read his *Researches on Electricity.* His genius was thus higher than that of Helmholtz, who fitted a phenomenon with an appropriate conception out of his store, as one might fit a bottle with a stopper (Peirce 1966, 272).

Mayo's "full-bodied experimental philosophy" (Mayo 1996, 444) is not attuned to the act of experimenting; it focuses rather on the end result: data and their statistical tests. For example, questions as to the interpretation of the experimental result do not arise in this framework. For another example, no theory of experiment is forthcoming in this approach (see, e.g., Radder 1995, 2003). Mayo indeed admits that her philosophy of experiment is limited:

> the statistical theory of experiment deals *only with certain kinds of experiments* insofar as their behavior may be characterized by certain parameters. A characteristic of key interest is the relative frequency with which an outcome occurs, or would occur, in a sequence of applications of the experiment in question (Mayo 1996, 161–162, italics has been added; cf., 164, 173).

She does therefore state that, "any planned inquiry in which there is a deliberate and reliable argument from error may be said to be experimental" (1996, 7). In Mayo's philosophical framework, experimental knowledge becomes completely statistical:

> experimental knowledge is knowledge of the probabilities of specified outcomes in some actual or hypothetical series of experiments. Its formal statement may be given by an *experimental distribution* (a list of outcomes and their associated probabilities), or by a standard "random" process such as a coin-tossing mechanism (1996, 12, italics in the original; cf., 162, 461).

Thus the errors upon which Mayo builds her error-statistical philosophy of experiment are not error at large but rather a specific and indeed limited kind of error, namely, error probabilities. Error probabilities are not probabilities of hypotheses,

but the probabilities that certain experimental results would occur, were one or another hypothesis be true about the experimental system (1996, 367).

As I have remarked, error covers multiple sins. What kind of error did Franck and Hertz commit in their Nobel winning experiment? (Hon 1989a) What happened in Ehrenhaft's experiments which made him conclude that there are subelectrons? (Hon 1989b, 485; Franklin 1981) What went wrong in Kaufmann's experiments so that he could speak decisively against the correctness of Einstein's special theory of relativity? (Hon 1995) How did Blondlot discover the existence of a new form of radiation—the N rays—which does not exist? (Nye 1980 and Hon 1989b, 493–494) Why was Lowell so convinced that the visible lines on the surface of Mars are in fact artificial canals for irrigation purposes? (Hon 1989b, 492–493) "The history of science," Maxwell observed, "is not restricted to the enumeration of successful investigation. It has to tell of unsuccessful inquiries, and to explain why some of the ablest men have failed to find the key of knowledge" (Maxwell 1871, 251). The probabilistic approach to the study of error is undoubtedly of considerable importance; of no less importance is the study of conceptual and physical circumstances in which errors in experimentation and generally in the search for knowledge may originate. "One must," as Wittgenstein demands, "reveal the source of error" (Wittgenstein 1979, 61). And to follow Kant's dictum, one must seek to disclose and explain sources of errors. Indeed, for Kant the disclosure of the source of error is of greater service to truth than the direct refutation of errors (Kant 1800/1992, 562). It is apparent that Maxwell, Wittgenstein and Kant speak of different kinds of error than the error probabilities of Mayo; theirs is the general phenomenon of error, not the one-sided account of error which Mayo expounds (cf. Hon 1998b).

Three distinct yet related themes of research could be identified in this domain of inquiry: (1) a study of the history of error in science and especially in observation and experimentation (e.g., Hon 1989c); (2) an epistemology of experiment that can inform a history of error *via* (3) a classification of types of error that reflects this epistemology (Hon 1989b, 2003). I proceed to outline this approach which I call "probing experiment with error" (Hon 1998a). The resulting typology of experimental errors is designed to contribute towards an epistemology of experimentation.

A Typology

Experimentation is a method designed specifically for obtaining physical knowledge. This method can be viewed like any other method which claims to secure knowledge. But what kind of method is it? What do we mean by experiment? I propose to use the concept of error as a means of probing experiment. The method lays bare the structure of experiment by studying its possible sources of errors.

According to Bohr, "by the 'experiment' we can only mean a procedure regarding which we are able to communicate to others what we have done and what we have learnt" (quoted in Honner 1987, 159). However, experiment is not just some

procedure by which one can communicate to others what one has done and what one has learnt from the result. Experiment—and this is the central point of the thesis—is a procedure, a physical process, which can be cast into an argument of a formal nature. A crucial characteristic of an experiment is that its result constitutes a claim to knowledge and it is this claim which distinguishes a mere procedure from an experiment. Being a claim to knowledge, the conclusion of an experiment may be seen as the result of a chain of reasoning concerning the behavior of some material system. The conclusion is connected then with a certain mode of reasoning. An experiment can be made to exhibit an inference from premises to a conclusion—the argument of the experiment.

How is it that a physical process we call experiment ends up in a claim to knowledge that may affect a theory? Take for example the following experimental outcome: "The electrostatic and electromagnetic properties of the cathode rays," concluded Hertz in 1883 his set of cathode ray experiments, "are either *nil* or very feeble" (Hertz 1883/1896, 254). The fact that this claim to knowledge, the result of a set of experiments carried out by a gifted experimenter, is considered today erroneous shows that there is an additional element to the physical process that makes up the experiment. The error we discern in Hertz's experiment cannot be associated with the physical process itself, with the course which nature takes within the constraints of the experiment. Rather, errors indicate claims to knowledge. An error reflects the existence of an argument into which the physical process of the experiment is cast. It is this hidden argument upon which the claim to knowledge is based.

The conclusion of experiment, the resultant claim to physical knowledge, is clearly prone to error. What is it for an experiment to result in an error? What is it for an experiment to contain an error? An experiment may be erroneous for all sorts of reasons associated with its different components. But whatever the reason in the final analysis an experiment is erroneous when it does not warrant its conclusion. This, however, is nothing else but saying that an erroneous experiment reflects a failed argument. In contrast to the Aristotelian position, we commonly hold with Newton that "Nature… is… always consonant to itself" (*Philosophiae Naturalis Principia Mathematica*, Rules of Reasoning in Philosophy, Rule III). Clearly, an error is ascribed to an argument, not to the actual physical process.

An argument is a sequence of propositions of a special form: each proposition is either an assumption, a premise, or one that arises through inference schema as a conclusion from propositions earlier in the sequence. An experiment, I claim, can be cast into a formal argument whose propositions instantiate partly states of affairs of material systems and partly inference schema and some lawful, causal connections. In other words, an experiment implies an argument the premises of which are assumed to correspond to the states of the physical systems involved; e.g., the initial conditions of some material systems and their evolution in time. These premises warrant the argument's conclusion. This is a fundamental point in characterizing the implied argument of a physical experiment: the experimenter aims at securing premises which correspond *accurately* to the actual physical situation of the experiment.[2] The requirement of accuracy puts this kind of argument apart from any other argument in which the premises may be conditional, hypothetical or counterfactual.[3]

An experiment therefore is a procedure, a physical process, that has a logical facet of a rhetorical force. Indeed, an experiment is often used as an instrument of persuasion for reaching agreement on certain situations.

An analysis of fallacies would not however suffice for the understanding of erroneous experimental results. What is required is a full fledged exposition of the problem of experimental error in its broadest sense, that is, beyond the mathematical, technical, narrow meaning. In other words, to illuminate the possible failures of an experiment one needs an epistemology of errors in experiment. Such a study will not only shed light on the nature of possible failures of experiment, but it will also reveal the structure of this method of acquiring knowledge—the method of experimentation.

Experimenters perform essentially two different tasks: they *prepare* a system and then they *test* it. A *preparation* is a procedure which is in principle completely specified. This is crucial for the correct completion of the experiment since on the basis of this knowledge rests the belief that the premises of the implied argument of the experiment correspond to its physical conditions. The second task, the *test*, starts like a preparation: it has a specified procedure which triggers the interaction between the prepared set-up and the object under study. The test includes another step, a crucial one, whereby *information*, which was previously unknown, is supplied to an observer, that is, the experimenter. This information is not trivial, not only because identical tests following identical preparations need not have identical outcomes (both in classical and quantum physics, though for categorically different reasons), but primarily because this information constitutes, after a suitable reduction, the sought *new* physical knowledge. Within certain constraints, experimenters are free to *choose* preparations and tests that they wish to perform—this is their prerogative. However, they are not free to choose the future outcome of a test. They are bound by the acquired information.[4]

The characterization of experiment as *preparation* and *test* is admittedly very general. Nevertheless, it provides an insight into the structure of the procedure of experimentation which is otherwise impossible to generalize. Beyond this general characterization, the experimental procedure becomes too varied to submit to any general deductive characteristic. Indeed, it appears that there cannot be an exhaustive survey of all experimental techniques. One can certainly characterize different general schemes of experiment such as the scattering technique which has been dominant in high energy physics throughout the twentieth century, or the technique of subjecting radiation to electric and magnetic field; each technique however has its own idiosyncratic features which may not be suitable to generalization.

Thus, however general, the characterization of the procedure of experiment in terms of *preparation* and *test* may be useful for the analysis of experiment. This characterization allows for a clear perception of the connection between the actual procedure of the experiment and its implied argument. The preparation stage is principally about presuppositions, whereas the test stage has to do with a mixture of premises and conclusions—the outcome of the experiment. It may be further observed that each of the two stages has two sub-categories. The preparation stage has theoretical and practical sub-categories while the test stage may be conceived of as the recording of the information and its processing.

We have seen that the experimenter aims at securing premises which correspond accurately to the states of the experiment. For that purpose one assumes in the preparation stage a theory which is considered correct; it underpins the experiment. This theory, the background theory and its "daughter" theories: the theories of the instruments and of the set-up itself, are therefore taken for granted. They are not tested by the experiment. Then there is the process of realizing these theoretical requirements in practice. Once the theory and its physical realization have been put to work in the preparation stage, information is allowed to flow to the recording instrument (natural or artificial). Finally, a process of reducing this information and interpreting the result takes place. This is the conclusion of the experiment—the outcome.

Thus four stages may be distinguished: (1) laying down the theoretical framework; (2) constructing the apparatus and making it work; (3) taking observations and readings; and (4) processing the recorded data and interpreting them. While stages (1) and (2) constitute the premise of the argument in the theoretical and concrete sense, the conclusion (4) is inferred from (3): the empirical information obtained through the test.

At each stage of the experiment different types of error may occur in the corresponding part of the implied argument. Clearly, these different types go beyond the standard dichotomy of systematic and random error. This dichotomy does not focus on the source of the error; rather, it examines the nature of the error by applying a mathematical criterion irrespective of the experiment. The criterion judges whether the estimation of the error is arrived at by a statistical analysis of repeated measurements or by non-statistical methods in which much depends on the judgment of the experimenter in allocating limits to the accuracy of the measurement. The former kind of error is random whereas the latter is systematic. Since the criterion is mathematical no distinction is observed *vis-à-vis* the source. Therefore, errors of different origins are grouped together. Thus, the error which one knows to have originated in a certain conceptual framework and under some physical conditions is transformed into a technical, mathematical term.

The present discussion of error in experiment seeks to go beyond the technical, narrow mathematical sense of experimental error, and to negotiate a new path towards the problem of error in experiment. The advantage of this approach—the proposed typology of experimental errors—is that it reflects faithfully the structure of the argument of experiment and may thereby shed light not only on the notion of experimental error but also on the nature of experimentation.

The proposed taxonomy consists then of four types of experimental error associated with the following categories:

1. Background theory;
2. Assumptions concerning the apparatus and its working;
3. Observational reports;
4. Theoretical conclusions.

Each of these four possible types of experimental error may be illustrated with case studies from the history of science (see, e.g., Hon 1989a, 1989b, 1995). A distinct characteristic of the proposed taxonomy is its focus on the source, rather than on the

resultant error. By concentrating on the definitions of different classes of source of error, the typology illuminates from a negative perspective the elements which are involved in experiment and their interrelations.

To conclude in the spirit of the metaphorical language of Francis Bacon (1561–1626) and following his idols of the *theatre*, I suggest to characterize these four kinds of experimental errors as idols of the *script*, the *stage*, the *spectator* and the *moral*. The image of theatrical play constitutes a convenient and useful metaphorical setting for experiment since, like a play enacted on stage, an experiment is the result of an activity that has truly "a show" at its centre. In an experiment, nature is made, if you will, to display a show on a stage conceived and designed in some script. The show is observed and registered by a human or automated spectator and, finally, interpretation is proposed with a view to providing a moral—that is, the outcome of the experiment as knowledge of the physical world.

Notes

1. For another criticism of Popper's position on learning from mistakes, see Chalmers 1973.
2. Accuracy refers to the closeness of the measurements to the "true" value of the sought physical quantity, whereas precision indicates the closeness with which the measurements agree with one another independently of their relations to the "true" value. Accuracy thus implies precision but the converse is not necessarily true (see Hon 1989b, 474).
3. Notice, however, that an experimental result may be correct while the premises are false. This would still make the experiment as a whole erroneous.
4. This is an ideal analysis. In practice, problems of reduction and questions of interpretation make the issue much more complex. It should be further noted that this study excludes the discussion of fraud in science. Fraud is irrelevant to the epistemology of experiment.

References

Affentranger, B. (2000) *The Spectacle of the Growth of Knowledge and Swift's Satires on Science* (New York: Allworth Press).

Almeder, R. (1999) "Recent Work on Error," *Philosophia* 27: 3–58.

Bates, D. W. (1996) "The Epistemology of Error in Late Enlightenment France," *Eighteenth-Century Studies* 29: 307–327.

Bates, D. W. (2002) *Enlightenment Aberrations: Error and Revolution in France* (Ithaca: Cornell University Press).

Biagioli, M. (1987) "Meyerson and Koyré: Toward a Dialectic of Scientific Change," *History and Technology* 4: 169–182.

Buchwald, J. Z. (ed.) (1995) *Scientific Practice: Theories and Stories of Doing Physics* (Chicago: Chicago University Press).

Chalmers, A. F. (1973) "On Learning from our Mistakes," *The British Journal for the Philosophy of Science* 24: 164–173.

Chalmers, A. F. (1978/2000) *What Is This Thing Called Science*, 3rd ed. (Buckingham: Open University Press).

Crocker, L. G. (1953) "The Problem of Truth and Falsehood in the Age of Enlightenment," *Journal of the History of Ideas* 14: 575–603.

Evans, G. R. (1998) *Getting it Wrong: the Medieval Epistemology of Error* (Leiden, Boston, Köln: Brill).
Fowler, H. N. (tr.) (1977) *Plato's Theaetetus*, the Loeb Classical Library, *Plato VII*, vol. 128 (Cambridge, Massachusetts: Harvard University Press).
Franklin, A. (1981) "Millikan's published and unpublished data on oil drops, "*Historical Studies in the Physical Sciences* 11: 185–201.
Galison, P. (1987) *How Experiments End* (Chicago: University of Chicago Press).
Giere, R. N. (1997) "Scientific Inference: Two Points of View," *Philosophy of Science (Proceedings)* 64: S180–S184.
Hertz, H. (1883/1896) *Miscellaneous Papers,* with an introduction by P. Lenard. D. E. Jones and G. A. Schott (trs.) (London: Macmillan).
Hon, G. (1989a) "Franck and Hertz versus Townsend: A Study of Two Types of Experimental Error," *Historical Studies in the Physical and Biological Science* 20: 79–106.
Hon, G. (1989b) "Towards a Typology of Experimental Errors: an Epistemological View," *Studies in History and Philosophy of Science* 20: 469–504.
Hon, G. (1989c) "Is There a Concept of Experimental Error in Greek Astronomy?" *The British Journal for the History of Science* 22: 129–150.
Hon, G. (1995) "Is the Identification of Experimental Error Contextually Dependent? The Case of Kaufmann's Experiment and Its Varied Reception," in Buchwald 1995, 170–223.
Hon, G. (1998a) "'If This Be Error': Probing Experiment With Error," in M. Heidelberger and F. Steinle, (eds.), *Experimental Essays—Versuche zum Experiment* (Baden-Baden: Nomos), 227–248.
Hon, G. (1998b) "Exploiting Errors," *Studies in History and Philosophy of Science* 29: 465–479.
Hon, G. (2003) "The Idols of Experiment: Transcending the 'Etc. List'," in Radder 2003, 174–197.
Honner, J. (1987) *The Description of Nature: Niels Bohr and the Philosophy of Quantum Physics* (Oxford: Clarendon Press).
Howson, C. (1997) "Error Probabilities in Error," *Philosophy of Science (Proceedings)* 64: S185–S194.
Jardine, N. (2000a) "Uses and Abuses of Anachronism in the History of the Sciences," *History of Science* 38: 251–270.
Jardine, N. (2000b) "Koyré's Kepler/Kepler's Koyré," *History of Science* 38: 363–376.
Kant, I. (1800/1992) *Lectures on Logic*, J. M. Young (tran. and ed.) (Cambridge: Cambridge University Press).
Kenaan, H. (1999) "Subject to Error: Rethinking Husserl's Phenomenology of Misperception," *International Journal of Philosophical Studies* 7: 55–67.
Koyré, A. (1945) *Discovering Plato* (New York: Columbia University Press).
Koyré, A. (1966) *Études D'Histoire de la Pensée Scientifique* (Paris: Presses Universitaires de France).
Koyré, A. (1966/1978) *Galileo Studies* (New Jersey: Humanities Press).
Krüger, L., L. J. Daston, and M. Heidelberger (eds.) (1989) *The Probabilistic Revolution: Ideas in History*, 2nd printing (Cambridge, Massachusetts and London, England: The MIT Press).
Mach, E. (1905/1976) *Knowledge and Error* (Dordrecht: Reidel).
Maxwell, J. C. (1871) "Introductory Lecture on Experimental Physics," in J. C. Maxwell (1890) *Scientific Papers*, W. D. Niven (ed.) (Cambridge: Cambridge University Press), ii, 241–255.
Mayo, D. G. (1996) *Error and the Growth of Experimental Knowledge* (Chicago and London: Chicago University Press).
Mayo, D. G. (1997) "Error Statistics and Learning From Error: Making a Virtue of Necessity," *Philosophy of Science (Proceedings)* 64: S195–S212.
Mellor, D. H. (1965) "Experimental Error and Deducibility," *Philosophy of Science* 32: 105–122.
Mellor, D. H. (1967) "Imprecision and Explanation," *Philosophy of Science* 34: 1–9.
Mirowski, P. (1994) "A Visible Hand in the Marketplace of Ideas: Precision Measurement as Arbitrage," *Science in Context* 7: 563–589.
Mirowski, P. (1995) "Civilization and Its Discounts," *Dialogue* 34: 541–559.

Nye, M. J. (1980) "N-rays: an episode in history and psychology of science," *Historical Studies in the Physical and Biological Sciences* 11: 125–156.
Peirce, C. S. (1966) *Selected Writings*, edited with an introduction and notes by P. P. Wiener (New York: Dover).
Plato, *Theaetetus*; see Fowler (tr.) 1977.
Pool, R. (1997) *Beyond Engineering: How Society Shapes Technology* (New York, Oxford: Oxford University Press).
Popper, K. (1959/1980) *The Logic of Scientific Discovery* (London and New York: Routledge).
Popper, K. (1963/1974) *Conjectures and Refutations: The Growth of Scientific Knowledge.* 5th ed. (London: Routledge and Kegan Paul).
Popper, K. (1997) *The Myth of the Framework* (London and New York: Routledge).
Popper, K. (1999) *All Life is Problem Solving* (London and New York: Routledge).
Quine, W. V. (1995) *From Stimulus to Science* (Cambridge: Harvard University Press).
Radder, H. (1995) "Experimenting in the Natural Sciences: A Philosophical Approach," in Buchwald 1995, 56–86.
Radder, H. (ed.) (2003) *The Philosophy of Scientific Experimentation* (Pittsburgh: University of Pittsburgh Press).
Schiller, F. C. S. (1908) *Plato or Protagoras? Being a Critical Examination of the Protagoras Speech in the Theaetetus with Some Remarks upon Error* (Oxford: Blackwell).
Schlich, T. (1993), "Making Mistakes in Science: Eduard Pflüger, His Scientific and Professional Concept of Physiology, and His Unsuccessful Theory of Diabetes (1903–1910)," *Studies in History and Philosophy of Science* 24: 411–441.
Sellars, W. (1961) "The Language of Theories," in H. Feigl and G. Maxwell (eds.) *Current Issues in the Philosophy of Science* (New York: Holt, Rinehart & Winston).
Shapin, S. (1994) *A Social History of Truth* (Chicago, IL: University of Chicago Press).
Sherry, D. (1997) "On Mathematical Error," *Studies in History and Philosophy of Science* 28: 393–416.
Sheynin, O. B. (1983) "Corrections and Short Notes on My Papers," *Archive for History of Exact Sciences* 28: 171–195.
Stigler, S. M. (1986) *The History of Statistics: The Measurement of Uncertainty Before 1900* (Cambridge, Massachusetts and London, England: The Belknap Press of Harvard University Press).
Swijtink, Z. G. (2000) "Error Theory," in A. Hessenbruch (ed.), *Reader's Guide to the History of Science* (London and Chicago: Fitzroy Dearborn Publishers), pp. 225–227.
Wittgenstein, L. (1979) "Remarks on Frazer's *Golden Bough*," in C. G. Luckhardt (ed.), *Wittgenstein Sources and Perspective* (Hassocks, Sussex: Harvester Press), pp. 61–81.

Error as Historiographical Challenge: The Infamous Globule Hypothesis

Jutta Schickore

Introduction

There is a repertoire of infamous errors that keep cropping up in standard histories of microscopy: the wheels attached to the head of rotifer, the little animals in semen, and the holes in the middle of blood corpuscles. The so-called 'globule theory' or 'globule hypothesis'[1] of organic matter occupies a prominent place on the list. This theory concerned the basic structure of bodily tissues. According to it, muscular and nerve tissue consist of tiny globules of regular size and shape. As the standard stories go, this erroneous view of organic elements was widely advocated among microscopists in the early decades of the nineteenth century. Only in the late 1830s, the cell theory offered a more correct description of muscular and nerve tissue.

The aim of my essay is twofold. First, I seek to revise the 'standard account' of the globule hypothesis. Secondly, I utilize the episode of globulism to consider the conditions for the use of the concept 'error' as an analytic tool for history and philosophy of science. What are the features of the historical record that merit the application of that concept? The episode lends itself well to this double purpose because we also have an alternative to the received view, namely, John Pickstone's historically contextualized reconstruction of the globule theory. This reconstruction takes into account the practitioners' theoretical assumptions that motivated their research. Remarkably – or perhaps not surprisingly? – Pickstone concludes that the globule hypothesis was not an error.

I begin by reviewing the standard accounts of the episode and demonstrate how these accounts explain the source of this so-called error and the way in which it was overcome. I confront this received view with Pickstone's narrative and show how he arrives at the conclusion that the globule hypothesis was not an error. My analysis uncovers the assumptions and implications that are tied to both accounts. I then survey various early nineteenth-century microscopical works on organic

J. Schickore (✉)
Indiana University, Bloomington, IN 47405, USA

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_3,

matter. My findings contrast with both the received view and Pickstone's alternative. I demonstrate that the proposed descriptions of globules differed widely and significantly in both content and scope. So actually it is very difficult to identify a plausible candidate for 'the' globule hypothesis. I also show that the factors that contributed to the formation of novel histological accounts were not so much technical or theoretical but first and foremost, practical. They had to do with the proper handling of the microscope and the prepared objects. In the final part of the paper, I consider whether or not the episode merits the application of the concept of error. I argue that it is crucial to distinguish between three perspectives on past science, appraisive, actors', and constitutive-analytic perspectives. The implications of the term 'error' change significantly in each perspective, and depending on the questions that we ask. In line with the overall theme of the present volume, I also show that the applicability of the term 'error' is limited. In some contexts and for some purposes, other concepts, in particular, 'going amiss', are the better conceptual tools for the analysis of the pitfalls in knowledge generation.

The Erroneous Globule Hypothesis in the Received View

The received view of the globule hypothesis was cemented in mid-twentieth-century historiography of microscopy. At that time, many historians regarded the globule hypothesis as part of the prehistory of cell theory, which took shape with Matthias Schleiden and Theodor Schwann's work in the late 1830s.[2] Some historians presented the microscopists' accounts of globules as the direct predecessor of cell theory (e.g., Baker 1948), others as digression from the straight path leading from eighteenth-century fiber theory to nineteenth-century cell theory (e.g., Berg 1942, Rather 1969). In both versions of the received view, however, the description of globules was presented as erroneous. It was generally assumed that this erroneous description had to be removed to make way for the correct – or at least, for a more correct – account of cells as basic structural units of plant and animal tissue.

To explain why the error had occurred and how it could be overcome, historians invoked technical developments and theoretical innovations. Some decades ago, it was not unusual for historians of microscopy to subscribe to strong technological determinism, according to which microscopy was shaped and driven by technological improvements. Such technology-centered histories of microscopy suggest that novel instrumental developments, especially the arrival of the achromatic microscope around 1830, enabled the practitioners to see more and better.[3] In this perspective, the formation of the globule theory was explained by the dark and distorted image that the pre-achromatic instruments produced. When optically superior microscopes came into use, the practitioners were able to identify that theory as erroneous.

John Baker's comprehensive article, *The Cell Theory, a Restatement, History and Critique* of 1948 outlines the career of the globule hypothesis along these lines. For Baker, globulism was the 'forerunner' of cell theory. Baker located the origins of globulism in the seventeenth century. He identified a number of so-called 'globulists', the majority of them working around 1800, when globulism reached its 'zenith':

Among them were Johann Friedrich Meckel, Everard Home and Francis Bauer, Carl Friedrich Heusinger, and Henri Milne Edwards. Baker's account of this episode is exemplary for the technological determinism that has informed large parts of the history of microscopy. Technological determinism makes a twofold claim: bad instruments are responsible for erroneous results and improved instruments make the error apparent and thus help overcome it. Baker conceded that some observations made around 1800 might have been of real globular elements. He suggested however that most descriptions were of optical artefacts, which were produced by bad instruments. Baker's account of Georg Prochaska's work is exemplary for this view. According to Baker, Prochaska described and depicted 'carefully the various appearances obtained as he focused his lens on the globules, and there can be no doubt that the figure shows merely the effect of spherical aberration. This error, together with haloes produced by lenses of small numerical aperture, must often have led the globulists astray' (Baker 1948, 117). In the same vein, Baker then characterized several globulists' works as 'not important', 'confused', and 'simply erroneous' (Baker 1948, 119). He even dated the refutation of the globule hypothesis in a particular year, 1827. It was in this year that Thomas Hodgkin and Joseph Lister provided 'some check on their [the globulists'] errors'. Baker writes:

> Using the improved microscopes designed by Lister, they [Hodgkin and Lister] found no globules, but only fibres, in striated muscle and in the muscle of arteries. They looked in vain for globules in nerves. They saw no globules in brain, but only very small particles, which they regarded as resulting from the disintegration of the tissue. [. . .] They were aware that their results differed from those of Milne Edwards, who was a friend of Hodgkin, and attributed the difference to the imperfection of Edwards's microscope.[. . .] The fact that the excesses of the globulists were exposed by Lister's microscope seems significant; for the particular advantage of his instrument was that spherical aberration was corrected and the 'ring' appearance round small particles thus reduced. His objectives, though not perfected by this time, must already have been good. The work of Hodgkin and Lister was a healthy and much-needed corrective (Baker 1948, 120–121).

In his 1953 monograph on Rudolph Virchow, Erwin Ackerknecht offered a more specific prehistory: the prehistory of his protagonist's field, cellular pathology. Drawing on Baker's work, Ackerknecht distinguished two globule hypotheses, both of which he presented as provisional versions of cell theory: Globule hypothesis or cell theory No. 1 was a version of the 'fiber theory' of organic matter that dominated eighteenth-century anatomy. According to this theory, the basic unit of bodily matter is the fiber. Cell theory No. 1 was 'the opinion, preferred toward the end of the fiber era by numerous authors between Prochaska (1797) and J. Berres (1837), that fibers originate from small globules. This opinion was obviously a compromise of traditional theories, actual observations, and optical illusions created by the poor microscopes of the time' (Ackerknecht 1953, 71). Globule hypothesis or cell theory No. 2, 'the' globule theory, took the globule to be a basic structural element; and again, Ackerknecht noted that some of that work 'has been dismissed as optical illusions due to poor instruments' (Ackerknecht 1953, 72). Both versions of the globule hypothesis were abandoned and the cell theory proper was developed when 'much improved' microscopes were introduced (Ackerknecht 1953, 72).[4]

Technological determinism was not the only framework that served to explain the career of the globule hypothesis. Other historians argued that this hypothesis was inspired by certain theoretical expectations that colored the microscopists' observations. These historians brought the apparent popularity of the globule hypothesis together with romantic theories of life. Because romantic biologists assumed that living organisms consist of homogeneous and simple organic elements, so the argument goes, they were inclined to lend credibility to their observations of globular particles. Once romanticism loosened its grip on biology, the practitioners were free to realize the incorrectness of their view and cast it aside. It is this version of the globule hypothesis and its fate that best reflects the grand narratives about the life sciences in the nineteenth century, according to which their advancement was a hard-won liberation from the spell of those futile *naturphilosophische* speculations that dominated the decades around 1800. In this version, the globule theory was a digression from the straight path that led from eighteenth-century fiber theory to early nineteenth-century cell theory. We find an illustrative example of this position in Alexander Berg's detailed history of the fiber theory of organic matter. Berg drew attention to the fact that for many microscopists in the eighteenth century, the fiber was the building block of organic matter. According to Berg, microscopists inspired by nature philosophy found that these fibers consisted of even smaller uniform elementary units of the organism, namely, globules, which were lined up like beads on a string. Berg explained that

> this theory of strings of beads was characteristic for the need for a unified basic account of the complicated bodily parts [...]. The fact that one did not proceed more rapidly cannot be explained by faulty technology, as *Studnicka* has tried. Rather, the reason can be found in the fact that the observations are influenced and in their interpretation even determined by the dominant theories (Berg 1942, 451).

This impediment of the evolution of the cell theory was finally overcome by the introduction of technically advanced microscopes (see Berg 1942, 453). Berg also revealed the source for this story about the origin of the globule hypothesis, and his reference is illuminating. The reader is pointed to Virchow's *Cellular Pathology.* The first chapter of this work offers a brief history of fiber and globule theories of organic matter. According to Virchow, early nineteenth-century microscopists believed that fibers were not the smallest elements of organic matter. Rather, they consisted of linearly lined up globules. It is suggested that the globule theory was not only due to incorrect applications of the microscope but also to lofty speculations, which were informed by nature philosophy. Virchow noted:

> The bad method, which was prevalent throughout the last and for a part of this century, that one observed with mediocre instruments in full sunlight brought about a certain dispersion of light in all microscopical objects, and the observer got the impression that he saw nothing else but globules. On the other hand, this view agreed with the nature philosophical [*naturphilosophischen*] ideas about the first origin of all forms (Virchow 1871, 22).

It is this suggestion that Berg took up, disregarding the fact that Virchow's presentation of the history of his own scientific field was already highly polemical. The

accusation that romanticism obstructed the advancement of science served him and many of his contemporaries to contrast his own 'empirical and unbiased' research projects favorably with the 'speculative and tainted' projects of his predecessors.[5]

While the standard mid-twentieth-century accounts of the globule hypothesis differ in that some of them explain globules as optical artifacts and others as theory-infused misperception and also in that some describe technological advancements and others theoretical developments as the crucial factors in the career of the globule hypothesis, the perspective of the narrative is very similar. In all versions of the received view, the globulists subscribed to an erroneous account of organic tissue, and the removal of this error eventually made way for considerable scientific progress, namely the formation of the cell theory. Labels like 'error', 'confusion', etc. are attributed in hindsight and in an appraisive fashion to claims to knowledge that were replaced by a more advanced account, one that is closer to the view that we accept today. Here, 'error', 'confusion', etc. are used as the complement of common success categories like 'rational' or 'progressive'.

In appraisive history, 'error' and related concepts are crucial for the reconstruction of past science because correcting errors makes a difference to the advancement of science. Errors are obstacles; they have a negative impact on scientific developments. To correct them, the sources of these errors – bad instruments, incorrect theories – need to be removed. Past errors construed in the perspective of an appraisive history of the present serve an important purpose for historians: They signify the epistemic distance between past and present science and mark the progress of the field.

A 'Saner and More Important Development'

In light of recent developments in historiography, many readers may feel uncomfortable with the received view of the 'erroneous' globule hypothesis. We have been made wary of appraising historical episodes by standards that were not the standards and conceptual frameworks of the actors' time. Instead, we seek to understand as well as possible the historical actors and their assessments of scientific knowledge claims on their own terms. Instead of measuring the work of past practitioners against later advancements of their discipline, we seek to establish what *they* considered to be an error, and how exactly *they* attributed success and failure to scientific beliefs and practices. But this discomfort with the use of appraisive categories in history raises an intricate historiographical question. What exactly are the consequences for the 'erroneous globule hypothesis' of our attempts to do justice to past microscopists' conceptual frameworks and evaluative standards? Does it still make sense to portray the microscopists' descriptions of globular elements as 'erroneous'?

John Pickstone's account of the globule hypothesis in his 1973 article 'Globules and Coagula' suggests a negative answer. Pickstone wants us to take seriously those early nineteenth-century 'patterns of thought which underlay this whole sequence of investigations'. If we do so, he says, the globule theory will 'emerge as a saner and

more important development than has hitherto been recognized' (Pickstone 1973, 337). In his article, Pickstone reconstructs the globule theory in the context of the chemical theories of organic matter that were prevalent at that time among physiologists (especially in France and Britain). He argues that we can understand the globule theory as a product of certain expectations that were derived from chemical models of tissue formation. Arguing against technical determinism, Pickstone points out – correctly, I think – that even though the early nineteenth-century microscopes (before 1830) were somewhat more powerful than older devices, it is implausible to assume that the possibility of observing globules had arisen only around 1830 (Pickstone 1973, 337). In other words, to make sense of the globule theory, we need to look elsewhere. We need to consider the theoretical assumptions that the practitioners brought to bear on their observations. Pickstone's article unearths the chemical theories of tissue formation that might have informed the observations made by the early nineteenth-century microscopists. Notably, based on his careful reconstruction of the historical context of the globule theory, Pickstone comes to the conclusion that globulism cannot be regarded as part of the prehistory of the cell theory: his portrayal of the chemical context of the globule hypothesis presents this tradition as 'a search for common units at a level that we see as subcellular' (Pickstone 1973, 356).

Pickstone accounts for the globule theory through a reconstruction of the practitioners' beliefs and assumptions, which, according to Pickstone, provided an incentive to look for globules. In this respect, his approach resembles Virchow's and Berg's. In another important respect, Pickstone's reconstruction also agrees with the received view: He identifies a particular account of organic tissue – the globule hypothesis – which allegedly dominated microscopy for a certain time, and which was then given up in favor of an alternative account. He analyzes this transition in terms of the factors that were relevant to the dismissal of that hypothesis. And here we find again a partial similarity to the received view: Although Pickstone does not attribute the globule hypothesis to 'bad' instruments, he does agree that ultimately it is the advancement of microscopy that 'undermined' the globule hypothesis (Pickstone 1973, 356). The crucial difference between the approaches – apart from the obvious difference that Pickstone offers a careful and detailed portrayal of the microscopists' concrete conceptual framework, while Berg's and Virchow's accounts of this episode are painted with broad brushstrokes – lies in the fact that Pickstone acknowledges the scientific merits of the contemporaneous framework of organic chemistry, while Virchow and Berg denigrate the microscopists' beliefs as *naturphilosophische* speculation and rather unfortunate aberration from the straight path of science. So we might say that Virchow and Berg found fault twice: once in the erroneous background assumptions, which, or so it is implied, impeded scientific advancement, and once in the erroneous descriptions of organic tissue that were inspired by these background assumptions. In contrast, Pickstone rehabilitates the background assumptions as well as the globule hypothesis as 'saner and more important development than has hitherto been recognized.' For the assessment of the globule hypothesis, the difference is critical. In Pickstone's reconstruction, the globule hypothesis was not an aberration or impediment but a reasonable and

indeed crucial step in the development of theories of organic matter. Note that even though Pickstone provides much more historical context, his account is not purely descriptive. It still has an appraisive dimension. The aim is to reconstruct the globule theory as rational and fruitful by showing how it fits into the past scientists' theoretical outlook and expectations.

At this point we are faced with the following alternative: First, the received view of the episode, according to which the globule hypothesis was an error that impeded the advancement of histological theories (and the cell theory in particular); secondly, Pickstone's reconstruction, according to which the globule theory was not an error but an important step in the development of histological theories. In both cases, 'error' serves as an evaluative category of appraisive history, complementary to success categories such as 'rational' and 'progressive'.

In what follows, I take a fresh look at the episode of globulism. The line of argumentation that I pursue is different from both the received view and Pickstone's alternative. I agree with Pickstone that one should not attempt to reconstruct the episode of globulism from the perspective of the 'more advanced' or 'more correct' cell theory and its history. But my description of the main features of the historical episode differs from his. My survey of the reports that were given of observations of organic tissue indicates that the microscopists encountered a plethora of conflicting results. Not all of them could be correct. A variety of accounts competed with each other. Pragmatic considerations regarding the application of microscopes were at least as important for the revision of descriptions of organic matter as theoretical or technological advancements. In the last part of my paper, I consider the episode in light of the concept of error. I argue, first, that we can highlight significant features of the episode when we recover the actors' usage of the term 'error'. Secondly, I show that other significant features of the episode can be captured neither in actors' terms nor with the appraisive category 'error' that the received view introduced. To acknowledge these features, we need to adopt a third analytic perspective. The concept of 'going amiss' is conducive to this purpose.

A Plethora of Observations

To simplify things, I concentrate on research that was carried out in the German lands and take only an occasional glance at investigations in France and Britain.[6] Needless to say, if we scrutinize handbooks of anatomy and articles on the structure of organic tissue from the 1810s, 1820s, and 1830s, we find ample evidence for globulism. But at the same time, it becomes obvious that 'the' globule hypothesis is quite hard to pin down. Several physicians, anatomists, and physiologists from all over the German lands, many of them well known and influential, contributed to the debates about the elementary parts of organic tissue. It is true that some microscopists' accounts appear as straightforward illustrations of the globule hypothesis, for example, the synoptic *Handbook of Human Anatomy* by Johann Friedrich Meckel, professor of anatomy at Halle, Germany. Meckel reported results

of his predecessors Prochaska, della Torre and Barba and stated: 'The structure of the nervous system is essentially everywhere the same. The last elements of form are globules and a half-liquid substance through which they are connected.' He added however that 'the observers' specifications of the shape and size of these globules and the degree of consistency of the connecting means diverge' (Meckel 1815, 265). A few years later, Karl Friedrich Burdach in Königsberg, already well known for his anatomical studies of the brain, described the nerve fibers explicitly as 'strings of beads, consisting of similar, regular globules, which are lined up' (Burdach 1819, 165). In the *System of Histology* of 1822, the anatomist Carl Friedrich Heusinger, then *außerordentlicher* professor at Jena, described fibers more generally as 'globules lined up through polar forces' and surmised that fibers originate from globules – adding that this hypothesis about the origin of the fibers could be *proven* only for the nerve fibers (Heusinger 1822, 115).

The distinguished member of the Royal College of Surgeons and Fellow of the Royal Society Everard Home and his co-worker Francis Bauer[7] also suggested that nerves consisted of 'many bundles of extremely delicate fibres, formed of minute globules connected together by a gelatinous substance' (Home 1821, 25). In 1823, the French M.D. Henri Milne Edwards, whose name is probably most frequently and most closely connected with the 'globule hypothesis', found that all kinds of animal tissues were composed of elementary globules, which looked the same everywhere, a result that was made accessible to the German-speaking community a few years later (Milne Edwards 1827). The physician Gottfried Reinhold Treviranus, renowned for the publication of the multi-volume work *Biologie oder Philosophie der lebenden Natur für Naturforscher und Aerzte* in the early 1800s, published a collection of anatomical and physiological essays in 1816 together with his brother Ludolf Christian. In the contribution that deals with the organic elements of the body, we read that the nerves are tubes filled with a viscous matter that contains delicate tubes and small globules, which are 'much smaller than blood globules', as well as irregular matter which 'seems to have developed from a unification of the globules' (Treviranus 1816, 129). In 1821, Carl Asmund Rudolphi, the first professor of anatomy and physiology at the newly-founded university in Berlin, reported 'small, irregular corpuscles' in the nervous substance, which were 'usually referred to as globules, while they appear to me much too soft, and too little separated, to take on such a definite shape' (Rudolphi 1821, 93). A few years later, the British physician Hodgkin and the wine merchant and microscope enthusiast Lister conducted research in histology with a new microscope that had been designed by Lister in collaboration with the London instrument maker William Tulley.[8] These two investigators simply denied the presence of globules in organic tissue. In 1830, the Leipzig professor of anatomy Ernst Heinrich Weber, elder brother of the physicist Wilhelm, again insisted that globules were integral parts of organic tissue but that they differed in size, and by no means were they all regular and perfectly spherical.

With the exception of Hodgkin and Lister, all of these microscopists observed globular elements, but of vastly different sizes and shapes. Notably, there were also reports of 'strings of beads', but again, the accounts differed. In his groundbreaking

work on the microscopical structure of the brain and nerves, published in 1833, the Berlin anatomist Christian Ehrenberg stated that according to his observations the nerves consisted of tubes, most of them smooth and cylindrical, some of them – the optical and auditory nerve and the nerves in the organ of smell – varicose (Ehrenberg 1833, 453). According to Ehrenberg, the cylindrical nerves contained a substance that consisted of 'small plump but not very regular particles', the nerve marrow (Ehrenberg 1833, 454). Ehrenberg also described smaller and bigger granules in the substance of the brain, but as the dominant element of the brain he identified the fibers. However, he stressed that other than the nerve fibers, the fibers of the brain were not simple cylindrical ones but 'resemble strings of beads' (Ehrenberg 1833, 452). And in 1835, Treviranus reported the results of his newest observations, which were made with a new, optically superior Plössl microscope (Treviranus 1835, VII). He had seen the nerves as rows of globules (Treviranus 1835, 34) and as fibers 'that often have the form of strings of beads' (Treviranus 1835, 35) And as late as 1836, Friedrich Arnold advocated a view very similar to Heusinger's, namely that 'the elementary parts of all structures of the body were originally globules or bubbles, that these line up in various ways and thus produce the sometimes in this, sometimes in that way shaped parts that many took for elementary parts' (Arnold 1836, xiii).

A Role for Error?

Even this brief survey has unearthed a multitude of descriptions of globules. Some accounts suggest that the globules are perfectly spherical or perfectly regular and everywhere the same, a few describe them explicitly as 'strings of beads', others combine this observation with a developmental hypothesis and claim that fibers originate from globules. Not all of the advocates of globules were romantics – Weber for example was an outspoken opponent of romantic thought – and some of the observations of beads on a string, for example Ehrenberg's and the later ones by Treviranus, were made with optically superior microscopes. Notably, this diversity of observations, descriptions, and theories is already indicated in the mid-twentieth-century histories: Baker, for instance, mentioned diverse claims advocated by globulists. But the emphasis of these earlier portrayals of the episode was on forging a progressive history of microscopy, and therefore the diversity of the microscopists' claims as such was not an issue.

It is perhaps not surprising that the picture of the episode that we obtain grows rather complex once we abstain from reconstructing the prehistory of the cell theory and focus instead on the practitioners' concrete efforts. One the other hand, if we stop looking at the discussions about the structure of organic matter from a presentist perspective and instead seek to do justice to the variety of findings that were initially announced, it seems that nothing is special about this episode. One would expect precisely this kind of debates in numerous scientific domains and epochs. Whenever a group of investigators seek to establish the facts of the matter in a new field, we may expect a fair amount of groping, scrambling, competing assumptions,

and dead ends. Nothing is wrong with that. And histology was a new field at that time, even though the microscope had existed for a couple of centuries: Only around 1800, the attention of anatomists and physicians moved from the organ to the tissue as the organizing unit of life and disease, and only then, preparation techniques were developed that helped make the tissue accessible for microscopical observation. So why not just acknowledge that the episode of the globule hypothesis was simply a period where the 'real' structure of organic tissue was at stake? Would it not be best to say that several practitioners contributed *tentative and preliminary* – rather than wrong – results to an ongoing investigative project, which, in the long run, established the nature of organic tissue? In short, why not say that no error occurred here? The project as such was not a failure; eventually the debates yielded a positive outcome. Of course scientists are perfectly justified and even required to try out various approaches and points of view, and in fact this is a common way of proceeding in scientific research.

So far, the outcome of my survey resembles Pickstone's narrative. The detailed historical analysis portrays these debates as a 'saner and more important' episode in the history of histology than the received view. However, if we look even more closely at how the practitioners around 1830 imposed order on the various accounts of organic tissue, it becomes clear that in a certain sense, we do have reason to resort to the concept of error in our rendering of the episode. Error plays an intriguing role in the practitioners' discourses about microscopic practice.[9] Let us consider in more detail how the practitioners introduced distinctions in their distribution of credibility over different accounts of organic tissue. In 1830, long before Virchow's polemics against globulism in the *Cellular Pathology*, Ernst Heinrich Weber assessed a particular version of the hypothesis, namely the 'string-of-beads account' of organic matter as erroneous. He did so in a very prominent place: in the widely read and influential fourth edition of Hildebrandt's *Handbook of Human Anatomy* that Weber had edited, thoroughly revised, and enriched with his own microscopical observations. In the chapter on the nervous system, Weber acknowledged that the recent investigations of nervous tissue had yielded an overabundance of results. Some of these he considered erroneous throughout, others he judged partially or nearly correct. What makes Weber's account so remarkable is not only the fact that he explicitly singled out a particular description of the nerves as 'erroneous' but also that he offered reasons why some practitioners had fallen into that error. Weber claimed that all those accounts, in which the globules were described as completely spherical, regularly shaped, and of the same size, were due to optical illusions. He even provided an innovative explanation for this phenomenon. It was not just the bad quality of his predecessors' instruments that produced these artefacts. Rather, it was the interaction of light with the microscope. In particular, the inflection and interference produced by direct or concentrated sunlight was 'very disturbing to the observation' and made distinct vision 'completely impossible; so that it is not the imperfection of our microscopes but rather the nature of light itself, which makes very bright illumination inadmissible, that sets narrow limits to the magnification of objects' (Hildebrandt 1830, 132). Past microscopists had fallen into error because they had unduly transgressed these limits.[10] Weber employed these arguments in a

lengthy critical review of earlier microscopical work, in which, as he claimed, the problematic consequences of the law of interference had been neglected. He showed the extent to which many seventeenth- and eighteenth-century microscopists, among them Leeuwenhoek, Monro and Fontana had applied too strong magnifications in combination with too strong illumination and had thus been misled in various ways. For this reason, Milne Edwards, Prevost and Dumas had seen muscle and nerve fibres as strings of beads. So Weber employed the notion of error in a critical fashion to highlight correct and incorrect methods of microscopical observation.

Note that while both Pickstone's and my reconstruction of the episode indicate that the historical record does not merit the application of 'error' as an appraisive term, Weber's contribution does call for the use of error as an actor's term in our narrative. It is significant that the notion of error serves the historical actor to make a methodological point, thereby of course also providing support for his own results. Needless to say, by taking up this usage of the term 'error' in *our* narratives, we do not automatically commit ourselves to writing an appraisive history of the present.

As to the merits of the concept of error, so far my analysis points to the conclusion that the notion of error is generally problematic as part of an appraisive history of the present and useful to the extent to which it reflects the assessments of the historical actors. If *they* use the term 'error' in their debates, then the reconstruction should reflect this fact. This conclusion is presumably largely uncontroversial. But note that even the sketch of the actors' assessments reveals some remarkable features of the debates. The reconstruction of Weber's argumentative moves leads to an interesting discovery about the role of error ascription in argumentation. In an article 'Accounting for Error', Michael Mulkay and Nigel Gilbert suggest that the term 'error' takes on a socio-psychological function in practitioners' debates. They argue that scientists account for their opponents' errors in psychological terms. While the practitioners portray their own 'correct' position as a cognitive phenomenon, which arises out of rational assessments of the available evidence, the opponents' erroneous beliefs are presented as due to the intrusion of distorting social and psychological factors (see Mulkay and Gilbert 1982). Mulkay and Gilbert's approach highlights an interesting feature of the debates in which Weber was involved. Weber did not refer to 'socio-psychological' factors that might have distorted other practitioners' results. Instead, he pointed to *practical* conditions that had produced the erroneous beliefs. Pragmatic explorations of the nature and limiting function of light were a crucial element in the management of the plethora of conflicting accounts of muscle and nerve fibres.

Can we draw any further conclusions about the role for error in the dynamics of such debates? The answer to this question depends on our epistemological outlook. In recent years, narrations of scientists' negotiations have served to substantiate social constructivist approaches to scientific knowledge. Social constructivists assume that in these negotiations the very distinction between 'true' and 'erroneous' factual statements is constructed. From this relativist point of view, those statements or beliefs that fail to gain general approval are – for precisely this reason – erroneous. The closure of the debate, it is implied, does not tell us anything about the facts of the matter – or indeed all there is to tell. Only if we settled for a social constructivist view

of scientific knowledge we could ascribe an epistemically productive role to error in precisely this sense.[11] Beyond the issue of social constructivism, it seems that the endeavor of 'accounting for error' is only worth noting if the actors' use of that term really made a difference to the debate. If the identification of error was instrumental for the closure of the histological debates, we would have reason to say that error played a formative role in this process of knowledge production.

At first glance, the historical record supports this move. Several microscopists followed Weber's verdict. Most probably inspired by Weber's explanatory account, Rudolph Wagner, physiologist at Göttingen, also identified the strings of beads as optical illusions. In 1832, Wagner reviewed Carl August Siegmund Schultze's textbook of comparative anatomy. The review reports Schultze's statement that nerve globules were present in very different animals; that they were completely spherical; and that all of them were of the same size (see Wagner 1832, 304). The significant passage follows in the next paragraph, which recounts Schultze's observations of muscular fibers. 'The author [Schultze] says: "The muscular fiber is a completely homogeneous mass, and in no animal have I been able to discover a trace of the globules of which it consists according to J. F. Meckel, Prevost and Dumas."' On the occasion of this observation, the review then brings up the problem of optical illusions. The text continues:

> As is well known, recently Bauer and Home, as well as Milne Edwards have also depicted the finest muscular fibers as strings of beads and as consisting of lined-up globules. Indeed, I must agree with Schultze's assumption; up to now I have not found any of those globules in the muscles at a magnification of 265 and 400; in general, the seeing of globules belongs to the most frequent optical illusions of the microscope, predominantly in strong magnifications, if one does not view the object quite in the correct focus (Wagner 1832, 304).

It is very likely that Wagner knew well and drew on Weber's writings at this point, because not much later he presented Weber's book explicitly as authoritative respecting the 'use, power, etc. of microscopes' (Wagner 1834, 49).

Other authors repeated – often to the letter – Weber's critique of earlier microscopists (e.g. Henle 1841, 139). In the 1830s, it became customary to restrict the magnification applied in anatomical investigations to 300–400 times. This magnification became the working limit of microscopy, as it were; it was referred to as the 'common' one. The trouble is that even though a number of the actors did take on Weber's view to characterize their opponents' views, we cannot say that this made the decisive difference in the discussion. Even though Weber's refutation of the string-of-beads account convinced some, it did not persuade everybody. The situation remained inconclusive for several years. Most notably, and most bewildering to the modern reader is the fact that Friedrich Arnold advocated in his 1836 Handbook the string-of-beads account of nerve fibers and supplied impressive pictures even though he had underwritten Weber's methodological cautions concerning light a few years earlier (cf. Arnold 1832, iv). The cases of Ehrenberg and Treviranus also show that Weber's refutation of the string-of-beads account was not decisive. In fact, Treviranus's observations complicate things even further because Treviranus did observe strings of beads, and he did agree with Weber that these were not genu-

ine, but he did not follow Weber's explanation. For Treviranus, the 'strings of beads' were due to artifacts of preparation, not to optical artefacts. In other words, even though they were not genuine – they did not represent the true form of organic tissue – they were real. Treviranus emphasized that this was due to the deterioration after death as well as to the treatment with water (Treviranus 1835, 38). Their true shape of the nerve fibers was cylindrical. These examples indicate that what holds for 'the' globule hypothesis also holds for 'the' string-of-beads account: a plethora of versions competed with one another.

In the case of the globule hypothesis the historical record does not give clear indications that the historical actors' identification of particular claims as erroneous brought the debates to a state of closure (not even a temporary one). And yet, if we just portray the episode in terms of 'preliminary' moves in the game of knowledge generation, we do miss an important, and, I suggest, epistemically productive aspect of the episode. We have seen that the practitioners encountered a proliferation of divergent results in the initial stage of their research endeavors when both the object and the methods of investigation were insufficiently delineated. The practitioners were acutely aware of the proliferation of conflicting accounts of the nerves. It was clear to them that not all the descriptions of nerve tissue that were advocated could be correct. This awareness was pivotal for the epistemic dynamics of the discourses. It had a profound effect on microscopical practice because it helped redirect the microscopists' attention from the objects of study to the means of investigation and inspired the negotiation of ways of going amiss in microscopic practice. These negotiations of correct and incorrect methods eventually chaneled the debates about the microscopical objects under study. This suggests an epistemic function for the practitioners' failings – and for the discourses about these failings – that is fundamentally different from the role of error as it is outlined in the received view. The received view and the notion of error that underpins it imply that the 'erroneous' globule hypothesis was an impediment or obstacle on the way to a more correct view and a more advanced stage of microscopical theory. My reconstruction of the episode suggests that a proliferation of conflicting accounts occurred, and that it was the fact that the practitioners had become acutely aware of this proliferation that was a crucial step for the closure of the debates.

While we would not want to say that some or all of the practitioners had fallen into error, we might very well say that something was going amiss. The proliferation of accounts was perceived as something out of course. The conceptual advantage of the notion 'something is going amiss' is that its application does not require us to identify concrete claims as 'correct' and others as 'erroneous'. Using this term, we are not committed to making specific appraisive judgements about particular claims to knowledge. Instead, we appraise whole sets of observations in terms of 'coherence/incoherence' or whole sets of practices (as they are described in the scientists' writings) in terms of 'uniformity/diversity' or 'stability/instability'.

But do we really need to introduce a new term to characterize the debates? Could we not frame them as confusions? After all, 'confusion' also refers to disordered conditions, and therefore it appears perfectly appropriate for the characterization

of the debates. – Nevertheless, there are two problems with this concept. First, it is ambiguous: The term confusion mainly refers to an *individual's* mental discomfiture or perturbation. This kind of confusion can certainly be epistemically productive,[12] but it is not the kind of confusion that made the globule episode so fruitful. The epistemic dynamic of this episode originated in the manifold of conflicting observations, which were put forward by diverse observers. Secondly, the term 'confusion' suggests *fusion*, the *blending* of distinct elements; and again, this is not what happened in the episode in question. Rather, the individuals put forward distinct descriptions of organic tissue, which proved inconsistent with each other.

The notion of controversy also suggests itself. Why not say that the globule episode was a controversy about the correct description of organic tissue? It seems to me that if we did so, we would miss an important feature of the discussions. The term 'controversy' implies that two or more people were involved in a prolonged antagonistic debate about whose views were correct. But in an important respect, the negotiations about the descriptions of organic tissue were *not* antagonistic: Many, if not all, *agreed* that the available histological descriptions did not yet have a firm observational basis and were insufficiently developed. Many, if not all practitioners *agreed* that something was amiss.

Earlier in this paper, I pointed out that presentist reconstructions of past scientific endeavours appraise science in terms of the evaluative standards that prevail in the historian's own time. These reconstructions assess claims to knowledge as well as past scientists' methods in terms of today's standards. In contrast, contextualized reconstructions seek to evaluate past scientists' claims and methods against the scientists' own standards. Following Pickstone, we may conclude that in this perspective, the globule hypothesis was not in fact an error. On the other hand, Weber's part in the episode suggests that we can fruitfully employ the term as an actor's category to portray past practitioners' assessments of their contemporaries' and predecessors' work. In this context, it makes perfect sense to ask for the reasons why particular historical actors – in the present case, Weber – had come to identify certain particular observations as erroneous, and how they accounted for error.

However, these options do not exhaust the range of possible approaches to the episode. My reconstruction shows that something *was* out of course in the decades after 1800. Because the microscopists were faced with a proliferation of mutually exclusive observations, it makes perfect sense to say that something was going amiss. Note that this analysis has again an appraisive dimension. For saying that 'something was going amiss' reflects our tacit assumption that observations of similar objects under similar circumstances should yield similar results: if this is not the case, something has gone amiss. In this context, we are introducing the term 'something is going amiss' as an analytic category that is not an actors' term. Nick Jardine has introduced the notion 'viciously anachronistic' as a label for those categories alien to the period in question that constitute 'historically incoherent interpretations of past deeds and words' (Jardine 2000, 252). 'Going amiss' is not viciously anachronistic in this sense. Of course, by saying that 'something

was going amiss' in early nineteenth-century histology, we do project onto the practitioners our assumption that observations of similar objects under similar circumstances should yield similar results. We attribute to the practitioners the same assumption. But the historical record supports this move. More importantly, even though the practitioners did not explicitly apply the concept of 'going amiss', the historical record of their works demonstrates that the realization that conflicting descriptions proliferated was productive. It led to a reorientation of histological practice.

Conclusion

It is a commonplace to say that scientific practice is infused by error. Hardly any philosopher, historian, or scientist would deny it. However, as my paper has shown, 'error' is a precarious concept, whose applicability is dependent on the historical and philosophical questions that we ask. We have seen that both as an appraisive and as an actor's term, the notion of error can be applied in a meaningful way. But to avoid misunderstandings, we need to attend carefully to the distinction between error as an appraisive and as an actor's category. We also need to be aware that there are different ways of appraising past science: histories of the present measure past scientists' works against the knowledge of today. Contextualized histories seek to measure past scientists' works against the standards of their own time. Still, to the extent to which these reconstructions exhibit inconsistencies, incoherence, and the like, they remain appraisive.

My discussion has also demonstrated that the exploration of error in scientific practice is not merely an assessment of putative failures in light of past and present standards. If scientific practice is error-ridden, an adequate analysis must capture scientists' struggles to identify, remove, or cope with errors as well as their struggles with all kinds of other challenges, including artifacts, noise, an overabundance of conflicting results, and dead ends. To capture the richness of these challenges, we need to devise novel conceptual tools. We need to engage in 'constitutive anachronism'. This term is inspired by another of Jardine's notions, namely, 'interpretative anachronism', which serves as a label for the application of 'categories from one period to deeds and works from a period from which those categories were absent' (Jardine 2000, 253). For the purpose of exploring error in scientific practice, however, the principal task is not interpretation. First and foremost, we need to *expound* a set of conceptual tools – hence *constitutive* anachronism. Nevertheless, constant engagement with scientists' works is required to develop sets of conceptual tools that can help analyze the dead ends, the noise, and the confusions that are integral parts of 'getting it right' in science. It is in this constitutive-analytic perspective that the notion of 'going amiss' proves useful as a conceptual tool.

We can go still further. The fact that 'something was going amiss' made an epistemic difference. For an extended period, the practitioners themselves were

quite aware of the fact that the state of affairs in histology was inconclusive and that the results obtained so far had to be reassessed. This shift of attention was of critical importance for the unfolding of the debates. In hindsight, we can also say that proliferations of claims to knowledge such as the proliferation of observations of globules were extremely fruitful. The practitioners encountered a proliferation of divergent results in the initial stage of their research endeavors when both the object and the methods of investigation were insufficiently delineated. And because they encountered a plethora of conflicting results, the investigative practices themselves became an issue. In one important sense, the ensuing discussion about methods and means was not antagonistic. Many, if not all, practitioners *agreed* that the histological observations of tissues were insufficiently secured. This agreement, born from proliferation, became a driving force in the debates. In particular, the fact that various competing accounts of nerve and muscular tissue were put forward motivated not only repetitions of observations but also a specific search for the factors of the setting that had an impact on the results, a search for potential sources of error. The increase of discordant findings redirected attention to the tools and circumstances that could affect the outcome of the research, and as a consequence, the anatomists began to look into the probable causes of the differences in their results. To analyze this, we need the nomenclature of 'going amiss'. This kind of 'going amiss', which may well be a common feature of knowledge generation in experimental practice is an epistemically productive force because it inspires concern with, and research on the tools of research. It is due to this insight that the story that I have presented is more than a simple correction of an error in the historiography of microscopy.

Acknowledgements Portions of this material were presented at the *HSS Annual Meeting* in Austin, Texas in November 2004 and at the workshop *Generating Experimental Knowledge* in Haifa, Israel in May 2005. I am grateful to Domenico Bertoloni Meli, Elizabeth Cavicchi, Igal Dotan, Melinda Fagan, Uljana Feest, Sander Gliboff, Hans-Jörg Rheinberger, and Gábor Zemplén for instructive comments on preliminary versions of this article.

Notes

1. Historians have used the two expressions interchangeably, and I will continue to do so in this essay.
2. Of course, cells had been described long before the 1830s, and many other microscopists contributed to the formation of cell theory (see Bechtel 1984, Parnes 2000). Schleiden and Schwann effectively synthesized much of this earlier work.
3. For an early version of this technology-centred approach, see Studnicka 1932.
4. See also L. J. Rather, who tried even harder to present the eighteenth-century works as prehistory of the cell theory. He claimed that Ackerknecht had not gone far enough and that the designation 'cell theory No. 1' should 'properly be given to the "fiber" theory'. He added that the fibers had a firmer observational basis than the globules, which were 'almost certainly optical artefacts due to spherical aberration' (Rather 1969, 200).
5. This move was quite common among mid-nineteenth-century scientists. We find such polemics also in the writings of the so-called 'organic physicists', the group of physiologists around Helmholtz (see, for example, Du Bois-Reymond (1860) 1912) or in Justus Liebig's work (e.g.

Liebig 1996, 50). Nicholas Jardine's detailed study on this topic shows how du Bois-Reymond and Virchow used history – and particularly the historical account of the history of anatomy and physiology in the early 1800s – to advance their own research agendas (Jardine 1997).

6. In this sense, my account complements Pickstone's, which concentrates on France and Britain.
7. The papers on microscopy that Home published in the 1820s merely report works of other people. Although Home was nominally the author of these papers, most of the microscopical investigations that the papers present had been carried out by Francis Bauer (on Bauer, see Meynell 1983).
8. For the context of this collaboration, see Schickore 2003.
9. For the history of microscopists' discourses about their practice, see Schickore 2007.
10. Weber's references to 'inflection' and 'interference' seem to indicate that the refutation of the globule hypothesis was not so much inspired by theoretical or technological progress but rather by the 'optical revolution'. This is not altogether surprising: After all, the period around 1830 was the time when the long-standing corpuscular theory of light was being replaced by the wave theory (see Buchwald 1989). And indeed, Weber was one of the early advocates of the wave theory of light – his *Wellenlehre*, published in 1825 together with his brother Wilhelm, gives evidence of this fact. But it would be misleading to make the fate of the string-of-beads hypothesis dependent on a new theory of instrumentation, namely the new undulation theory of light. The microscopists' appropriation of optics for the purposes of the methodology of microscopy was in fact purely qualitative. In the anatomical *Handbook*, Weber did not bother with theoretical explanations of the terms 'inflection' and 'interference'. He merely noted that the microscopist would encounter problems if the illumination was 'very strong' and the magnification was 'very considerable' (Hildebrandt 1830, 132).
11. See, for example, Schlich 1993 on the dispute about diabetes between Eduard Pflüger and Oscar Minkowski. For another case study in the constructivist spirit see Ashmore 1993, which concentrates on the reception of the discovery of N-rays.
12. For examples of this kind of productive confusion see Cavicchi 1997 and Graßhoff and Nickelsen (this volume).

References

Ackerknecht, E. 1953. *Rudolf Virchow. Doctor, Statesman, Anthropologist*. Madison: The University of Wisconsin Press.

Arnold, F. 1832. *Anatomische und physiologische Untersuchungen über das Auge des Menschen*. Heidelberg und Leipzig: Neue akademische Buchhandlung von Karl Groos.

Arnold, F. 1836. *Lehrbuch der Physiologie des Menschen*. Zürich: Orell, Füssli und Compagnie.

Ashmore, M. 1993. "The theatre of the blind: Starring a Promethean prankster, a phoney phenomenon, a prism, a pocket and a piece of wood." *Social Studies of Science* 23: 67–106.

Baker, J. R. 1948. "The cell theory, a restatement, history and critique. Part I." *Quarterly Journal of Microscopical Science* 89: 103–125.

Bechtel, W. 1984. "The evolution of our understanding of the cell: a study in the dynamics of scientific progress." *Studies in History and Philosophy of Science* 15: 309–356.

Berg, A. 1942. "Die Lehre von der Faser als Form- und Funktionselement des Organismus: Die Geschichte des biologisch-medizinischen Grundproblems vom Kleinsten Bauelement des Körpers bis zur Begründung der Zellenlehre." *Virchows Archiv für pathologische Anatomie und Physiologie und für klinische Medizin* 309: 333–460.

Buchwald, J. Z. 1989. *The Rise of the Wave Theory of Light. Optical Theory and Experiment in the Early Nineteenth Century*. Chicago: The University of Chicago Press.

Burdach, K. F. 1819. *Vom Baue und Leben des Gehirns*, vol. I. Leipzig: In der Dyk'schen Buchhandlung.
Cavicchi, E. 1997. "Experimenting with magnetism: ways of learning of Joann and Faraday." *American Journal of Physics* 65: 867–882.
Du Bois-Reymond, E. 1912. "Gedächtnissrede auf Johannes Müller" [1860], pp. 135–317, *Reden*, vol. I, 2nd edition. Leipzig: Veit & Comp.
Ehrenberg, C. G. 1833. "Nothwendigkeit einer feineren mechanischen Zerlegung des Gehirns und der Nerven vor der chemischen, dargestellt aus Beobachtungen von C. G. Ehrenberg." *Annalen der Physik und Chemie* 28 (104): 449–473.
Henle, J. 1841. *Allgemeine Anatomie. Lehre von den Mischungs- und Formbestandtheilen des menschlichen Körpers*. Leipzig: Leopold Voß.
Heusinger, C. F. 1822. *System der Histologie*. Eisenach: Bärecke.
Hildebrandt, F. 1830. *Handbuch der Anatomie des Menschen. Erster Band: Allgemeine Anatomie*, vol. I, 4. Braunschweig: Verlag der Schulbuchhandlung.
Home, E. 1821. "The Croonian Lecture: microscopical observations on the following subjects. On the brain and nerves; showing that the materials of which they are composed exist in the blood. On the discovery of valves in the branches of the vas breve, lying between the villous and muscular coats of the stomach. On the structure of the spleen." *Philosophical Transactions of the Royal Society of London* 111: 25–46.
Jardine, N. 1997. "The Mantle of Müller and the Ghost of Goethe: on the entanglements of the sciences in history". In: Kelley, D. R. (Ed.), *History and the Disciplines: the Reclassification of Knowledge in Early Modern Europe*. Rochester: University of Rochester Press, pp. 297–318.
Jardine, N. 2000. "Uses and abuses of anachronism in the history of the sciences." *History of Science* 38: 251–270.
Liebig, J. v. 1996. "On the study of the natural sciences". In: Schirmacher, W. (Ed.), *German Essays on Science in the 19th Century*. New York: Continuum, pp. 46–56.
Meckel, J. F. 1815. *Handbuch der menschlichen Anatomie. Erster Band. Allgemeine Anatomie*, vol. I. Halle (Saale): Hallisches Waisenhaus.
Meynell, G. 1983. "Francis Bauer, Joseph Banks, Everard Home and others." *Archives of Natural History* 11: 209–221.
Milne Edwards, H. 1827. "Mikroscopische Untersuchungen über die innerste Structur der thierisch-organischen Gewebe." *Notizen aus dem Gebiete der Natur- und Heilkunde* 355: 33–41.
Mulkay, M. and G. N. Gilbert 1982. "Accounting for error: how scientists construct their social world when they account for correct and incorrect belief." *Sociology* 16: 165–183.
Parnes, O. 2000. "The envisioning of cells." *Science in Context* 13: 71–92.
Pickstone, J. V. 1973. "Globules and coagula: concepts of tissue formation in the early 19. Century." *Journal of the History of Medicine* 28: 336–356.
Rather, L. J. 1969. "Some relations between eighteenth-century fiber theory and nineteenth-century tissue theory." *Clio Medica* 4: 191–202.
Rudolphi, K. A. 1821. *Grundriss der Physiologie*, vol. I. Berlin: Dümmler.
Schickore, J. 2003. "Cheese mites and other delicacies: The introduction of test objects into microscopy." *Endeavour* 27: 134–138.
Schickore, J. 2007. *The Microscope and the Eye: A History of Reflections, 1740–1870*. Chicago: University of Chicago Press.
Schlich, T. 1993. "Making Mistakes in Science: Eduard Pflüger, His Scientific and Professional Concept of Physiology, and His Unsuccessful Theory of Diabetes (1903–1910)." *Studies in History and Philosophy of Science* 24: 411–441.
Studnicka, F. K. 1932. "Aus der Vorgeschichte der Zellentheorie von H. Milne Edwards, H. Dutruchet, F. Raspail, J. E. Purkinje." *Anatomischer Anzeiger* 73: 390–416.
Treviranus, G. R. 1816. "Ueber die organischen Elemente des thierischen Körpers". In: Treviranus, G. R. and L. C. Treviranus (Eds.), *Vermischte Schriften anatomischen und physiologischen Inhalts*, vol. I. Göttingen: Röwer, pp. 117–143.

Treviranus, G. R. 1835. *Neue Untersuchungen über die organischen Elemente der thierischen Körper und deren Zusammensetzung*, vol. I, in: Beiträge zur Aufklärung der Erscheinungen und Gesetze des organischen Lebens. Bremen: Heyse.

Virchow, R. 1871. *Die Cellularpathologie in ihrer Begründung auf physiologische und pathologische Gewebelehre*, 4th edition. Berlin: Hirschwald.

Wagner, R. 1832. "Rezension: Systematisches Lehrbuch der vergleichenden Anatomie." *Litterarische Annalen der gesammten Heilkunde* 24: 299–309.

Wagner, R. 1834. *Lehrbuch der vergleichenden Anatomie*. Leipzig: Voss.

Part II
Learning from Error

Learning Without Error

Erez Braun and Shimon Marom

Learning is a process of growing success in a fixed environment. One can only speak about learning when behavior noticeably increases the efficiency with which information is processed so that desirable states are reached, errors are avoided, or a portion of the world is controlled[1]. The concept of error plays a key role in general learning theories, and reinforcement learning in particular. In these theories[2] a dedicated entity is invoked, whose function is to compare the state of the system to a desirable state, and to produce an error signal that drives the system to change. That such is the case in machine reinforcement-learning protocols, one cannot argue. But, what about learning in biological systems? We submit that, if not for deep philosophical reasons, the schematics portrayed for reinforcement learning in machines cannot hold for biological systems. For an error signal to be produced, states should be measured and compared; but since biological states are practically infinite object that are not local in time nor in space[3,4], the scales and standards required for measurements and comparisons do not exist. We provide examples for state-space immensity[5] at three levels of biological organization that are intimately related to the subject matter: molecules, cells and behavior. We then comment on the impacts of this immensity on learning and on the practice of experiments in biology.

As a representative of state space immensity at the molecular level, consider the number of possible states in which particular proteins, the voltage-gated ionic-channels family, may reside. These proteins are responsible for shaping the time-amplitude envelope of neuronal and synaptic signals, and their state space is described as highly relevant for neuronal activity (the interested reader is referred to the authoritative monograph by B. Hille[6]). Molecular biologists and physiologists have convincingly shown over the past two decades that the number of types of ionic channel proteins that a single neuron expresses at any given point in time is in the order of ten. Each type of channels is composed typically of five to ten sub-types that form combinatorial structures that have different functional consequences. Depending on that combinatorial structure, the channel proteins are extensively engaged in

E. Braun (✉)
Network Biology Research Laboratories, and Departments of Physics (E.B.) and Medicine (S.M.), Technion, Haifa 32000, Israel
e-mail: erez@physics.technion.ac.il

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_4,

interactions with ligands and with other proteins, interactions that have significant functional consequences. Furthermore, recent estimates and measurements indicate that even a bare, stand-alone ionic channel protein, has intrinsic to it a large number of possible states[7]. Other examples may be given for molecular immensity within the domains of synaptic functions, second messengers and related signal transduction pathways.

The molecular immensity exemplified above does not disappear at higher levels of brain organization. Consider, for instance, a cortical neuron in its network: It is known that the number of synapses impinging upon a single neuron in the human cortex is well beyond one hundred. In fact, estimates are that each cortical neuron is affected by the states of thousands of other neurons (the interested reader is referred to the monograph by Abeles[8]). This means that, even for the most simple case, in which the spatial and temporal attributes of a single neuron are not considered, and even if one assumes that synaptic inputs come in packets of several hundred of correlated synapses, the number of possible single-neuron states is immense. Indeed, careful analyses of neural activity time series, recorded from neurons in wake animals, show that these neurons have no preference for any limited set of uniquely defined states. Rather, their activity characteristics are consistent with a continuum-of-states model. Strong claims were made, supported by exhaustive statistical analyses, pointing to the curious fact that the number of neuronal states, reflected in the activity of neurons, is practically limited only by the experimental conditions and the rituals of statistical analyses[9].

What about overt behavior? Unlike the case of neurobiology, where the space of possible brain states is defined, the language is agreed upon and the problem is of identifying relevant states, in psychology there seems to be no consensus on the actual definition of the state-space itself[10]. This alarming situation is further complicated by the fact that whereabouts precise analysis is feasible, experimental findings indicate that behavior cannot be decomposed to a fathomable number of uniquely defined states. Consider, for example, an experimental psychology field that is (arguably) most approachable for scientific analysis, namely- the study of memory. In 1885, Ebbinghaus reported his seminal introspective study of human memory. He demonstrated that while the retention of nonsense syllables decreases as time elapsed from the initial learning increase, the rate at which that decrement occurs monotonically slows down, that is to say, not fixed. Ebbinghaus intuited that the mathematical form of the forgetting function is logarithmic. Unlike the exponential function commonly used to describe relaxation data, which would imply a fixed rate of forgetting in that case, Ebbinghaus's logarithmic function implies a rate of forgetting that depends on the time elapsed since the learning. Thus, if one accepts the physical-chemistry principle equating a single state with a single rate, Ebbinghaus's interpretation implied no unique memory states because no unique rates are found[11]. Notwithstanding a transient belief in the short-term memory theory during the early 1970's, cognitive psychologists now have ample evidence in favor of Ebbinghaus's interpretation. In fact, re-analysis of Ebbinghaus's data, taken together with a host of new data, indeed confirm that memorizing and forgetting cannot be described by a fathomable set of uniquely defined rates[12], and therefore no fathomable set

of uniquely defined states are involved. Thus, we see that in the behavioral level of organization the problem is twofold. First and foremost, behavioral states are ill defined, leading to a conceptual alienation. Second, behavioral entities that are approachable for precise analysis reveal an immense space of possible states.

The examples given above show that immensity of state space reigns at every level of biological observation, from molecules to behaviour. We conclude that the scales and standards required for the identification of a state and hence an error cannot be constructed in a biological system. This is the case already at the molecular level and definitely at higher levels (cells, organs, organisms and populations), which are aggregates of molecules. No error assignment can be exercised under such conditions. While the above examples are from the discipline of neural sciences, the picture is practically identical in many other biological systems (e.g. immune system, genetic networks and development).

Another possibility to construct scales and standards is by attributing functionality to a biological system. In that case error would be defined in terms of deviations from the "standard" functionality. While functionality is the hallmark of biological systems, it is an alien concept for the other natural sciences. In that respect, biology is closer to engineering sciences and man made machines. However, in contrast to engineering, biology is an historical science; presently observed configurations reflect accumulation of accidental events over evolutionary time scales, selection processes, multi-functionality, an immense number of entailments between functions, redundancies and overlap within and between levels of organization. These facts, taken together with the above-mentioned immensity of degrees of freedom preclude the possibility of adapting a given functionality as a standard for error detection. As a result, assigned functionality reflects the point of view of a given observer rather than that of the "designer". For instance, going back to the channel protein mentioned above: It is known that the protein can function only within a narrow voltage range around a set point determined by the gradient of ionic concentrations across the cell membrane. From the point of view of electrical functionality one may assume that the ionic concentration has been optimized to support the activity of the ionic channel proteins. This functionality of ionic concentrations, however, cannot serve as standard for error detection because the set point is similar in practically all cell types, including those that are not generators of electrical signals. The fact is that the ionic concentration gradients are key determinants of many, unrelated physiological processes within the same cell; hence the requirement for evolutionary selection. In order to understand the design principles behind ionic concentrations one needs to uncover the entire historical path and the set of all the involved interactions. Indeed, understanding the functional relations between the components of the biological system is possible only from the evolutionary point of view.

From the above we are again forced to conclude that lack of standard makes the concept of error irrelevant in the context of biological systems. This conclusion introduces serious problems in attempts to understand the biology of learning in its wider sense. For whatever definition we use for learning, a measure is required for the gap between present and desired configuration; in other words, a measure of the error is

required. Under these conditions such an error measure is translated to driving force that modifies the configuration of the system, aiming at error reduction. Lack of scales and standards thus presents us with a challenge to understand learning in the biological context. How does the immune system learn to identify pathogens? How does a colony of ants learn to construct nests, or to form patterns of tracks towards food sources? How do bacteria swim upstream food gradients given a single bacterium small scale? In fact, how do we, human beings, learn? What are the mechanisms allowing for adaptive reconfiguration of the immense number of entities involved in learning?

The common feature in all learning phenomena is the existence of exploration in configuration space. What makes learning in biological systems unique is the fact that, unlike man-made machines, the driving force for the exploration does not scale with the gap between present position and desired one. Rather, it is only dictated by local measures, irrespective of its distance from target.

For example, let us consider learning in the neural system. We know, from every day experience, that it is widely accepted to use evaluative-concepts in descriptions of learning in psychology; i.e., "appropriate" behaviour is "rewarded", "right" actions are "positively reinforced" and "wrong" are "negatively reinforced" (i.e. "punished"). Such language usage implies that in learning, surely, error measures are used and therefore standards are required. At the beginning of the twentieth century, the modern terminology of learning was established. Rules of association by simultaneity and temporal sequences involved in instrumental conditioning were defined. It became clear that the concepts of reinforcement, reward, punishment, are extremely useful in describing and controlling behaviours. Attempts to understand how the concept of reward is realized in a biological world that lacks standards were made by eminent psychologists such as Hull[13] and Guthrie[14] over fifty years ago and even earlier by Freud[15] and James[16]. The resulting learning theories, which may collectively be referred to as Drive Reduction theories, stress the effect of reward on the driving stimulus. Specifically, the reward acts to reduce the stimulus that drives the exploration process. This reduction is based on local cues and precludes the acquisition of new stimulus-response entailments. Sharpening the stimulus-response entailment, in turn, is achieved through a selection process. Such description of learning in neural systems classifies the operation of the brain as a Darwinian process, similar to the other above-mentioned biological systems. That is, no separate neural rewarding entity is postulated or needed for shaping behaviour. In fact, one may find texts that explicitly reject mapping of evaluative behavioural concepts to defined brain entities, suggesting that the concept of error does not belong to the neural system itself, but rather to the larger complex that contains the environment, the system and the observer. Here is, for instance, what Guthrie said in his presidential address to the American Psychological Association in 1946:

> Psychologists who think in terms of punishment and reward have almost uniformly neglected to note how the animal at the time responded to the punishment or to the reward, and the role this played in subsequent behavior. The resulting generalization is inevitably an attempt to link the intentions of the experimenter (intentions to reward or punish) with good or bad behavior on the part of the animal. Punishment and reward are, objectively viewed, stimuli acting on the animal's sense organs, and their effect must be mediated through the

> animal's nervous system and appear in muscular contraction or glandular secretion. Since levers and loops and mazes are not innervated, the operations of these devices are incidental to the actual learning which the living animal performs.[17]

In spite of the above, the prevailing trend is to describe biological processes, including brain functions, in mechanistic terms of error detection, non-local driving forces and optimization processes[18]. Indeed, parts of biology may be described in such terms; the price being loss of the global, system point of view. It is acknowledged that mechanistic approaches to biology reflect present technological forefronts and are effective for specific applications (e.g. medical treatments). A most notable example is the interaction between the impressive technological developments of computing machines and brain research. While, as mentioned above, the general learning theories of the early 1920's explicitly advocated avoidance of attempts to map evaluative concepts to specific brain structures, nowadays such mapping dominates neuroscience. This shift reflects the dominance of the computer paradigm in brain research. Most algorithms used for effective machine learning are supervised ones, where an additional source of information, of knowledge of the error, dictates the drive and directs the learning process. Interestingly, when cognitive psychology, heavily relying on computational theories, entered the arena and practically removed behaviourism and general learning theories from the scene, it brought with it the error-based algorithmic computational approach.

History teaches us that the duality of mechanistic and Darwinian aspects will continue to drive the biological research. Jerne describes this historical pattern in a text written in 1967:

> Looking back into the history of biology, it appears that wherever a phenomenon resembles learning, an instructive theory was first proposed to account for the underlying mechanisms. In every case, this was later replaced by a selective theory. Thus the species were thought to have developed by learning or by adaptation of individuals to the environment, until Darwin showed this to have been a selective process. Resistance of bacteria to antibacterial agents was thought to be acquired by adaptation, until Luria and Delbrück showed the mechanism to be a selective one. Adaptive enzymes were shown by Monod and his school to be inducible enzymes arising through the selection of pre-existing genes. Finally, antibody formation that was thought to be based on instruction by the antigen is now found to result from the selection of already existing patterns. It thus remains to be asked if learning by the central nervous system might not also be a selective process; i.e., perhaps learning is not learning either.[19]

This duality presents a challenge to experimental biologists. Setting up experimental designs aimed at exposing mechanistic aspects of a given biological system is a natural extension of prevailing paradigms in engineering and physical sciences. However, uncovering the Darwinian aspects of biological systems requires new experimental concepts. The experimental design, in such cases, should allow the observed system to control its driving forces based on interactions with the environment. In other words, standards reflecting the experimental constraints should be eliminated. Results from such experiments will enable development of comprehensive understanding the unique aspects of biology, and may serve as a basis for a paradigm shift in engineering.

Notes

1. Klaus Krippendorff's Dictionary of Cybernetics (URL= http://pespmc1.vub.ac.be/ASC/LEARNING.html)
2. E.g., Sutton and Barto (1998)
3. Rosen (1991)
4. A term introduced in Elsasser (1987)
5. Elsasser (1987)
6. Hille (1992)
7. E.g., Toib et al. (1998); Ellerkmann et al. (2001); Gilboa et al. (2005)
8. Abeles (1991)
9. E.g., Teich et al. (1997)
10. E.g., the alienation between the concepts of cognitive psychology and those of psychodynamics.
11. Interestingly, the modern concept of scale-free distribution does fit Ebbinghaus's description; scale-free distributions are often interpreted as indicating (practically) continuum of states. See also note 7 and 9 above.
12. E.g., Wixted and Ebbesen (1997)
13. Hull (1943)
14. Guthrie (1946)
15. Freud (1895/1966)
16. James (1890)
17. Guthrie (1946).
18. E.g., Hollerman and Schultz (1998)
19. Quarton et al. (1967), p. 204

References

Abeles, M. (1991). Corticonics: Neural Circuits of Cerebral Cortex. Cambridge: Cambridge University Press.

Ellerkmann, R. K., V. Riazanski, et al. (2001), J Physiol 532(Pt 2): 385–97

Elsasser, W. M. (1987). Reflections on a Theory of Organisms: Holism in Biology. Baltimore: The Johns Hopkins University Press.

Freud, S. (1895/1966). Project for a Scientific Psychology. The Standard Edition of the Complete Psychological Works of Sigmund Freud. Pre-Psychoanalytical Works and Unpublished Drafts. J. Strachey. London: The Hogarth Press. 1: 283–360.

James, W. (1890). Principles of Psychology. New York: Dover Publications, Inc.

Gilboa, G., R. Chen, et al. (2005), J Neurosci 25(28): 6479–89.

Guthrie, E. R. (1946). "Psychological facts and psychological theory." Psychol Bull 43: [Presidential address of the APA, Evanston, Illinois, 1945].

Hille, B. (1992). Ionic Channels of Excitable Membranes, Sunderland, MA: Sinauer.

Hollerman, J. R. and W. Schultz (1998). Nat Neurosci 1(4): 304–309.

Hull, C. (1943). Principles of Behavior. New York: Appleton-Century-Crofts.

Quarton, G. C., et al. (1967) eds., In the Neurosciences: A study program (Rockefeller University Press), pp. 200–205

Rosen, R. (1991). Life Itself. New York: Columbia University Press.

Sutton, R. and A. Barto (1998). Reinforcement Learning: An Introduction. Cambridge, MA: MIT Press.

Teich, M. C., C. Heneghan, et al. (1997), J Opt Soc Am A Opt Image Sci 14(3): 529–546.

Toib, A., V. Lyakhov, et al. (1998), J Neurosci 18(5): 1893–1903.

Wixted, J. T. and E. B. Ebbesen (1997), Mem Cognit 25(5): 731–739; and references therein.

Living Extremely Flat: The Life of an Automaton; John von Neumann's Conception of Error of (in)Animate Systems

Giora Hon

> Proofreading or "editing" has been suggested in DNA replication…but a detailed description of its chemical kinetic basis is lacking. The problem is thus to find a simple quantitative model containing the essential features of proofreading scheme.…These circumstances allow the construction of a simple mechanism of "kinetic proofreading."[1]
>
> John J. Hopfield, 1974

Half a century ago, in January 1952, in a lecture delivered at the California Institute of Technology, John von Neumann (1903–1957) envisaged the synthesis of reliable organisms from unreliable components. This was not a science-fiction talk, calling for imaginative creations in the spirit of Ridley Scott's *Blade Runners*. It was a carefully argued scientific paper in which von Neumann sought to prove the existence of a self-reproducing universal computer. The paper constitutes an important contribution to the consolidation of the theory of automata. Von Neumann did not conceive of cellular automata as mathematical objects for pure investigation; rather, he considered the new algorithm a means for treating in detail the problem of how to make machine reproducible.[2] The realization that cellular automata can demonstrate that "arbitrarily complicated mathematics could be performed within a system whose basic organization is thoroughly rudimentary,"[3] is a testimony to the success of von Neumann's idea. Indeed, his construction shows that "a small set of local rules acting on a large repetitive array can result in a structure with very complex behavior. The von Neumann construction thus immediately suggests how an organ with behavior as complex as the brain's can be specified from limited genetic information."[4]

To get the basic terms clear, cellular automata are "abstract dynamical systems that play a role in discrete mathematics comparable to that played by partial differential equations in the mathematics of the continuum."[5] These dynamical systems

G. Hon (✉)
Department of Philosophy, University of Haifa, Israel
e-mail: hon@research.haifa.ac.il

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_5,

consist of arrays of computing elements characterized by discreteness in space, time and state values, whose architecture and design followed initially studies of the nerve systems of mammals: the fact that "the cerebral cortex is composed of a large number of local neural assemblies that are iterated throughout its extent."[6] An essential feature of this computational technique is that the functions computed are functions of the internal states of the computing elements *and* the inputs from neighboring elements. Thus the role of a cellular operation is to transform an array of data displayed in a discrete space at time *t,* into an array of data at time $t+1$. At this time each element of the array has a value which is determined not only by its initial state, but also by the values of its nearest neighbors. Operations are assumed to occur in discrete time with each step in time being a *generation*, *iteration*, or *cycle*. Further, changes in all computing or processing elements, are taken to be simultaneously. The action, in other words, of all the elements in a cellular array is synchronous. This action is governed by a *transition rule* or *transform* that uses as its independent variables the states of the particular computing element and its neighbors. Finally, the transition rule is considered *local* (no action-at-a-distance), *uniform* (the same rule applies to all sites at all times) and *deterministic*: any given configuration of the states of the elements of the array has just one successor configuration for a given transform. Note crucially that no centralized authority, so to speak, governs the evolution of the system which, on the contrary, evolves as per local interactions between a single cell and its neighboring cells. Thus, no general predictive procedure is possible, that is, there is neither an analytical expression nor a short-cut in the computational process—the system simply has to exhaust its runs. In other words, the evolution of such systems effectively defines the most efficient simulation of their behavior.[7]

For further clarification one may draw a comparison between cellular automata and the traditional model of computation, the Turing machine. The former scheme of calculation is parallel while the latter is serial. Cellular automata have no "head" to "read" a sign and put it in relation to a subsequent sign; rather, the computation proceeds in parallel across the entire lattice—the multiple cells that comprise the automaton. Moreover, cellular automata have no halting states and therefore it is difficult to separate in such schemes the dynamics of the system from the computation.[8] This means that the dynamics becomes an expression of the computation and consequently the simulations which is expressed in the computation can be seen—in some cases, directly—in the dynamics of the automaton.

Notice that in the Turing machine there is a clear separation between on the one hand the structural part of the computer which is fixed, and on the other hand the data which are variable and do not belong to the material structure of the computer. In other words, the computer cannot operate on its own matter; it cannot extend or modify itself, or build other computers. This is not the case with cellular automata where, as we have seen, it is difficult to distinguish between the dynamics of the system and the computation itself. This suggests the following characterization of these two different schemes: while the Turing machine is predominantly spatial, the cellular automata are essentially temporal.[9]

These features make cellular automata conducive to simulating complex dynamics and especially behaviors of living systems, since little modeling is required.

The algorithm of cellular automata is therefore useful for simulating real complex systems such as physical fluids, molecular dynamical systems, natural ecologies, weather, neural networks, military command and control networks, economy and many others.[10] Though von Neumann was known as a leading physicist and mathematician, he was also involved—in his capacity as an advisor to many government agencies—in many of these fields characterized by dynamical complexity. In this context he directed his attention to reductionistic explanation of certain aspects of biology. Explicit physical considerations are lacking in his work on cellular automata. He recognized and indeed emphasized a central feature of cellular automata, namely, that unlike the rigidity of the Turing machine here, in cellular automata, the distinction between the computing devices and data is blurred: construction and computation are two possible modes of activity of this algorithm. In other words, the plasticity of cellular automata—so characteristic of the living system—caught the attention of von Neumann.

The operational success of von Neumann's theory of cellular automata amounts to a proof that the following possibility is viable, namely, the successful abstraction of "the set of primitive logical interactions necessary for the evolution of the complex forms of organization essential for life."[11] Put differently, the physiological fact that the cerebral cortex consists of a very large number of local neural assemblies that are iterated throughout its extent, is successfully represented by an array of computing elements and their rules of transition. The success of the theory is in showing the possibility that such a structure of numerous simple elements is capable of complex behavior, as the brain amply exhibits, "*without the need to invoke* region-to-region variability, long range interactions, stochastic components, or mysticism."[12]

The success of this proof of possibility should not surprise us since we have known the answer from the outset: living systems reproduce themselves, and they consist of some basic discernable elements. The presupposition that these systems are biochemical machines leads to the expectation that they should be describable by some algorithm. The success is then in finding an algorithm that captures these features which the theory of cellular automata can simulate.[13]

My interest in cellular automata does not lie, however, in the success of this computing technique in simulating life phenomena. Rather, I am interested in the approach that von Neumann took which is rarely elaborated in the literature. Von Neumann commenced his paper by stating that its subject matter is

> the role of error in logics, or in the physical implementation of logics—in automata-synthesis. Error is viewed, therefore, not as an extraneous and misdirected or misdirecting accident, but as an essential part of the process under consideration—its importance in the synthesis of automata being fully comparable to that of the factor which is normally considered, the intended and correct logical structure.[14]

This instructive statement is placed up-front in the introductory section of von Neumann's essay of 1956: "Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components." It is a significant opening remark. It puts on a par the negative concept of error with positive elements of knowledge. To formulate

it differently and in practical terms, von Neumann noted that computing structures require reliability and therefore the occurrence of error should be addressed head-on and indeed at the outset of the project. The complexity of the brain, its dexterous performance and robustness, served for him as the prime example which points not only towards possible successful designs, but also to the treatment of failures.[15] The conception of failure of machines and living systems is at the center of this paper. To anticipate my findings, structurally we may benefit enormously from the analogy between living systems and cellular automata, but the nature of error, or failure, in computing systems transpires to be starkly different from failures in the living systems. Put another way, the occurrence of error points to differences rather than to similarities between living systems and cellular automata. A brief philosophical analysis of the notion of error in general will facilitate a clear understanding of these differences.

Von Neumann expressed dissatisfaction with the way error had been treated: "unsatisfactory and *ad hoc*" are his words. He thought that,

> error should be treated by thermodynamical methods, and be the subject of a thermodynamical theory, as information has been, by the work of L. Szilard and C. E. Shannon.[16]

He then admitted that his work fell short of this conception, but added that he intended his discussion of error to contribute toward this approach.

I will not pursue this physical approach to error; rather, I will direct attention to the core of the problem, to what I call the epistemic phenomenon of error. Against this background I will examine the striking difference between error of inanimate systems and that of the living. I shall conclude by suggesting that this difference may have consequences for the conception of experimentation in the biological domain.

I begin then with the epistemic phenomenon of error. According to David Hume (1711–1776) there are seven different kinds of philosophical relations: "resemblance, identity, relations of time and place, proportion in quantity or number, degrees in any quality, contrariety, and causation." Hume divides these relations into two classes. The first class comprises those relations that depend entirely on the ideas which we compare, and the second those which may be changed without any need for adjustment. To the former belong the four relations: resemblance, contrariety, degrees in equality, and proportions in quantity or number; and to the latter the remaining three relations: identity, the situations in time and place, and causation.[17] Having presented these relations and classified them in these two groups, depending on the nature of the underlying idea, Hume states:

> All kinds of reasoning consist in nothing but a *comparison*, and a discovery of those relations, either constant or inconstant, which two or more objects bear to each other.[18]

Hume italicized "comparison" and placed this discussion of philosophical relations and their underlying notion of comparison in his analysis of knowledge as part of his first book on human nature, that is, *On the Understanding*.

We need not enter into an argument with Hume about the kinds of philosophical relations and their classification; rather, at stake here is comparison—a central procedure for attaining understanding. Taking mathematics as the paramount example for his claim, Hume observes that we can carry on in algebra and arithmetic "a chain of reasoning to any degree of intricacy, and yet preserve a perfect exactness and certainty." This, he explains, is due to the fact that in this kind of reasoning we possess

> a precise standard, by which we can judge of the equality and proportion of numbers; and according as they correspond or not to that standard, we determine their relations, without any possibility of error.[19]

From this analysis we may surmise that for error to be identified as such a context must be established in which procedures of comparison could be developed and indeed applied. Such procedures logically require that a standard must be available to allow for the comparison to proceed so that an error could be determined. In other words, a fundamental characteristic of error is the *recognition* of a discrepancy in a comparative procedure. It is essential to underline "recognition" since otherwise an error would not be acknowledged as such.

What do we claim to know when we identify an error? We discern a divergence from a certain standard—a discrepancy. I have suggested elsewhere that the nature of the discrepancy and its reason may shed light on the object under study.[20] Following up this approach, my goal here is to draw consequences from the contrast between discrepancies identified in inanimate systems that are designed to simulate live organisms on the one hand, and claims of errors pertaining to living systems on the other. Von Neumann's pioneering papers on computing machines and cellular automata present a rich case for such a study.

In his seminal paper of 1946, "On the principles of large scale computing machines," von Neumann, together with Herman H. Goldstine, addressed the broad issue: "to what extent can human reasoning in the sciences be more efficiently replaced by mechanisms?"[21] Von Neumann and Goldstine observed that in highly complex fields that are based on non-linear partial differential equations such as fluid dynamics there had arisen a computational gap that generations of mathematicians had not succeeded in bridging. According to the authors, most experiments in these fields are "of a quite peculiar form": they are designed not to verify proposed theories but to replace a computation from an unquestioned theory by direct measurements. Wind tunnels, for example, are used as computing devices of the so-called analogy type to integrate the non-linear partial differential equations of fluid dynamics. The construction of large scale computing machines was partially motivated by this *impasse*. As the authors put it: "many branches of both pure and applied mathematics are in great need of computing instruments to break the present stalemate created by the failure of the purely analytical approach to non-linear problems."[22]

The machines which von Neumann and Goldstine considered belong to the digital, or counting type. These machines treat real numbers as aggregate of digits and

they are distinct from the analogical, measurement type. In analogical machines a real number is treated as a physical quantity, e.g., the intensity of an electrical current or the voltage of an electrical potential. The machines of the analogical type tend to be of a one-purpose character, specialized for a given task. This stands in contrast to the digital machines which are essentially all-purpose.[23]

One aspect of the design of the digital machines which von Neumann and his collaborator set to address right at the outset was the question of stability; the issue of error is at the center of this discussion.[24] For my argument it is important to note that von Neumann analyzes the issue of error in computing machines before he discusses "the input-output organs", "the memory organ" and "the coding of problems"—the sections that in the paper follow the discussion on error. Thus, the issue of error is presented before attention is given to the architecture and the underlying principles of these machines.

Von Neumann discerns two principal types of error. The first type pertains to malfunctions: "the device functions differently from the way in which it was designed and relied on to function."[25] Von Neumann adds that this type has its counterpart in human mistakes, both in planning and in actual human computing. Malfunctions are quite unavoidable in machine computing and they require checking. However vital this form of checking to the running of computing machines, von Neumann chooses not to be concerned with it. Rather, he focuses on the other type of error which arises even when the machine works perfectly well according to plan. Under this heading von Neumann distinguished three kinds of error.[26]

The first kind has to do with the fact that all data of empirical origin is approximate. Any uncertainty of the input, be it associated with the data or with the background theory, that is, approximate differential equations, will reflect itself as an uncertainty of the results. Based on well-known mathematical analyses, it could be shown that the size of the divergence due to this source depends on the size of the input errors and the degree of continuity of the mathematics involved. Von Neumann remarks that this kind of error pertains to any application of mathematics to nature and therefore is not peculiar to the computational approach. He therefore did not pursue it further.[27]

The second kind of error under the heading of functioning as planned, deals with the specific nature of digital computing. All continuous mathematical procedures, like integrations of differential equations, must be replaced in digital computing by elementary mathematical operations, that is, they must be approximated by a succession of the basic arithmetical operations of addition, subtraction, multiplication and division. The resulting deviation from the exact result is due therefore to truncation errors that express the discrepancy between the original continuous problem and its digital transform. However, von Neumann observes that this kind of error can be kept under control by familiar mathematical methods and are usually—so he remarks—not the main source of trouble. He therefore "passes them up, too," as he comments, at least for the time being.[28]

The third kind of error, the last one in von Neumann's enumeration, is the most crucial. It has to do with the fact that, irrespective whether the input is accurate or

approximate, "no machine, no matter how it is constructed, is really carrying out the operations of arithmetics in the rigorous mathematical sense." And he continues,

> there is no machine in which the operations that are supposed to produce the four elementary functions of arithmetic, will really all produce the correct result, i.e. the sum, difference, product or quotient which corresponds precisely to those values of the variables that were actually used.[29]

In analogical machines this is the result of representing the variables by physical quantities and the arithmetical operations or any other operation by physical processes. Such processes are invariably affected by uncontrollable uncertainties and physical fluctuations inherent in any physical instrument. Von Neumann resorts here to a term which he borrowed from communication engineering that has since then gained currency. "These operations," he writes, "are contaminated by the *noise* of the machine."[30] Analogical machines always include in their performance of an arithmetic operation an unknown quantity which represents the random noise of the mechanism of the physical processes involved. It is paramount for the success of the operation to minimize this quantity.[31]

In digital machines the reason for this kind of error is different. A digital machine must work with a definite number, which may contain many digits, but ultimately it must have a fixed, finite value. The capacity of the machine determines this value and thus its limit. Arithmetical operations conducted on a given number will normally result in more digit numbers than the machine would be able to represent with its own finite structure. A new term is therefore introduced which is known as the *round-off* error. Although this term is not a random variable and can be in fact determined in every particular instance, its determination is so complicated and its variations throughout its instances in a given calculation is so irregular that it can be considered to a high degree of approximation a random variable.[32] Von Neumann therefore refers to this third kind of error in both analogical and digital machines as *noise*, and observes that, "there is ample evidence to confirm the view, that in complicated calculations… this source of error is the critical, the primarily limiting factor."[33]

Faults in large scale computing machines

Malfunctions: Mistake	Functioning according to plan: Error
	Uncertainty in the input: theory and data
	Uncertainty due to digital representation; truncation error
	Noise & round-off error

In 1948, two years after the presentation of his research on large scale computing machines, von Neumann delivered a paper on "The General and Logical Theory of Automata."[34] It was clear to von Neumann that in spite of the fact that natural organisms are, as a rule, much more complicated and subtle than artificial automata, there is a fruitful reciprocal relation between these distinct systems. While some regularity in living organisms could be instructive in the thinking and planning of automata, the experience with automata could be to some extent projected on the

interpretation of natural organisms.[35] The latter point is at the center of my interest. Although von Neumann acknowledged the different conception of error in the two systems, he thought they are in some sense related. I call this claim into question; indeed, I attempt to refute it.

Since the living system is immensely complex, von Neumann suggests a reasonable and indeed by now a common approach of two moves based on the following presupposition. The organism may be viewed as made up of parts which are to a certain extent independent, elementary units. The first move is then to identify the structure and function of such elementary units individually. The second move consists in seeking an "understanding how these elements are organized into a whole, and how the functioning of the whole is expressed in terms of these elements."[36]

In the first move von Neumann retains the traditional distinction of structure and function as the underlying heuristics.[37] He disposes of this first step by applying what he calls the *Axiomatic Procedure*:

> Axiomatizing the behavior of the elements means this: We assume that the elements have certain well-defined, outside, functional characteristics; that is, they are to be treated as "black boxes." They are viewed as automatisms, the inner structure of which need not be disclosed, but which are assumed to react to certain unambiguously defined stimuli, by certain unambiguously defined responses.[38]

This procedure is a powerful heuristic device that underlies physiological studies. Not surprisingly, von Neumann chooses to concentrate on the second move, where issues of formalism—logical as well as mathematical—are at stake.

As in the other papers, here too the issues of precision and reliability receive attention right at the outset. Von Neumann remarks that normally one would expect of a machine that "the larger the number of operations required to produce a result, the smaller will be the significant contribution of every individual operation."[39] Thus the occurrence of error in automata will matter only to the extent of the fraction of the total number of steps which are required for the completion of the task. This however does not hold for computing machines. In computing machines any step—whatever the number of operations—is as important as the whole result. To put it bluntly in von Neumann's own words: "any error can vitiate the result in its entirety."[40] Computing machines have to perform billions of steps in a short time and no error is permitted in a considerable part of the procedure. In fact, the demand is that no error should occur anywhere in the entire procedure. In this sense, a computing machine is an exceptional artificial automaton, but it is this feature, according to von Neumann, that makes this automaton most suitable for a comparison to the functioning of a natural organism.

By comparing a cellular automaton with a living organism, von Neumann identifies processes of digital and analogical nature. While the nerve impulse seems to function in a binary way and thus well suited to digital representation, other functions of the living system are mediated in a continuous fashion in what von Neumann calls "humoral media".[41] Specifically, he discerns both processes in the central nervous system, that is, digital as well as analogical. The organism exhibits composite functional sequences which "go through a variety of steps from the

original stimulus to the ultimate effect—some of the steps being neural, that is, digital, and others humoral, that is, analogy." Furthermore,

> These digital and analogy portions in such a chain may alternately multiply. In certain cases of this type, the chain can actually feed back into itself, that is, its ultimate output may again stimulate its original input.[42]

The complexity of the living organism is due partly to this intricate combination of different kinds of process, in contrast to computing machines which in the present state of the art are purely digital. And von Neumann remarks that in drawing an analogy between the living organism and large scale computing machines he attends only to the digital aspect of the living system—an oversimplification which is however heuristically productive, and especially so when the device—be it a neuron or a vacuum tube (von Neumann, it should be noted, wrote this paper before the invention of the transistor)—is considered a "black box" with a schematic description.[43]

The parallel function of the two key elements, that is, the nerve cell and the vacuum tube, has thus been drawn. It reflects the correspondence between the building blocks of the nervous system and those of the automata with computing capability. Von Neumann turns now to what he considers a crucial drawback, in fact the stumbling block in the development of automata, namely, the rigidity of the formalism: the available mathematical-logical theories had been too rigid to be conducive to the operational requirements of automata. In particular, the length of "chains of reasoning" had to be considered as well as failures that are part and parcel of a working machine. Thus,

> The operations of logic (syllogisms, conjunctions, disjunctions, negations, etc., that is, in the terminology that is customary for automata, various forms of gating, coincidence, anti-coincidence, blocking, etc., actions) will all have to be treated by procedures which allow exceptions (malfunctions) with low but non-zero probabilities.[44]

Von Neumann imports his analysis of error from the large scale computing machines to his studies of automata. He expected this theory to be less combinatorial and more analytical, akin to the character of thermodynamics as Boltzmann treated it. Von Neumann discerns here a theoretical limitation which is of much importance to the point I am seeking to make. At stake is error checking procedure.

We have seen von Neumann analyzing possible kinds of error in large scale computing machines. For him errors and their sources "need only be foreseen generically, that is, by some decisive traits, and not specifically . . . in complete detail."[45] However, a malfunction in artificial automata must be detected, as soon as it occurs, otherwise these machines would be useless. Effort should be made to identify the error, by say mathematical means or automated checks, to isolate the faulty component that caused the error, and put it then aright or replace it altogether. This is why designers compartmentalize machines. As Walter Elsasser (1904–1991) explains:

> If a system is sufficiently compartmentalized so that errors are prevented from spreading, their consequences may be limited to one compartment for a very long time. If this is not done the consequences of the error tend to spread over the whole system owing to the extensive interconnection of various processes by mutual feedback. Designers of electronic

> computers therefore have a pronounced tendency to compartmentalize their systems as much as possible, partly in order to prevent the spreading of errors and partly to be able to track them down more readily in case they occur.[46]

Notice that the diagnosis is effected from without and the faulty component is replaced by agents external to the system. But over and above the corrective measures that may be taken, the error itself may be identified in the first place only against a known standard or criterion. It is this identification which subsequently allows for insulation and rectification. Therefore, as von Neumann puts it,

> we are trying to arrange the automata in such a manner that errors will become as conspicuous as possible, and intervention and correction follow immediately.[47]

The quick intervention is important to prevent further errors setting in. It is a common experience that machine which has begun to malfunction rarely will restore itself, and more probably go from bad to worse.

This is not the case of the living system; in von Neumann's words, "the organism obviously has a way to detect… [malfunctions] and render them harmless."[48] Note that von Neumann regards this observation as indisputable: he says "obviously". An organism, for example, the living cell, is presumed to have a way of detecting on its own, that is, from within, malfunctions and treat them accordingly. Therefore this system must

> contain the necessary arrangements to diagnose errors as they occur, to readjust the organism so as to minimize the effects of the errors, and finally to correct or to block permanently the faulty components.[49]

And in the case of the living system there is little evidence of compartmentalization. Thus, according to von Neumann, the entire organism appears to make the malfunctions as inconsequential as possible, and to apply corrective measures. In other words, "organisms are constructed to make errors as inconspicuous, as harmless, as possible."[50] In sum, while the engineer seeks to make the error as conspicuous and distinct as possible and react swiftly with external means to eliminate it before further errors set in, the alien designer of the living system has equipped the system with an internal faculty that can diagnose a malfunction and render it as inconspicuous as possible in a relatively long time—so von Neumann's argument runs.

I have underlined the success of cellular automata in obtaining complexity that evolves from rudimentary, elementary machinery in parallel to that of the living system. But when it comes to disturbances and interferences there appear to be major qualitative differences—the flexibility of cellular automata is not sufficient for capturing the plasticity of the organism in handling faults. To use a metaphoric language, automata live an extremely flat live. At stake are the very elements of the cellular automata: the number of states variable in a given cell, the number of cell neighbors and the sensitivity of the transition rule to the environment. The difficulties in capturing the versatility of the living system may be characterized respectively as

robustness to perturbation, that is, stability, then variability, and finally sensitivity (or rather insensitivity) to changes in the transition rule.[51]

Consider robustness:

> Alteration of the state of a single unit of the von Neumann machine typically leads to catastrophic failure; [by contrast] malfunction of a single neuron or neural assembly should have no measurable effect.[52]

The successful operation of the von Neumann construction is due to choosing a discrete substrate in space, time, and state variable. This success is obtained however at a very high price since the automaton is much more vulnerable to disturbances than, say, differential equations whose continuous substrate is conducive to the treatment of perturbations. How many states are required in order to obtain robustness in cellular automata? It may well be that increasing the number of states would not after all result in robustness.

Then there is the issue of variability. It is the variability at the level of the individual neuron which the von Nuemann machine cannot accommodate, for it would fail catastrophically were the interacting neighboring cells of the automaton be of a too varied nature. Again, the question of number arises: how many neighboring cells it would take to achieve variability, a feature which is natural, so to speak, in the living system.

Finally, it may be at times beneficiary to the living system to be insensitive to the environmental changes; by comparison, it is not at all clear how a cellular automaton can ignore changes in the transition rule. These three elements: stability, variability and sensitivity may constitute terminal problems for the designer of cellular automata in the attempt to depict fundamental features of the living system.

Such difficulties render the comparison of computing inanimate machines and living systems problematic; but how does error fare in this comparison? I return to the distinction which von Neumann draws between modes of checking and rectifying errors in artificial automata and organisms. Recall that the engineer seeks to make the error as conspicuous as possible in the shortest time possible, quite the opposite to the common practice, as it were, of the living system. Now, how is error made conspicuous, or for that matter, inconspicuous? Von Neumann's analysis is based on the presupposition that knowledge of what the machine is supposed to do and how it is designed to accomplish it is given. As I have argued, a comparison procedure makes the discrepancy apparent. Thus, it is this *given* knowledge of goals and means that makes the identification of error possible. This procedure of comparison should work also for the living system. Von Neumann characterizes the relevant background knowledge—the "operating conditions"—in the living system as "normal"; that is, the operating conditions

> represent the functionally normal state of affairs within the large organism.…Thus the important fact is not whether an organ has necessarily and under all conditions the all-or-none character—this is probably never the case—but rather whether in its proper context it functions primarily, and appears to be intended to function primarily, as an all-or-none organ.

And von Neumann adds candidly,

> I realize that this definition brings in rather undesirable criteria of "propriety" of context, of "appearance" and "intention." I do not see, however, how we can avoid using them, and how we can forgo counting on the employment of common sense in their application.[53]

Indeed, it is impossible to see how such terms can be avoided—this is the kern of my claim. Von Neumann's revealing remark harbors important consequences, but he does not draw them. Knowledge of these "operating conditions" is in effect the standard against which error may be discerned and if criteria such as "propriety", "appearance", and "intention" are undesirable then on what grounds could a fault in the living system be identified at all as such, namely, a fault?

The problem is compounded by the fact that the living system lacks accuracy. Karl Lashley (1890–1958)—the American psychologist who brought into focus the controversy between localization and holistic emphasis of brain function—posed this problem to von Neumann in the discussion on the theory of automata.

> In the computing machines, the one thing we demand is precision; on the other hand, when we study the organism, one thing which we never find is accuracy or precision. In any organic reaction there is a normal, or nearly normal, distribution of errors around a mean. The mechanisms of reaction are statistical in character and their accuracy is only that of a probability distribution in the activity of enormous numbers of elements. In this respect the organism resembles the analogical rather than the digital machine. The invention of symbols and the use of memorized number series convert the organism into a digital machine, but the increase in accuracy is acquired at the sacrifice of speed. One can estimate the number of books on a shelf at a glance, with some error. To count them requires much greater time. As a digital machine the organism is inefficient. That is why you build computing machines.[54]

This statistical approach is usually associated with the belief in the existence of overall laws of large scale nerve stimulation and composite action, but in living systems there are often single elements, a neuron, that may control a whole process.[55] How could we then determine the governing law of this single cell? What will be considered "appropriate" of its behavior or, for that matter, what is its "intention"? Put concisely, we have to determine the "value" system of this neuron in order to identify an error in its function.

This train of reasoning underpins Warren S. McCulloch's graphical response to von Neumann's theory of automata. McCulloch, of the well known McCulloch-Pitts model of the neuron (1943), is recorded rejoining:

> I confess that there is nothing I envy Dr. von Neumann more than the fact that the machines with which he has to cope are those for which he has, from the beginning, a blueprint of what the machine is supposed to do and how it is supposed to do it. Unfortunately for us in the biological sciences—or, at least, in psychiatry—we are presented with an alien, or enemy's, machine. We do not know exactly what the machine is supposed to do and certainly we have no blueprint of it. In attacking our problems, we only know, in psychiatry, that the machine is producing wrong answers. We know that, because of the damage by the machine to the machine itself and by its running amuck in the world. However, what sort of difficulty exists in that machine is no easy matter to determine.[56]

Note that the standard of comparison to which McCulloch refers is coherence, that is, what appears to McCulloch and his co-workers in psychiatry as self-preservation and efficient adaptability to the world—be it either the physical or the social world, or indeed both realms. But surely this is just one interpretation, one possible mode of evaluating the objective that this system, namely, the human being, is supposed to accomplish.

The claim that the living systems lacks a known standard, which in turn undermines—so I have argued—the possibility of determining error in this context, may be formulated for clarity sake by using the notion of "teacher", an agent knowledgeable of the system so that it can supervise its performance. An artificial automaton has to have a teacher, the designer who oversees the functioning of the machine. The teacher, by definition, possesses knowledge of the standard that the automaton has to maintain. In principle, the teacher could be decoded and the instructions be taught automatically. The crucial point, however, is that the teaching comes from without, externally to the system. Note that the teacher is not capable of doing what the machine does, it only oversees the functioning of the machine. Indeed, as Lashley pointed out, this is why we build such machines. Thus, we may ask, how does the teacher know that the end result of millions of calculations is correct? The teacher can supervise the procedure but cannot check the result itself. Von Neumann's solution is degeneracy, namely, apply another machine; so he calls this procedure, "multiplexing".[57]

> Connect . . . three . . . machines in such a manner that they always compare their results after every single operation, and then proceed as follows. (*a*) If all three have the same result, they continue unchecked. (*b*) If any two agree with each other, but not with the third, then all three continue with the value agreed on by the majority. (*c*) If no two agree with each other, then all three stop.[58]

This system that comprises three machines will obtain correct results unless two of the three machines err simultaneously, for which the probability, according to von Neumann's calculation, is one in 33 million. Notice how von Neumann proceeds: he applies a comparative procedure. Once again the key is comparison and in this case each result is compared to the other in an attempt to achieve consensus, albeit machine produced consensus.[59]

Thus far machines and automata and their required instructor; but does the living system has a "teacher"? If the answer is negative, or if we do not have access to it, then in such systems the determination of malfunctions, and generally of errors—a process to which von Neumann refers as "obvious"—would be logically impossible. In this sense the foregoing discussion of error in living systems is in fact unfounded.[60]

Granted, living systems possess organs that have identifiable functions whose ultimate goals and standards may be determined as "normal". This brings us, however, directly to the function-structure problematic distinction.[61] But note that these organs are mostly peripheral, located as they are at the interface between the living system and its environment. Consider, however, the cell itself, or its constitutive elements—the fundamental building blocks of life. The determination of function

ceases then to be clear and consequently knowledge of the standard, that is, the norm, may be missing altogether. I claim that in these cases it is not clear at all what does it mean to impute error to the system, and indeed to call a certain building block faulty.

This philosophical worry does not disturb practitioners from further inquiring into biology in the spirit that von Neumann inaugurated half a century ago. A good example is the work of John J. Hopfield who in the 1970s developed an algorithmic scheme which he called "kinetic proofreading", and later on in the 1980s demonstrated how physical systems could pick up features of neural networks and simulate the function of memory purely by computation. Hopfield speaks of "reading" the genetic code with few mistakes. He considers the understanding of how small error rates are achieved in the living systems as one of the fundamental general problems of biosynthesis. Admittedly, he writes that he examines the issue "from a phenomenological point of view." Still, his proofreading procedure which is based on energy levels presupposes the concept of error as a primitive that needs no explanation, certainly not a technical one, and one remains perplexed with respect to the definition of this basic concept, let alone imputing it to organism.[62]

The two related points, namely, lack of a teacher (or ignorance of it) and processes that are in principle not accurate, constitute a categorical difference between large scale computing machines and artificial automata on the one hand and living systems on the other. To be sure, the comparison between the two systems is productive as von Neumann amply showed. However, the comparison may be misleading when it comes to the conception of error. In fact, given the argument I have presented concerning the epistemic phenomenon of error, the attribution of error to animate systems may be in itself erroneous.

The question now presents itself whether the application of the experimental technique in biology—as we have come to know it, say, in biophysical experiments—should take stock of this consequence. So far it appears that this has not been the case and practitioners such as Hopfield have no hesitation to attribute error, e.g., misreading, to the living systems, and indeed to its constitutive elements. In conclusion, I suggest drawing the consequence so that to avoid the undesirable criteria of "propriety" of context, of "appearance" and "intention", as indeed von Neumann described the problem. A new mode of experimenting is called for that acknowledges this difficulty, but this I leave for another story.

Acknowledgment I thank Jutta Schickore for incisive comments and Andrea Loettgers for drawing my attention to the work of J. J. Hopfield.

Notes

1. Hopfield 1974, 4135.
2. Thatcher 1970. Cf. Kendall and Duff 1984, 1. For historical background, see Abraham 2000, Ch. III: "From Neural Networks to Self-Reproduction: John von Neumann and Automata Theory."

3. McIntosh 1990, 105.
4. Victor 1990, 205.
5. Toffoli and Margolus 1990, 230.
6. Victor 1990, 205.
7. Kendall and Duff 1984, 11–12; Ilachinski 2001, 7. Cf. Wolfram 1986, 1. For historical background, see Toffoli and Margolus 1990, 231–232.
8. Culick et al. 1990, 357.
9. For further discussion, see Sutner 1990, 389–390.
10. Kendall and Duff 1984, 11–12.
11. Ilachinski 2001, 3.
12. Victor 1990, 205 (emphasis in the original).
13. Ilachinski 2001, 571.
14. Von Neumann 1956/1963, 329.
15. Kendall and Duff 1984, ix.
16. Von Neumann 1956/1963, 329.
17. Hume 1739–1740/1978, 69–73; Bk.1, pt.3, 1, 2.
18. Ibid., 73 (emphasis in the original).
19. Ibid., 71.
20. Hon 1998, 466.
21. Goldstine and von Neumann 1946/1963, 2; Bródy and Vámos 1995, 495.
22. Ibid., 4; ibid., 497.
23. Goldstine and von Neumann 1946/1963, 8–9.
24. Ibid., 13–14.
25. Ibid., 15.
26. Ibid.; Bródy and Vámos 1995, 508.
27. Goldstine and von Neumann 1946/1963, 16; Bródy and Vámos 1995, 508–509.
28. Ibid., 509.
29. Ibid.
30. Goldstine and von Neumann 1946/1963, 16 (emphasis in the original); Bródy and Vámos 1995, 509.
31. Bródy and Vámos 1995, 531; cf. von Neumann 1951/1963, 293–294.
32. Ibid., 533; cf. ibid., 294–295.
33. Goldstine and von Neumann 1946/1963, 17; Bródy and Vámos 1995, 510.
34. Von Neumann 1951/1963; Bródy and Vámos 1995, 526–566.
35. Ibid., 288–289; ibid., 526–527.
36. Ibid., 289; ibid., 527.
37. I have argued elsewhere (Hon 2000) that this is the source of much misconception of the living system.
38. Von Neumann 1951/1963, 289.
39. Von Neumann 1951/1963, 292; Bródy and Vámos 1995, 530.
40. Ibid. Von Neumann qualified this remark, adding that the claim is not absolutely true; probably only 30 per cent of all steps made are of this nature.
41. Von Neumann 1951/1963, 296.
42. Von Neumann 1951/1963, 296; Bródy and Vámos 1995, 534. Cf. Von Neumann 1956/1963, 368–369, 372.
43. Ibid., 296–298; ibid., 534–536. See also ibid., 368–369, 372, 375–376.
44. Ibid., 304; Bródy and Vámos 1995, 542.
45. Von Neumann 1951/1963, 324; Bródy and Vámos 1995, 562.
46. Elsasser 1966, 40.
47. Von Neumann 1951/1963, 305–306; cf. Bródy and Vámos 1995, 543–44.
48. Von Neumann 1951/1963, 305; Bródy and Vámos 1995, 543.
49. Ibid.
50. Ibid., 306; Bródy and Vámos 1995, 544.
51. Victor 1990.

52. Ibid., 206.
53. Von Neumann 1951/1963, 298.
54. See von Neumann 1951/1963, 324; Bródy and Vámos 1995, 565.
55. See von Neumann 1956/1963, 369.
56. See Von Neumann 1951/1963, 319; Bródy and Vámos 1995, 557.
57. Von Neumann 1956/1963, 347 and 353–368 (§§ 9, 10).
58. Von Neumann 1951/1963, 322. For a detailed technical analysis see von Neumann 1956/1963, 347–353.
59. Burks, Goldstine and von Neumann 1946/1963, 68–70. Note that this procedure does not allow for diagnosis.
60. Canguilhem's study of the normal and the pathological focuses on this difficulty from a different perspective (1978/1991).
61. See Hon 2000.
62. Hopfield 1974; 1980; 1982.

References

Abraham, T. H. (2000). "Microscopic cybernetics: mathematical logic, automata theory, and the formalization of biological phenomena, 1936–1970." Ph.D. thesis, Institute for the History and Philosophy of Science and Technology, Toronto, Canada: University of Toronto.

Bródy, F. and T. Vámos eds. (1995). *The Neumann Compendium*. Singapore: World Scientific.

Burks, A. W., H. H. Goldstine and John von Neumann (1946/1963). "Preliminary Discussion of the Logical Design of an Electronic Computing Instrument." Report prepared for U.S. Army Ordnance Department, 1946. Reprinted in von Neumann 1963, 34–79.

Canguilhem, G. (1978/1991). *The Normal and the Pathological*. New York: Zones.

Culick, K. II, L. P. Hurd and S. Yu (1990). "Computation Theoretic Aspects of Cellular Automata." In Gutowitz 1990, 357–378.

Elsasser, W. (1966). *Atom and Organism: A New Approach to Theoretical Biology*. New Jersey: Princeton University Press.

Goldstine, H. H. and J. von Neumann (1946/1963). "On the Principles of Large Scale Computing Machines." Unpublished manuscript, printed in von Neumann 1963, 1–32. See also Bródy and Vámos 1995, 494–525.

Gutowitz, H. (1990). *Cellular Automata: Theory and Experiment*. North Holland. *Physica D*, **45**: 1–3.

Hon, G. (1998). "Exploiting Error." *Studies in History and Philosophy of Science* **29**: 465–479.

Hon, G. (2000). "The Limits of Experimental Method: Experimenting on an Entangled System—The Case of Biophysics." In M. Carrier, G. J. Massey and L. Reutsche eds., *Science at Century's End: Philosophical Questions on the Progress and Limits of Science*. Pittsburgh: University of Pittsburgh Press, pp. 284–307.

Hopfield, J. J. (1974). "Kinetic Proofreading: A new Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specifity." *Proceedings of the National Academy of Science* **71**: 4135–4139.

Hopfield, J. J. (1980). "The energy relay: A proofreading scheme based on dynamic cooperativity and lacking all characteristic symptoms of kinetic proofreading in DNA replication and protein sythesis." *Proceedings of the National Academy of Science* **77**: 5248–5252.

Hopfield, J. J. (1982). "Neural networks and physical systems with emergent collective computational abilities." *Proceedings of the National Academy of Science* **79**: 2554–2558.

Hume, D. (1739–1740/1978). *A Treatise of Human Nature*. L. A. Selby-Bigge ed., 2nd edition. P. H. Nidditch ed., Oxford: Clarendon Press.

Ilachinski, A. (2001). *Cellular Automata: A Discrete Universe*. Singapore: World Scientific.

Kendall, P., Jr. and M. J. B. Duff (1984). *Modern Cellular Automata: Theory and Applications*. New York and London: Plenum Press.

McIntosh, H. V. (1990). “Wolfram’s Class IV Automata and a Good Life.” In Gutowitz 1990, 105–121.

Sutner, K. (1990). “Classifying Circular Cellular Automata.” In Gutowitz 1990, 386–395.

Thatcher, J. W. (1970). “Universality in the von Neumann cellular model.” In A. W. Burks ed., *Essays on Cellular Automata*. Urbana: University of Illinois Press.

Toffoli, T. and N. H. Margolus (1990). “Invertible Cellular Automata: A Review.” In Gutowitz 1990, 229–253.

Victor, J. D. (1990). “What can Automaton Theory Tell us about the Brain.” In Gutowitz 1990, 205–207.

Von Neumann, J. (1951/1963). “The General and Logical Theory of Automata.” In *Cerebral Mechanisms in Behaviour*. The Hixon Symposium. New York: Wiley, 1951. Reprinted in von Neumann 1963, 288–328.

Von Neumann, J. (1956/1963). “Probabilistic Logics and the Synthesis of Reliable Organisms From Unreliable Components.” In C. E. Shannon and J. McCarthy eds., *Automata Studies*. Annals of Mathematics Studies, No. 34. Princeton, N. J.: Princeton University Press, pp. 43–98. Reprinted in von Neumann 1963, 329–378.

Von Neumann, J. (1963). *Collected Works*, A. H. Taub ed., vol. 5: *Design of Computers, Theory of Automata and Numerical Analysis*. New York: Macmillan.

Wolfram, S. (1986). *Theory and Applications of Cellular Automata*. Singapore: World Scientific.

Part III
Concepts and Dead Ends

Experimental Reorientations

Hans-Jörg Rheinberger

Thoughts to Begin With

In his preface to Claude Bernard's *Introduction à l'étude de la médecine expérimentale*, François Dagognet characterizes the epistemological attitude of the French nineteenth-century physiologist as one of "incessant rectification," and he adds in the spirit of Bernard:

> The most insignificant fact eventually destroys the most solid hypothesis. Theory for its part can play the role of a springboard, but also that of an obstacle. One discovers less with ideas than against them[…]To be sure, one has to put questions to the life process, but everything depends on being attentive to the answers that are delivered on the margins or even outside the expected discourse.[1]

In a similar vein, Gaston Bachelard has claimed that "scientific thinking is essentially a rectification of knowledge,"[2] meaning that there is scientific spirit only as long as there is science in the making. Now, everything depends on what one means by the term rectification. Where is "rectification" located in the coordinate system created by "verification" and "falsification"? Both verificationist and falsificationist logics of research were developed in great detail in the second half of the nineteenth and the first half of the twentieth centuries, and they were generally associated with inductivist and deductivist perspectives, respectively.

With my remarks, I do not want to add another layer to that discussion. Nor do I want to critically assess the merits and shortcomings of these different positions. What appears to me to be more fruitful is to explore what it could mean and how far it might take us to approach the dynamics of research from a *rectificationist* perspective. But then, what we need is to talk about what the possible meanings of the notion of rectification are. At first glance, what it means appears to be straightforward: "Rectification" is associated with the idea of the

H.-J. Rheinberger (✉)
Max Planck Institute for the History of Science, Berlin

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-88393-3_6,

correction of an error. It is about straightening out, getting it right. Bachelard, in his *Le nouvel esprit scientifique* claimed that "scientifically, one thinks of truth as the historical rectification of one long error."[3] The context of the quote cannot be fully reconstructed here; but it suggests that scientific truth is not to be thought of as a perennial achievement, but rather that it must be thought of as an ongoing *process*. Bachelard himself saw the process of knowledge acquisition as inexhaustible. Truth itself then, in science, is a historical entity. It is historically rather than logically constituted. But again, how shall we characterize this process more precisely? What does it mean to see the development of the sciences as the rectification of one long error? What, then, about our present knowledge? Is it possible to read the notion of rectification in a non-normative and deflationary manner?

In this paper, I would like to discuss a broader context of *reorientation* within which the notion of rectification gains meaning. In particular, I would like to talk about the reorienting forces of experiment. Such forces of reorientation appear to be inherent in productive experimental arrangements. It also appears that the elimination of errors in the strict sense of the word, that is, of correcting assumptions that simply and plainly turn out to be—quantitatively or qualitatively—false, is neither the rule, nor is it at the center of the experimental process. Nor is the plain and simple corroboration of a hypothesis at the center. We could perhaps say that these are boundary cases that, if taken as central, fail to give an adequate picture of the dynamics of experimental reorientation. Experimenters are usually working in a landscape where almost nothing is either black or white, and almost everything consists of shades of gray. It happens much more frequently that findings regarded as prominent at a particular time recede into the background at another time; that questions regarded as important remain unsolved and eventually become marginalized, if not forgotten altogether, whereas others move center stage; that technical impasses force one's interest in a different direction. This includes the possibility that certain experimental avenues chosen may lead to dead ends and become abandoned either temporarily or forever. It is seldom acknowledged but certainly commonplace among working scientists that the bulk of their experimental efforts do not lead anywhere. They do not lead to promising findings, but they also do not lead to the clear-cut falsification of sharply delineated assumptions. They simply do not amount to anything that could be worth reporting. And yet these efforts enrich the experimental experience and therefore are integral to the experimental approach. The whole enterprise would not function without these efforts that end in suspense. No exploration of experimental spaces would be possible without them.

Bachelard developed the concept of the "epistemological obstacle" for what he conceived as an inherent slowness of the thinking mind in its engagement with objects of knowledge.[4] Clarity in the realm of the empirical sciences was, for him, of a necessary belatedness. In moving a step forward we do not get a clear picture of the future, we only come to see a little bit more clearly what is behind us. The emergence of knowledge in the realm of the empirical sciences remains in the mental space of an inevitable confusion, a space of groping and

errantry, as a condition for the precipitation of epistemological acts. In contrast to many of his contemporary colleagues in the philosophy of science, Bachelard did not exclude this space of non-comprehension from the grip of epistemology. Rather, he declared it to belong at its center. In terms of the epistemological objects proper, he talked about their resistance against the effort of knowing as the essence of any empirically grounded, experimental research process. The positivity of new knowledge announces itself in the negative form of a resistance. Resistance means that there is something that does not fit, something that intrigues the mind. In terms of the experimental arrangement, findings do not fall into place. In the experimenter's mind, this translates into disquietude, agitation, trouble, disturbance, and uneasiness. Understanding the dynamics of research means understanding the structure of experimental resistance. Resistance induces reorientation. With an analysis of modes of resistance in experimental setups, we may be able to make a step into this landscape of gray shades between yes and no, between the positive and the negative, between plain truth and plain error. A history of instances of resistance will tell us, I hope, some useful lessons about experiments going right or wrong.

In this paper, I will present two historical examples of exploratory reorientation[5]. The first is taken from nineteenth-century physiology and deals with the discovery by the aforementioned Claude Bernard that animals not only break down but also synthesize sugar in their bodies. The example covers a short decade of work between 1840 and 1848. The second example leads us into molecular biology of the twentieth century. It traces the development of *in vitro* techniques for analyzing subcellular particles and covers a long half-century between 1910 and 1965. The examples are taken from different time periods. They cover rather different fields of work, although both of them belong to the life sciences; and the experiments considered are embedded in widely different research technologies. Yet there is a recurrent pattern that surfaces again and again, whether we choose a microscopic perspective and follow the weekly motions of a particular laboratory researcher, or whether we choose a macroscopic perspective and consider the development of a whole research field involving two generations of scientists and several different laboratories. We could call that recurrent pattern, with Ludwik Fleck, the founder of a history of experimental cultures and one of the most important sociologists of science of the past century, the "Columbus effect": One looks for India, and what one finds is America.[6]

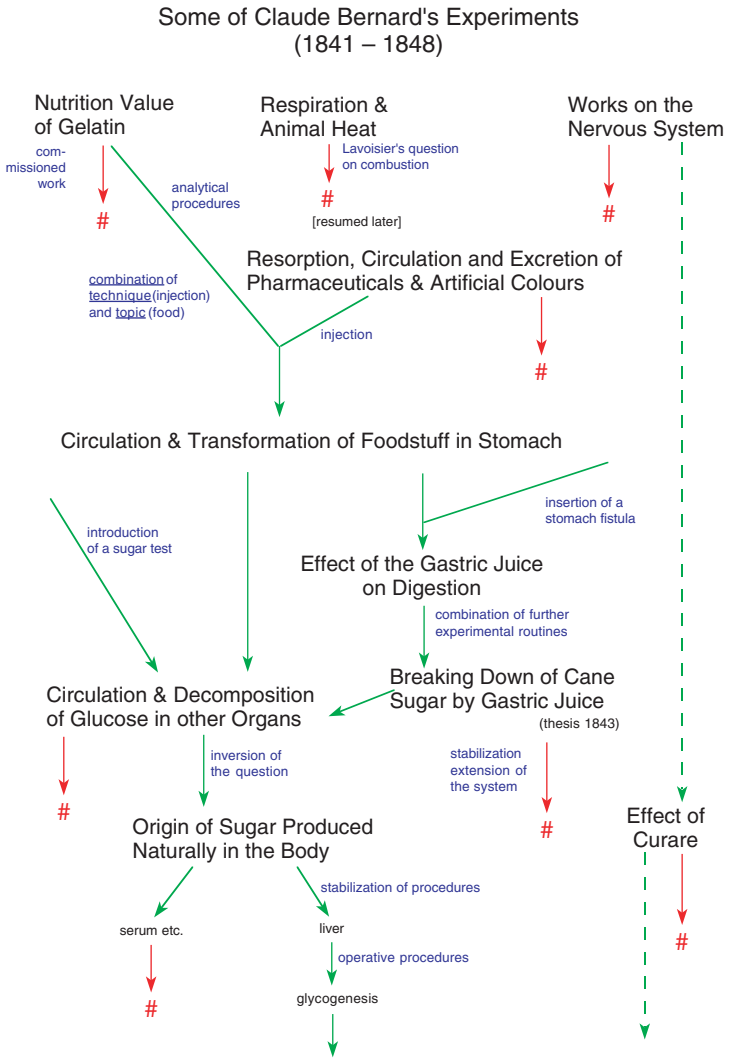
Bernard flattered himself that he was said to "find things that [he] never looked for."[7] As an experienced experimenter, he did not take this as an insult, but rather as a compliment. He saw in acts of reorientation the foundation and starting point of all research. Research is about new knowledge, but new knowledge can arrive only as an event. The genius of the experimenter consists of being aware of such events. As *new* knowledge, it can—by definition—not be anticipated. Bachelard argued in a very similar manner when, in his early *Essai sur la connaissance approché*, he claimed: "History of science teaches that every big step toward an ultimate reality has shown that this reality always announced itself from a completely unexpected direction."[8]

Claude Bernard and the Demonstration of Glycogenesis in the Liver[9]

Let me now come to my first example. It traces the path by which the French physiologist Bernard arrived at an observation that became fundamental to nineteenth-century animal physiology and that cemented his fame as a gifted experimenter. It was the realization that the liver of higher animals produces and stores sugar. Mirko Grmek and Frederic Holmes, whom I follow here as far as the outline of the story is concerned, have given us a detailed investigation of Bernard's unpublished early laboratory notes.[10] They have convincingly shown that the published papers of the French physiologist from the 1840s alone do not allow us to recover the groping laboratory and learning process that Bernard underwent, first, as a preparator and assistant of François Magendie at the Collège de France in Paris, and then, as an independent young researcher working, among other places, in the laboratory of the chemist Théophile Pelouze. These were the years during which he developed a completely new method of investigating nutrition and digestion processes within a living animal's body.

To institute a strictly experimental medicine had been the outspoken program of Magendie, who had studied with Xavier Bichat and worked as professor of physiology at the Collège de France since 1835. Bernard took up the experimental orientation of Magendie, but he gave nineteenth-century physiology a particular twist that resulted in a threefold demarcation from tradition. First, he detached it from the tendencies of a doctrinaire physicalism. Second, he turned the attention of the physiologist to the processes as they went on within the organism, and with that, came to stand in opposition to the chemistry of his time, which, as practiced by Justus von Liebig and Jean-Baptiste Dumas, understood itself as the leading physiological science and was mainly interested in the analysis of metabolic inputs and outputs that could be measured outside the body. And third, he emancipated physiology from comparative anatomy and morphology, the leading biological sciences of his time. It is in this experimental physiological program that Bernard's discoveries inscribed themselves, and it is through these discoveries that the program acquired its shape. But this is not to say that he had, from the very beginning, a definite idea of either the change of this broader framework or of the particular findings that would result from his efforts. Rather, what we see is a process of groping, of bouncing against obstacles, of taking detours, and of launching new starts. If Bernard later said that in order to become a productive physiologist, "one must have groped around a thousand times,"[11] this is not mere hyperbole, it is the lesson of his own life as an experimenter. Figure 1 shows a rough sketch of Bernard's experimental activities in the seven years between 1841 and 1848.

After his medical exam, Bernard joined the staff of the Collège de France as an auxiliary assistant in 1840, and from 1841 onward, served as preparator for Magendie. In addition to this job, he came to be involved in experiments devised and conducted by the so-called Gelatin Commission. This commission was an investigative body that had been installed by the Academy of Sciences at the instigation of



the government. Its task was to determine the nutritive value of gelatin. Gelatin had been used as a cheap substitute for meat in France's hospitals and public asylums for some time, and it was a highly contested measure. Bernard's early physiological investigations on nutrition—essentially feeding tests with animals—found their starting point in this accidental assignment, and they gained a visibly individual profile only in the course of the coming years. In addition, and together with Magendie, Bernard was busy with work on respiration and animal heat. What he had in mind was testing the chemical theories of animal respiration from Lavoisier to Dumas, which were based on an analogy to combustion, in direct experiments on the living

animal. Although these experiments had an explicit aim and were conceived based on a successful experimental tradition, they did not lead to any new results worth mentioning.

In September 1842, Bernard started to work on a new line of experiments that revolved around the uptake and the excretion of drugs in experimental animals. For this purpose, Bernard made use of the injection technique developed by Magendie. When an animal suddenly died after he had administered two different, individually harmless substances at different points of the body, he realized that he could use the technique to follow the path of different substances and identify the place where they reacted. Thus, as Holmes phrased it, a technical accident turned into a "primary object of the investigation itself."[12] Bernard then introduced different test substances, for instance, two substances whose combination resulted in a color reaction, separately from each other and at different places, into the bloodstream of his animals. His quite general and loosely defined aim was to find out where the substances came together in the animal body, and in this way, to learn something about the specific distribution and reaction specificity of drugs in the organism. Other than pointing to the stomach as a privileged organ of chemical reaction, this series of experiments did not lead him to any significantly new knowledge in pharmaceutical physiology. However, it set Bernard on a track that moved his work, if only incidentally at the beginning, in a new direction. For the same injection technique could also be used to follow the fate and transformation of *nutrients* after their application into and subsequent distribution through the body. This change of perspective did not occur abruptly. It was rather the eventual result of a combinatorial laboratory practice in which a current and well-mastered experimental procedure was applied to a current, but not-so-well-mastered problem in a different context. It was the nutrition problem that had occupied Bernard a couple of years before, when he was working for the Gelatin Commission. But for the time being, Bernard arrived at no clear-cut results on this front either.

Inspired by an actual chirurgical invention of Nicolas Blondlot, Bernard introduced an additional surgical technique into his arsenal: the insertion of a stomach fistula. It allowed him to concentrate on the action of gastric juice. With the help of the fistula, he could retrieve gastric juice and also test its influence on foodstuffs, such as cane sugar, outside the stomach. Again, the change of technique was not abrupt and resulted from collaboration with Charles-Louis Barreswil who, at that time, had joined Magendie as a preparator. From then on, however, Bernard relied less on questions that were picked up casually, and more on those that, in one way or another, had to do with one of the several experimental series carried out earlier. One could perhaps say, in picking up a term that Bernard later coined for the inner environment of the organism, that he slowly built up something like an "internal milieu" for his experimental regime.

In his gastric investigations, Bernard obviously did not rely on the then-current theory of digestion as a process of mere separation and redistribution of prefabricated organic substances. He rather seems to have assumed that gastric juice was responsible for bringing about certain chemical reactions. But nothing shows that he was aware of the contemporary characterization of the gastric ferment pepsin

in Germany. The productivity of Bernard's regime at this point resided neither in a new and explicit theory nor in a novel technique, but rather in the *combination of the means* with which he pursued the phenomena of the digestion of nutrients. He behaved exactly like the tinkerer whom Claude Lévi-Strauss, in his *La pensée sauvage*, characterized with the following words:

> The rule of the game is to be able to do with 'the means on board,' that is, with an always finite amount of tools and materials that are, in addition, heteroclite, because their composition does not stand in a definite relation to a particular actual project, nor with a defined project *tout court*, but rather represents the contingent result of all those opportunities that arise from renewing or enriching the arsenal, or maintaining it with the remnants of earlier constructions and deconstructions.[13]

It was the coupling of two techniques, the combination of digestion analysis in the test tube and comparative injection, that allowed Bernard to find an opening for the pursuit of the specific alterations and pathways of nutrients and digestion products in the body. An experimental system for the localization of metabolic processes appeared on the horizon. It is obvious so far that the dynamics of this exploratory pathway are to be understood less as a process of error correction but rather as a meandering survey of uncharted territory, with shorter or longer ventures in different directions, most of them broken off, in search of a few landmarks from which to proceed into the unknown.

Bernard summarized his work on digestion in his medical dissertation of 1843 on the decomposition of sucrose by means of gastric juice.[14] During the following four years, Bernard solidified his experimental protocols, but nothing turned up that led him beyond the first results. There was an experimental machinery that spun beautifully but remained mute. From 1844 onward, Bernard started to look into the effects of the arrow-poison curare, as a follow-up to his earlier experiments on the nervous system.

Over several years, no important result turned up on any of Bernard's ventures. They were years of no results, but yet they were decisive for the acquisition of what Fleck once called the "experiencedness" of the experimental scientist.[15] What Bernard accumulated in his daily efforts to build up and consolidate his *in vivo* system for the analysis of nutrient digestion was an open horizon of problems around one established fact, the decomposition of sugar. None of these problems could immediately be solved, and a wealth of observations, for the time being, did not join to form a coherent picture. It was only years later that, out of this temporary mess, a series of insights into animal metabolism resulted that comprised, besides the transformation of sugar, the cleavage of fats and proteins by the juice of the pancreas, and the specific paralyzing effects of curare. When he came back to it at a much later time, Bernard also contributed important insights to the abandoned field of respiration and animal heat. If one considers the mere chronology of Bernard's experimental publications, his discoveries must appear like a "chain of logically successive, genially conceived experiments," as one observer put it.[16] But historically, they were all the incalculable emanations springing from the channeled autocatalysis of an experimental system set in motion and playing out its forces of reorientation.

Early in the summer of 1848, Bernard started with a series of experiments that, within a few months, would result in one of his best-known contributions to animal

physiology. After conducting a series of experiments to assay the effect of pancreatic juice and the combined effect of bile and gastric juice, Bernard came back to the old, abandoned question of whether different kinds of sugar behave differently in circulation. But the question had by then been slightly altered: Was there an organ or a specific place in the organism where different sugars were preferentially decomposed? Several injection experiments pointed to the lungs as a place for the decomposition of grape sugar (glucose). However, when Bernard tested isolated lung tissue, considerable amounts of grape sugar remained intact in the vessel even after a long incubation. And more assays of the same sort soon suggested to Bernard that glucose disappeared in the liver as well. Once more, incongruities accumulated. Bernard realized that there was no way to proceed without a quantitative determination of these residual amounts of sugar. He interrupted the series of *in vitro* assays and repeated the injection experiments. But nothing new came from them; they remained inconclusive. Technical difficulties, such as reliable sugar determination, were a part of the problem. Another was the assumption, implicit in the experimental disposition, that the different sugars were indeed decomposed in a specific organ immediately upon their arrival.

The revolution of June 1848 interrupted the experiments for a short time, when the barricades prevented Bernard from getting into his laboratory. But at the beginning of July, again shifting to the test tube, he tested the influence of fibrin and of blood serum on sugar. Two days later, he injected a dye into a fasting dog, and recovered it from the urine. In parallel, he executed a sugar reaction with different body fluids taken from the starved animal. To his great surprise, the reaction with the blood serum was positive, although intestinal probes had assured him that the dog could not have absorbed the sugar from the digestive tract. Where did this sugar come from? “This is all very strange and requires further experiments for clarification,” Bernard noted into his laboratory journal.[17] This was the point where the very question became reversed. Somewhere in the body, there was obviously a hidden source of sugar that did not dry up even after a prolonged period of starvation. The question that he had followed so tenaciously and for so many years, namely, where sugar was *decomposed* in the body, all of a sudden was transformed into the question of its provenance in the absence of an external source, that is, its *composition*. The prevailing doctrine of physiological chemistry at the time was that the composition of sugar took place exclusively in plants, whereas animals were only able to decompose it. On the basis of the ruling theory, the new question therefore could not even have arisen; it would not have made sense in the existing conceptual framework of the time. A major reorientation for all of nineteenth-century physiology followed from this finding. The animal body, from now on, began to be viewed as a site not only of catabolic processes but of anabolic processes as well: as a place where substances were not only used up, but also built up.

From that moment on, Bernard concentrated all his experimental energy on the identification of the organ that delivered sugar to animals that had been starved or fed a sugar-free diet. But once more, results remained mixed and inconsistent. Two months of feverish experimentation followed, in the course of which Bernard tried to harmonize the different parameters of his system. Among the parameters was

the kind of diet, the duration of the starving period, the method and the time point of taking the blood, the locus and amount of blood withdrawn, the transition to a semi-quantitative sugar test, and finally, the clamping of the portal vein before the resection of the liver. On the 21st of October 1848, Bernard presented his results to the *Société de Biologie*, a society that he had founded the same year together with like-minded colleagues. As Holmes has pointed out, the manuscript of the lecture is a masterpiece of modern experimental prose. All traces of the ramifying and ultimately bottle-necking experimental groping process; the whole lingering course of the emergence, dislocation, flanking and by-passing of obstacles; the whole internal calibration of the experimental arrangement; the breakdown of the old question; the reorientation of all resources to the new question; in short, everything that characterizes the real process of research, the explorative wit of the experimenter and the differential dynamics of the experimental system—all that was sacrificed for the sake of the clear formulation of a hypothesis and its subsequent corroboration by an ingenious and decisive experiment.

And yet, the goal that Bernard continued to pursue in the summer of 1848, namely, to measure the extent to which sugar was decomposed in different organs and body fluids, should only have led him, given the test systems at hand, to abandon the project. His operative attentiveness, however, the implicit maxim of an experimental regime laboriously acquired in passing through many failures, enabled him finally to turn an oddity within an established experimental framework into an unprecedented event that opened a new horizon. He had cast his experimental net such that it led him to questions that, at the start, he would not have been able to pose. The only *a priori* decision under which Bernard's experimental regime took shape was, as Holmes has reminded us, his decision, against the dominant chemical tradition when studying the phenomena of nutrition, to privilege "the operation table over the test tube."[18] But as we have seen, this by no means excluded test tube experiments, in which body fluids reacted with body tissues, as supporting measures in his endeavor.

Slightly less than two decades later, Bernard published his *Introduction to the Study of Experimental Medicine*, a methodological treatise that became "something like the bible of scientific medicine."[19] As in his presentation before the *Société de Biologie* in 1848, the idealized representation of the experimental method he gave there compressed his own groping and meandering procedure into the straitjacket of an experimental philosophy that hid the actual process of discovery behind a logical exposition that hardened, in the latter half of the nineteenth century, into the standard presentation of a scientific publication: hypothesis, experiment, scientific result and, ideally, follow-up hypothesis. All the additional motions that no longer appear necessary for explaining a finding that made its career, remain suppressed. The (dis)order of research falls prey to the logic of representation. The epistemological consequence has been, for a long time, the marginalization of the research experiment that, despite the emphatic insistence on the empirical, not only characterized positivism but also logical empiricism and critical rationalism of the early twentieth century. Against this backlash, I want to insist that the productivity of a complex research endeavor depends on its capacity for orchestrating a polyphonic

texture of experimental operations within which the contingent, the unthought-of, the unprecedented can take on meaning.

Bernard's *Notes* have been preserved. They allow the historian to reconstruct the path of his experimental reasoning, which finds itself, at times, turned on its head in the published record.[20] In the concluding paragraph of the *Introduction*, we read: "Science steps forward only through new *ideas* and through the creative and original power of thought."[21] In the *Cahier rouge*, however, the notebook which accompanied Bernard's work between 1850 and 1860, we read: "All discoveries proceed necessarily from a newly observed *fact* whose meaning one starts to explore."[22] And slightly farther down: "At the point where there is nothing more to know, one must *find*."[23] These are notes taken in the pauses of daily research, in which Bernard felt himself unobserved and distant from the verdict of contemporary philosophers of science like Auguste Comte, and therefore freed from the coercion of self-stylization. Against Comte, he clung to the belief that "it is the vague, the unknown which moves the world."[24] And in such moments, he did not hesitate to see in the development of his science, physiology, nothing more than "a succession of evolving facts that follow each other in time, but do not necessarily engender each other. It is," he continued, nothing more than "a chain whose individual links do not possess a relation of cause and effect, neither to the ones that precede nor to the ones that follow."[25] In these notes, the conviction finds its expression that the course of research cannot be thought of as a causal succession, and yet, the links form a chain. The addition of each link remains dependent on those productive moments in which something is exposed that cannot be deduced from the foregoing stage. In such historical moments of indeterminacy, the question of right or wrong is suspended. The problem is not to determine whether a particular claim can be justified or not, but rather to create conditions under which facts, and claims concerning these facts, come into being.

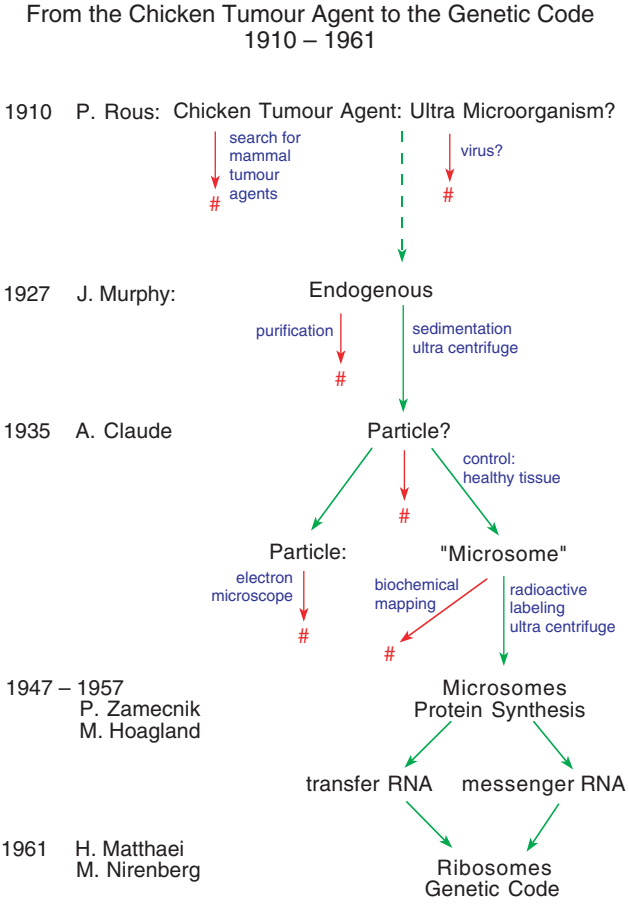
The micro-genesis of Bernard's laboratory work between 1841 and 1848 exposes such a process of manifestation. Through a series of starts and failures, of approaches abandoned, reconstructed, and modified again and again, Bernard gradually established a system whose characteristic configuration consisted of making physiological reactions within the living organism accessible to experimentation. Within the space of representation thus established, questions happened to be generated that would not have been possible in the context of either a continued measurement of mechanical parameters or even of a refined input–output chemistry. With the construction of this local system, physiology moved to the center of biology in the second half of the nineteenth century.

From the Chicken Tumor Agent to the Genetic Code

The decisive question is where the knowledge that we do not yet have comes from. The following, second example retraces a line of research in its material continuity, and at the same time conceptual discontinuity over half a century, from early

virology to cytomorphology to biochemistry and, finally, molecular biology. These disciplinary characterizations are indicative of major conceptual shifts, but from the perspective of following research objects along their meandering path from one stage to the next, they lose their organizing function for the narrative.

The story as it is told here is highly compressed, and it is roughly schematized in Fig. 2. A more extended version has been given elsewhere.[26] Nevertheless, I hope that the core of the dynamics by which research was driven are still recognizable. In our context, this story is the macroscopic counterpart to the microscopic analysis of the early laboratory years of Claude Bernard. Despite its much wider time horizon, it is also an open-ended story, a story without prior knowledge about its outcome, and, as in Walter Benjamin's description of the angel of history, it is a story that unfolds in breaks before the eyes of the observer who moves, driven by the blows of a storm called progress, backwards into the future.[27] "The question what to do, and why," says Michel Serres, "in these things like in others, knows only local,



differential, restricted answers. It cannot be integrated. We have no global answer and we will perhaps never have one."[28]

In 1910, oncologist Peyton Rous isolated a cell-free, filterable agent from the sarcoma of a Plymouth Rock hen that had been brought to his laboratory at the Rockefeller Institute in New York. The filtrate, when injected in healthy chicks, was able to induce the same malignant growth in their muscles. Rous thought of the agent as a virus. According to the ideas of the time, a virus was an ultra-microbial disease germ that could only be negatively characterized using the means at hand: It could not be seen in the light microscope, it was not retained by bacterial filters, and it did not grow on sterile culture media. The community of medical oncologists received Rous' report with utmost skepticism, and Rous himself was not able to identify, in the course of the following years, similar agents in other groups of higher animals, not to mention humans. Stuck at an impasse and discouraged by his colleagues' suspicion, he turned away from the subject. The analysis of the active substance labeled "chicken tumor I agent" entered into a latency period during which the tumor was kept at the Rockefeller Institute and faithfully handed down from one generation to the next. But it had lost its status as an object of immediate scientific concern.

It was only at the end of the 1920s that the agent was reactivated by James Murphy, a student of Rous and at that time director of the oncological laboratory at Rockefeller. But it was reactivated under a different perspective. Together with Albert Claude, a young medical postdoctoral student from Belgium, Murphy wanted to demonstrate that the agent was not, as Rous had suspected, an extraneous viral disease germ, but rather an endogenous cellular component that had undergone a malignant transformation, perhaps an enzyme of sorts.

Over half a decade, Claude tried to purify and characterize the chicken tumor agent by means of biochemistry, with no great success and with ambiguous results. It was only in the middle of the 1930s that new movement came into the field due to the introduction of a novel technique. Claude started trying to sediment the putative malignant cellular entity by means of a high-speed centrifuge. Within a year, his efforts resulted in a concentration of the agent by a factor of about 3,000, which appeared to be a quantum leap in comparison with the enrichment by a factor of about 20 that he had achieved with conventional techniques before. A second insight resulted from the new procedure. Since the agent could be sedimented readily in the gravitational field, it had to be a particle, and one of considerable size. That was not quite the thing to expect if Murphy's assumption was correct that an endogenous enzyme gone wild was involved in the malignant growth. But this was only a first link in a chain of displacements that followed from this step. In his sedimentation experiments, Claude had included a control. In parallel to the malignant samples, he centrifuged the cell sap of healthy tissue. To his surprise, the control did not differ, either in its sedimentation behavior or in its chemical constitution, from the malignant sample, with the exception, of course, that it was not infectious. Here again we have a "Columbus effect." Claude had looked for an endogenous tumor agent. What he found were cytoplasmic granules derived from normal cells. Within their mass, the tumor agent remained hidden. It resisted becoming a manipulable scientific object. At this point, the tumor agent was again put aside. Within the

confines of the existing system, its nature remained elusive. Instead, Claude went on to use the potential of high-speed centrifugation to arrive at a more physically and biochemically differentiated picture of sedimentable cytoplasmic particles.

Besides the nucleus, there was another class of particles that had been visualized by light microscopy within the cell. They were known under the name of mitochondria. Thus it appeared to be obvious, to Claude, to identify his centrifugal fraction with mitochondria, with fragments thereof, or early developmental stages of them. It soon turned out, however, that this assignation did not hold. He had been led astray by the collective biological wisdom of the day. The two spaces of representation, that of traditional light microscopy and that of centrifugal fractionation, could not be harmonized without friction. By introducing different speeds of sedimentation, Claude succeeded in separating a bigger fraction of particles from a smaller one. The bigger could now be associated more safely with mitochondria, the smaller had not yet been characterized. He recalled his earlier assumption and termed his novel fraction "microsomes." In terms of biochemistry, the granules contained mainly lipids and proteins. A smaller, yet considerable fraction of pentose nucleic acids was invariably present, too. Claude tried to identify the granules with a specific cellular function. But he failed to arrive at a clear-cut pattern. The available enzyme tests, however, could successfully be applied to the mitochondria, which in the course of the 1940s revealed themselves as the centers of cellular respiration. What worked with one class of particles did not work with the other. The research program ramified. Impasses on one path were compensated with steps forward on another.

In the early 1940s, Claude had the chance to integrate the newly developed technology of electron microscopy into his system of cytoplasmic particle exploration. His hope was that electron microscopy would accomplish what light microscopy had failed to do, namely to force the microsomes into the realm of the visible. But here again, results came from the mitochondria, while the microsomes tenaciously remained invisible. Electron microscopy, however, allowed the relegated tumor agent of Rous to enter the stage again. What centrifugation had been unable to achieve became possible with electron microscopy. Under the electron microscope, malignant tissue appeared dotted with small electron-dense particles that had no counterpart in healthy controls.

The series of no results with microsomes was only broken when still another technique was introduced into the efforts of characterizing subcellular components *in vitro*. Claude's work had been cytomorphologically oriented, and his biochemistry consisted of standardized enzyme tests applied to the different fractions of his centrifuge runs. After World War II, radioactively labeled compounds started to revolutionize physiological and biochemical analysis. In the late 1940s, several groups, among them Paul Zamecnik's at the Massachusetts General Hospital in Boston, introduced radioactive amino acids into the analysis of protein synthesis. The context was again one of oncology. At issue was the biochemical characterization of growth in cancer tissue as compared to normal tissue. The transition from the use of tissue slices to a cell-free system brought the fusion of a new, radioactivity-based analysis of metabolism with the subcellular cytomorphology developed during the preceding decade by Claude and his Rockefeller colleagues. The resulting experimental systems were

no longer dependent on the principle of quantitative purification. With the help of radioactive markers, specific metabolic reactions could also be followed in mixtures of compounds. To achieve such conditions, however, was itself not a trivial task, as the early history of *in vitro* protein synthesis research amply shows. One could even read the establishment of such systems as a history of the exclusion of reactions not belonging to the protein synthesis pathway. In the course of setting up a system, crude fractionation of the cell sap became crucial. In this way the microsomes left Claude's hands and entered into the realm of protein synthesis research. Eventually, they were also visualized under the electron microscope as electron-dense particles of a fraction that was essential for sustaining protein synthesis. They turned out to consist roughly of half RNA and half proteins, and in due time were renamed ribosomes.

For about a decade, from 1945 to 1955, the field of protein synthesis technically evolved without a final decision between two alternative conceptions: that of protein synthesis as a reversal of proteolysis, or that of protein synthesis as a different, phosphate-energy-driven reaction. It was only in the middle of the 1950s that this question was solved, when Mahlon Hoagland and Paul Zamecnik demonstrated that the amino acids of their system became activated by the nucleotide adenosine triphosphate (ATP) and that these activated amino acids condensed on the microsomes to form polypeptides. But beyond the resolution of a question that had been in suspense for a long time without hampering the research process, a decisive shift of perspective at this point was again brought about by the occurrence of an unprecedented event. Zamecnik, in trying to demonstrate the additional synthesis of ribonucleic acids in his protein synthesis system via the incorporation of the radioactive nucleotide ATP, found by way of a control that small, soluble ribonucleic acids in his system became labeled by amino acids instead. This RNA had been recognized for several years, but was taken to be a contamination of the soluble enzyme fraction that could not be purified away for the time being.[29] Within a couple of years, this RNA–amino acid hybrid shifted the perspective of work with the protein synthesis system from questions of biochemistry to questions of molecular biology. Soluble RNA revealed itself as a mediating carrier of genetic information from nucleic acids to proteins. The protein synthesis system was reconfigured accordingly. A few years later, systems of this kind became the stage on which the genetic code was deciphered. But interestingly enough, they remained a battlefield for decades to come with respect to the elucidation of the molecular details of peptide bond formation, precisely the question that had dominated, if not its very beginnings, then the early steps of its proliferation.

Concluding Remarks

Despite their historical distance and despite the considerable differences in time scale, the two examples given in this paper show remarkable similarities with respect to the reorienting forces of the experiment. It is as if the research process

were of a fractal nature. Both in the minute detail and in the grand lines, we get the impression of a highly non-linear endeavor. It is the nature of exploration that it moves on the borderline between knowledge and ignorance. There is a space of pre-conceptuality in research. Perhaps no one has better captured the subtle dialectics of this balancing act than Johann Wolfgang Goethe, who stated in his *Maxims and Reflections*: "You never go further than when you no longer know where you are going."[30] There is positivity in ignorance, then, in a state where concepts are lacking. Projections can help in such situations, but do not grant access to successful avenues. Along any particular research trajectory, impasses appear to be the rule. Ramifications are multiple, with most twigs ending nowhere. The route on which the research endeavor will continue will only be known *ex post facto*. Old questions that inspired fierce experimental efforts can get lost and forgotten. Reorientations are the order of the day. They may lead in directions completely different from the one chosen at the beginning, but they can also, after the establishment of new experimental conditions, lead back to issues once abandoned and thus reveal themselves as detours.

Instances of both possibilities are present in the two case studies. A Darwinian tree of divergence might therefore not be the best picture for the process. A meshwork of shorter or longer paths at times diverging from each other and ending nowhere, but at others again merging into each other might give a better image. As mentioned at the beginning, clear-cut demonstrations or refutations of clear-cut assumptions are not excluded, but they appear to be rather extreme and infrequent boundary cases of a process that is dominated and driven by situations that come to lie in between. These appear to make up the bulk of the research process. To realize that one has been wrong is not the rule. Not being able to decide whether one is wrong or right is, according to the available evidence, the everyday reality of the researcher. To find out what researchers do in this middle ground, how they behave when they do not have secure ground under their feet, how they manage to "think with their hands" or to follow, as it were, their intuition, is therefore a valuable effort towards a historical understanding of the dynamics of empirically driven research. This effort situates itself in the context less of a narrowly conceived epistemology of error, but rather of a broader epistemology of *errancy*. In this broader perspective, the different forms of going amiss will have their place, but within a more encompassing typology of what I call, for want of a better term, the forms of experimental reorientation.

Notes

1. François Dagognet, Préface à Claude Bernard, Introduction à l'étude de la médecine expérimentale. Flammarion, Paris 1984, p. 18.
2. Gaston Bachelard, Le nouvel esprit scientifique (1934). Presses Universitaires de France, 10th edition, Paris 1968, p. 173.
3. Bachelard 1968, p. 173.
4. Gaston Bachelard, La formation de l'esprit scientifique (1938). 6th edition, Vrin, Paris 1969.

5. On experiment and exploration, see Friedrich Steinle, Experiments in history and philosophy of science. Perspectives on science 10 (2002), pp. 408–432.
6. Ludwik Fleck, The Genesis and Development of a Scientific Fact (1935). The University of Chicago Press, Chicago 1979, p. 69.
7. Claude Bernard, Cahier des notes 1850–1860. Edited by Mirko Grmek, Gallimard, Paris 1965, p. 145.
8. Gaston Bachelard, Essai sur la connaissance approcchée (1928). Vrin, Paris 1987, p. 284.
9. This and the following section are based on Hans-Jörg Rheinberger, Historische Beispiele experimenteller Kreativität in den Wissenschafteen. In Walter Berka, Emil Brix and Christian Smekal (eds.), Woher kommt das Neue? Kreativität in Wissenschaft und Kunst. Böhlau, Wien 2003, pp. 29–49.
10. Mirko Grmek, First steps in Claude Bernard's discovery of the glycogenic function of the liver. Journal of the History of Biology 1 (1968), pp. 141–154. Frederic Lawrence Holmes, Claude Bernard and Animal Chemistry. Harvard University Press, Cambridge 1974.
11. Claude Bernard, Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux (1878/1879). Vrin, Paris 1966, p. 19.
12. Holmes 1974, p. 129.
13. Claude Lévi-Strauss, La pensée sauvage. Plon, Paris 1962, p. 27.
14. Claude Bernard, Du suc gastrique et de son rôle dans la nutrition. Rignous, Paris 1843.
15. Fleck 1979, p. 96.
16. Nikolaus Mani, Die historischen Grundlagen der Leberforschung, II. Die Geschichte der Leberforschung von Galen bis Claude Bernard. Schwabe & Co., Basel 1967, p. 369.
17. Holmes 1974, p. 420.
18. Holmes 1974, p. 443.
19. Karl Rothschuh, Introduction to Claude Bernard, Einführung in das Studium der experimentellen Medizin. Barth, Leipzig 1961, p. 11.
20. Compare the pioneering work of Mirko D. Grmek, Le raisonnement expérimental et les expériences toxicologiques chez Claude Bernard. Droz, Genève-Paris 1973.
21. Claude Bernard, Introduction to the Study of Experimental Medicine. Translated by Henry Copley Greene, Henry Schuman, 1949, p. 226, emphasis added.
22. Bernard 1965, p. 121.
23. Bernard 1965, p. 135, emphasis added.
24. Claude Bernard, Philosophie. Manuscrit inédit. Edited by Jacques Chevalier, Editions Hatier-Bouvin, Paris 1954, p. 26.
25. Bernard 1954, p. 14.
26. For the earlier part of this story see, Ilana Löwy, Variations in meaning in discovery accounts. Historical Studies in the Physical and Biological Sciences 21 (1990), pp. 87–121; see also Hans-Jörg Rheinberger, Cytoplasmic particles. The trajectory of a scientific object. In Lorraine Daston (ed.), Biographies of Scientific Objects. University of Chicago Press, Chicago 2000, pp. 270–294.
27. Walter Benjamin, Über den Begriff der Geschichte, IX. In Gesammelte Schriften, Band I.2. Suhrkamp, Frankfurt am Main1974, pp. 697–698.
28. Michel Serres, Hermes V. Le passage du nord-ouest. Editions de minuit, Paris 1980, p. 127.
29. On contamination, see also the contribution of Christoph Hoffmann to this volume.
30. Johann Wolfgang v. Goethe, Maxims and Reflections. Penguin Books, London 1998, p. 108, Nr. 901.

Concepts from the Bench: Hans Krebs, Kurt Henseleit and the Urea Cycle

Kärin Nickelsen and Gerd Graßhoff

Introduction

On 30 November 1932 Sir Frederick Gowland Hopkins, in his capacity as President, chose an unusual subject for the Royal Society of London's Anniversary Address: the main scientific topic of his lecture did not concern the recent success of an eminent scientist of the time, as was usually the case; rather, he referred to a young German doctor called Hans Adolf Krebs, whose findings had greatly impressed him. He quoted them as an important example of success in the new discipline of biochemistry:

> The facts as revealed have just that degree of unexpectedness – if I may use the phrase – which was to be expected in a biochemical phenomenon. I often find myself compelled to assert that, though biochemical events are, of course, limited by chemical possibilities, they are not safely to be predicted by chemical probabilities, even when these are strong.[1]

The unexpected facts referred to in these lines was the discovery of the urea cycle. Up to this point, Hans Krebs had enjoyed the unspectacular existence of a hospital doctor who, in addition to his clinical obligations, also managed to do some research work.[2] Together with Kurt Henseleit, his research assistant, Krebs had published his findings on urea biosynthesis in the first half of 1932 in three short papers, all entitled "Inquiries into the Formation of Urea in the Animal Body".[3] Almost immediately afterwards, he was promoted to the status of a *Privatdozent* at the University of Freiburg and received praise from the discipline's leading figures. Carl Neuberg, for example, congratulated Krebs on his "wonderful successes" and enthusiastically concluded his letter with "You have achieved great things!";[4] a short time later Neuberg put Krebs's name forward for a chair in physiological chemistry at the University of Münster.[5] Jakub Parnas asked Krebs if he would take on one of his collaborators as a research fellow in his laboratory; Otto Meyerhof wanted Krebs to lecture

K. Nickelsen (✉)
History and Philosophy of Science, Institute of Philosophy, Sidlerstr. 5, University of Berne, 3012 Berne, Switzerland
e-mail: nickelsen@philo.unibe.ch

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_7,

at Heidelberg; and Max Planck invited Krebs to give a talk at Harnack House, the official lecture venue and social centre of the Kaiser Wilhelm Institutes in Berlin.[6] The driving force behind Planck's invitation was Krebs's former teacher Otto Warburg, who only two years before had dismissed Krebs from his laboratory in Berlin-Dahlem. (According to Krebs, Warburg did not think that he had "sufficient ability for a successful research career".[7]) Warburg obviously completely changed his view of Krebs's ability after the latter's discovery of the urea cycle: he offered him a new post in Berlin when Krebs lost his appointment in Freiburg after the Nazis had come to power[8] and in later years he even referred to Krebs as his "favourite pupil".[9]

However, this splendid triumph – the discovery of the urea cycle – was, from a certain perspective, the outcome of a research project in which Krebs much of the time was stuck in confusion. For months he had not even been able to reproduce effects that he should have been able to obtain by following standard assumptions – that is, obtaining the synthesis of urea from amino acids via the ammonia stage. In fact, Krebs had had trouble producing any urea at all. And when he did finally produce the substance in sufficient quantities, it occurred under totally unforeseen circumstances, so that Krebs was at a loss as to how to make sense of it. But this appearent dead end turned out to be productive: for it was only from this point that Krebs was finally able to break away from allegedly well-founded standard assumptions – something that one does not undertake lightheartedly – and eventually come up with a surprising solution to the problem, which included formulating a new concept of how some substances react with each other.

As is well known, Krebs and Henseleit proposed that the reaction path of urea synthesis forms a cycle that includes the combining of three amino acids (see Fig. 1). This was the first cyclic reaction path to be elucidated and aroused an enthusiastic response from the scientific community. However, contrary to widespread belief,

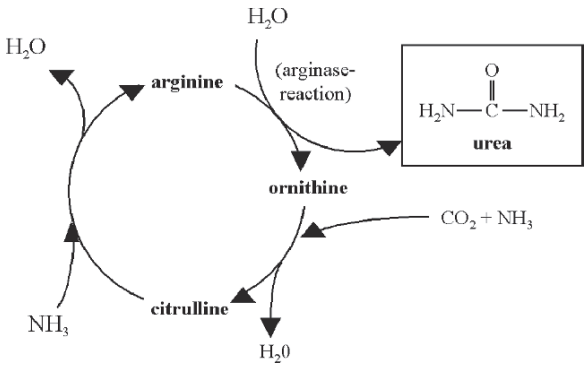


Fig. 1 The urea cycle. Ornithine (on the *right*) absorbs one molecule of carbon dioxide and one molecule of ammonia, which react together to form citrulline (thereby releasing water). Citrulline absorbs another molecule of ammonia to form arginine (again, releasing water). Arginine hydrolysis finally releases one molecule of urea and thereby regenerates ornithine; this last step requires the presence of the enzyme arginase

the notion of a "cycle" or even of a "reaction cycle" was *not* a new concept in 1932, as we shall see later on. The new concept that grew out of the seemingly hopeless situation Krebs found himself in was something else entirely. Before Krebs and Henseleit presented the cyclic reaction path in their first paper of 1932, they had discussed at length another subject, namely the role of the amino acid ornithine in this path and concluded: "According to these experiments ornithine acts in urea synthesis in the living cell like a catalyst."[10] This was repeated in their third paper in which they stated that "the action of ornithine in the living cell resembles a catalytic action".[11] Speaking of a catalyst or of a catalytic action in connection with a simple organic substance such as ornithine – this was new in 1932. Indeed, it was so new that two years later, in 1934, the Hungarian-American biochemist Albert Szent-Györgyi almost apologized for proposing that, in carbohydrate oxidation, succinate also acted catalytically, admitting that it "may appear strange to assign such a simple substance as succinic acid a catalytic function".[12]

The novelty of his proposal made Krebs extremely cautious in his formulation – indeed, he did not give ornithine the full status of a catalyst. Other researchers who consequently picked up the idea were not as hesitant,[13] and Krebs's own caution also disappeared as time went by. His Nobel lecture (Krebs was awarded the Nobel Prize for physiology or medicine in 1953 for the discovery of the citric acid cycle, also known today as the "Krebs cycle") included mentions of many substances that act as catalysts – not only succinate, as Szent-Györgyi had already proposed, but also fumarate and citrate. It was this insight, Krebs maintained in his lecture, which provided the crucial evidence for assuming a cycle in the case of citric acid as well.[14]

Thus we are confronted with a situation in which, during the greater part of their project, things went awry for the players, until they came to a dead end at a crucial moment. Eventually, however, they arrived at a creative solution to the problem. The question to be followed in this paper is how Krebs and Henseleit came up with the notion that ornithine acts "like a catalyst" in the urea cycle or, to put it in more general terms, how their conceptual work interacted with their experimental practice.

The Setting and the Protagonists

In 1931 Hans Krebs was a young and aspiring but not yet particularly outstanding researcher. Born in Germany in 1900, he completed his medical studies (carried out at Göttingen, Freiburg im Breisgau and Berlin) in 1923, and thereafter practised for some time as a doctor. In 1926 Otto Warburg appointed Krebs as a scientific assistant at the Kaiser Wilhelm Institute for Biology in Berlin-Dahlem, where Krebs remained until 1930.[15]

This period with Warburg proved to be of major importance to Krebs. At the institute he acquired the necessary skills and techniques to carry out his own research: a stock of physiological, chemical and methodological knowledge, including the

so-called tissue-slice technique that Warburg had developed for his research into cell oxidation. Warburg had found that very thin slices of human or animal tissue – 0.2–0.4 mm thick, that is, ten to twenty layers of cells – could be kept alive for some time in a saline solution and still be capable of carrying out all the metabolic processes that they would perform in the living body: a crucial precondition for studying all those in vitro processes that require living cells. Everyone working in Warburg's laboratory had to master this technique. And one of Krebs's first tasks there was to describe systematically the tissue-slice technique and the relevant manometric methods in an article for one of the laboratory manuals (published in 1928) of the time.[16] So, if anyone was familiar with the tissue-slice technique and all its inherent possibilities and limitations, it was Krebs.

However, working with the demanding Warburg was not easy, and in 1930 Warburg dismissed his research assistant without helping him find a new job – not an easy task in Weimar Republic Germany. Research work in biochemistry was largely unpaid at this time, so Krebs went back to practising as a hospital doctor, first in Altona, near Hamburg, and one year later in Freiburg im Breisgau, where the newly appointed professor Siegfried J. Thannhauser had offered Krebs a more attractive post. In addition to his clinical obligations, Krebs was given the chance to set up a small laboratory, equipped with two pieces of "Warburg apparatus", which were indispensable for applying the tissue-slice technique. The first project on which Krebs embarked in this laboratory – starting in June 1931, that is, almost immediately after having taken up this new post – was his inquiry into the biosynthesis of urea: that is, the series of chemical reactions by which mammals convert their uptake of nitrogen to urea. Krebs was assisted in his work by Kurt Henseleit, a medical student who had never worked in a professional laboratory but wanted to do some research for his M.D. thesis.

Krebs's Path to the Cyclic Solution

How did Krebs arrive at his reaction path of urea biosynthesis and how did he envisage the catalytic function of ornithine described above? The following sections outline the path of Krebs's discovery, with special emphasis placed on the methodological strategies that he applied at different stages of his research.[17] This reconstruction is based on an analysis of the laboratory notebooks kept by Krebs and Henseleit, all of which have survived. These records provide the only historical source of their work entirely free from any retrospective distortions.[18] In later years, Krebs repeatedly told the story of his discovery himself, in autobiographical accounts as well as in many interview sessions with the historian Frederic L. Holmes.[19] Although valuable, these retrospective accounts tend to include historical fallacies: by projecting later knowledge and insight into earlier phases of one's own work, retrospective accounts are not as reliable as direct contemporary evidence.[20] This means that one must try to reconstruct Krebs's guiding principles and working hypotheses from his experimental work alone.

The Problem

According to Krebs, the subject of urea was chosen because its biosynthesis was supposed to be a comparatively simple process: the molecule is rather small and symmetrically structured (see Fig. 2).[21] Additionally, Krebs had a well-established body of existing knowledge on which he could rely, since urea synthesis was a long-standing problem of the discipline – no minor asset from a beginner's perspective. It had repeatedly been demonstrated, for example, that the rate of urea synthesis in living animals increases when they are fed additional proteins or isolated amino acids. By the 1870s amino acids were firmly established as intermediates in the reaction path from proteins to urea; it had also been found that urea was primarily synthesized in the liver.[22] However, the details of the pathway remained obscure – not least owing to methodological difficulties. Metabolic processes were at this time usually studied by applying the so-called "perfusion method", which had been developed by Carl Ludwig and his collaborators in the 1860s. This method enabled researchers to isolate single organs from an animal's body and keep them alive for some time. Substances that were assumed to stimulate metabolic processes were guided through these organs and the results were measured by determining the composition of the excurrent fluid. Although this method was a great improvement on earlier procedures, it still proved inadequate for elucidating the actual steps of a mechanism. Thus, as for other processes of intermediate metabolism, several hypotheses were proposed for the path of urea formation in the liver, ranging from the dehydration of ammonium carbonate to a biological version of Friedrich Wöhler's test-tube preparation of urea from ammonium cyanate. However, no conclusive evidence was available for any of these hypotheses. Krebs took up the challenge and decided that the tissue-slice technique he had acquired in Berlin might circumvent the methodological difficulties of the perfusion method – a decision that was to bring him the first major success of his career.

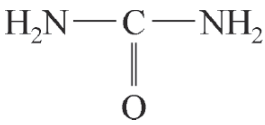


Fig. 2 The structural formula of Urea

The Standard Hypothesis

In 1930 urea formation was generally thought to be part of the decomposition of proteins in the animal body. It was widely held that proteins from nutrition were broken down into amino acids, and were consequently deaminated. During this process, considerable amounts of toxic ammonia are produced. In order to prevent the body from being harmed, the ammonia was then converted by the liver to the far

less toxic urea, which was finally excreted. This view was clearly expressed, for example, in Otto Neubauer's textbook on protein metabolism, the standard reference work on this subject at the time:

> By establishing amino acids as regular intermediates in protein decomposition and by proving their deamination, the problem of urea synthesis was greatly simplified, since the main issue can now be formulated as follows: *How is urea formed from the ammonia released by the amino acids?*[23]

This hypothesis can be represented in a two-step-process, assuming X and Y to represent additional, so far unknown, reactants:[24]

$$\text{Amino Acid} + \text{X} \rightarrow \text{NH}_3\ [\text{Ammonia}] + \text{Y}$$

$$\text{NH}_3 + \text{Y'} \rightarrow \text{Urea} + \text{Z}$$

For his first experiments Krebs adopted this hypothesis without modification – he had no reason to do so – and tried to figure out which molecules could act as precursor substances before the ammonia stage. Did various amino acids act differently? Were there other nitrogenous substances that might feed ammonia into the last part of the reaction sequence?

If ammonia was an intermediate on the pathway from precursor substance to urea, its addition should increase urea formation, at least at the same rate as the precursor.[25] Consequently, in the first stage of their research Krebs and Henseleit tested a number of amino acids and other nitrogenous organic substances (for example, pyrimidines) and compared their rate of urea formation with that of ammonia. Every experiment was diligently recorded in their laboratory notebooks; Fig. 3 shows a typical example of their records. On the top part of the page the title of the experiment was given, which usually referred to the substances and processes to be tested. The experimental procedures that were carried out before the tissue slice was planted in the vessel were described below the title; additionally, Henseleit specified the experimental parameters, such as, for example, the temperature and the period of time during which the slices were exposed to the substances in the solution. A number of columns document various parameters for each vessel, and the final results of urea determination can be found at the bottom of the page, underlined twice. Occasionally, the findings of the experiment were summarized, together with a few additional comments.

The first experiments of this stage were performed by Krebs, about a month before Henseleit joined the project. After having established the rough experimental set-up – including a crucial refinement of the measuring process[26] – Krebs tested the effect of potential urea precursors for the first time on 8 July 1931: the amino acid alanine and one of its derivatives, phenylalanine. Alanine was a fairly obvious choice, since it is one of the smallest, simplest and cheapest of all amino acids; phenylalanine might have been included to test the additional influence of a side chain, since its structure consists of alanine plus a phenyl ring. The outcome was indecisive: although alanine increased urea formation, it did so only slightly; the addition of

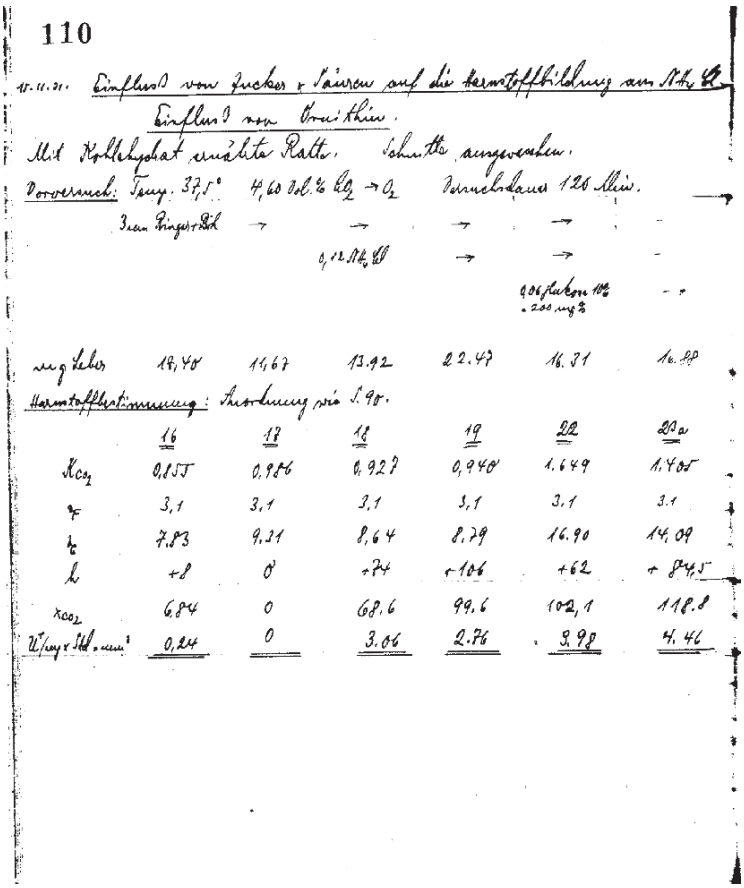


Fig. 3 The page from Henseleit's laboratory notebook that documents the crucial experiment of 15 November 1931 in which the ornithine effect first emerged

phenylalanine had even less effect. "Very little U^+ [urea] formation", reads Krebs's comment on this experiment (see Fig. 4 for the respective entry in his notebook). The standard to which Krebs evaluated the urea formation as "little" can be taken from the lower part of the page. Here, Krebs wrote down the following equation:

$$C_3H_7O_2N + 3\,O_2 + 2\,H_2O = 3\,CO_2 + 3\,H_2O + 1\,NH_4OH$$

This reaction scheme summarized the processes that should have occurred in the test tube if a complete degradation of alanine to ammonia had taken place – corresponding to the first part of the standard hypothesis. For every six moles of O_2 consumed in the reaction, two moles of ammonia should have been formed, as Krebs notes in the next lines, which should then have released one mole of urea. However, the conversion rate of alanine to urea in the test tube proved considerably lower than

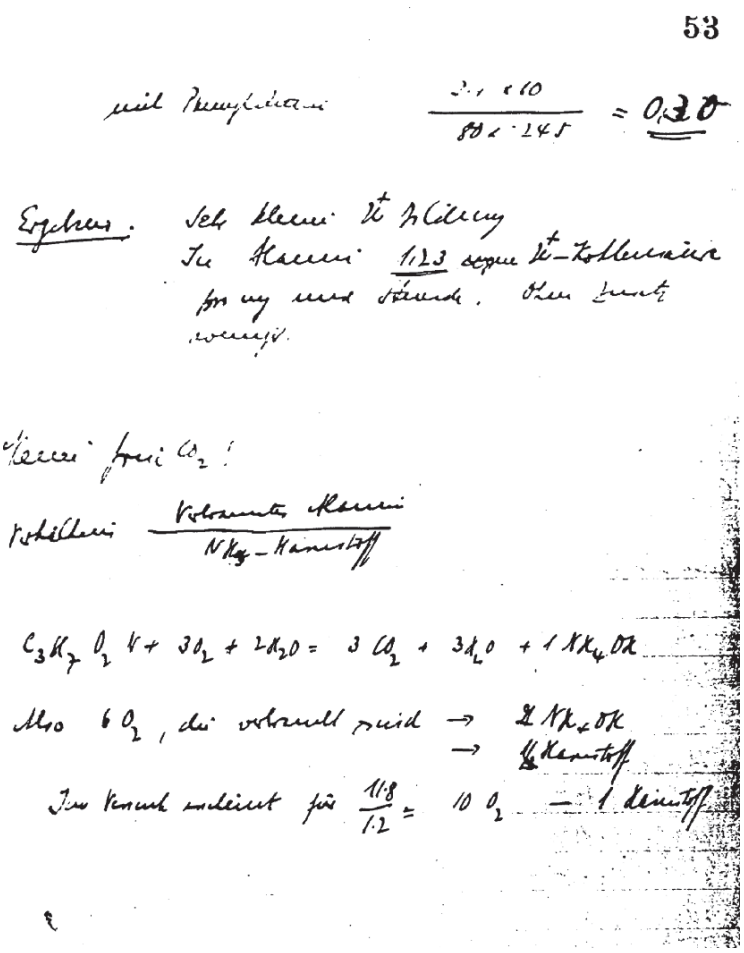


Fig. 4 A page from Krebs's laboratory notebook: above, a comment on the outcome of the alanine experiment; below, his calculation for the theoretical rate of urea formation from alanine

this hypothetical value: only for every ten moles of consumed O_2, had one mole of urea been synthesized. Krebs repeated his testing of alanine the following day, this time directly comparing the results with those of the tested ammonia. Again, the findings were disappointing: adding alanine to the tissue hardly stimulated any urea formation; adding ammonia did to some extent but not nearly as strongly as expected.

Refining the Set-up

Having obtained these results, Krebs (and later Henseleit) then conducted a series of experiments, carried out in August 1931, in which the formation of urea from

ammonia was explored in detail. Krebs and Henseleit tested whether aerobic or anaerobic conditions were required to produce urea (3 August), whether the formation took place only in the liver or also in other tissues (4 and on two occasions on 10 August), whether the addition of glucose influenced the reaction and, if it did, how (6 and 11 August). They also varied the concentration of ammonia (7 August), the pH of the solution (8 August), tried to use serum instead of Ringer's solution as an alternative medium for the reaction (10 August) and exposed the tissue for a longer time to the nitrogenous medium (11 August), to name the most important parameters at this point.

These experiments may be interpreted as attempts to find out as much as possible about the second part of the standard hypothesis, namely how ammonia reacts to form urea:[27] a kind of "explorative experimenting", to cite the term coined by the historian of science (and co-editor of this volume) Friedrich Steinle. However, we would argue that they should be considered to be attempts to optimize the experimental set-up for the reaction, which Krebs and Henseleit eventually wanted to explore: namely, the formation of urea from nitrogenous substances. The values Krebs had obtained for urea formation from alanine and phenylalanine were surprisingly low; even for ammonia, which was an undisputed precursor on the pathway from proteins to urea, his results were unsatisfying. This observation might indicate that some additional, circumstantial factor was missing. Krebs knew that the occurrence of a metabolic reaction not only depended on the presence of the reactants but also on many other factors, including, for example, the temperature, certain percentages of oxygen and carbon dioxide, the composition of the reaction medium, and so forth. Krebs's and Henseleit's varying of the parameters was, therefore, presumably motivated by their attempt to improve their insight into the general requirements of the experimental set-up. These variations allowed for the possibility that alanine was one of the urea precursors for which they were looking but that its effect was overshadowed by some missing factor. Only after Krebs and Henseleit had tested the most obvious parameters and had slightly adapted the set-up and measuring procedure accordingly did they return to their search for nitrogenous precursors.

This was not the last time that they tried to optimize the conditions of their experiments.[28] It is interesting to note that Krebs and Henseleit regularly repeated these variations in parameters whenever any of their results could not be reproduced or after a sequence of disappointing results on one topic.

In Search of a Precursor

In the following weeks, Krebs and Henseleit continued along the same lines. From August to November, they tested the following nitrogenous substances as potential precursors in the reaction path (listed in order of their first appearance in the notebooks): ammonia, alanine, phenylalanine, glycine, thymine, thymosine, uridine, uracile, cysteine, arginine, methylamine, cystine, ammonia carbamine, ammonia cyanate, asparagine and aspartate. Despite continuously refining the experimental

set-up, none of the substances yielded as much urea as ammonia – an amount that was already paltry when compared with theoretical expectations.

There was only one exception to the generally disappointing results: the addition of the amino acid arginine on 21 October 1931 yielded the first substantial rate of urea formation, although the result did not impress the two researchers. It had long been known that arginine decomposed to urea and ornithine under the influence of the enzyme arginase.[29] However, it was widely held that this reaction could not explain the main pathway of urea formation from proteins for two reasons. First, the low frequency of arginine in proteins could not account for the high amount of urea produced. In quantitative terms, only about 6 per cent of the urea produced by the body could originate from a direct hydrolysis of the nutrients' arginine.[30] Second, the arginase reaction also occurred in tissue that had been ground, whereas urea formation in liver tissue, stimulated by the supply of proteins, amino acids, or ammonia, required the action of living cells and ceased as soon as the cell structure was damaged or had died.[31] Therefore, the arginase reaction was generally thought to be a subsidiary pathway of urea biosynthesis, quite distinct from the main mechanism.[32]

The (Alleged) Reactant Combining Effect

Arginine remained the exception. No other substance was able to induce any substantial urea production. The overall results of Krebs's and Henseleit's experiments thus clearly demonstrated that their set-up still lacked something. As mentioned previously, they had repeatedly tried to vary the standard parameters; furthermore, they had attempted to stimulate the metabolism by adding glucose or other carbohydrates. However, neither had had any effect. So, on 26 October Krebs and Henseleit tried something new: instead of only testing a potential precursor substance and ammonia separately (and comparing the results afterwards), they set up an experiment in which the effect of a combination of ammonia and an amino acid was tested. For this new approach, Krebs and Henseleit went back to the simple amino acid with which they had started their research, alanine, and set up three vessels: the first with only alanine, a second with only ammonia, and a third with both alanine and ammonia.

In an interview with Holmes, Krebs recalled that he had had an inkling that one of the nitrogens in urea "comes from ammonia and the other comes directly from amino acids".[33] Although Krebs did not recall so in later years, this mysterious hunch could well have been inspired by Neubauer's textbook, in which the possibility of a degradation of amino groups together with a carbon atom is discussed (although Neubauer himself considered this path implausible).[34] But since he had got nowhere with the standard hypothesis, Krebs perhaps felt he had no alternative but to resort to considering non-standard possibilities. There is no need to assume that Krebs had much confidence in Neubauer's hypothesis, but at least he could try to eliminate it.

Yet Krebs found the following: on its own ammonia yielded very little urea (Q_H[35] = 1.13), alanine alone produced a little more (Q_H = 1.59), while the combination of these two substances yielded a quotient of 3.98, that is, more than the mere sum of the isolated results. A new working hypothesis seemed to present itself:

$$\text{Amino Acid} + NH_3 \rightarrow \text{Urea} + Z$$

In the following few days Krebs and Henseleit tried to confirm this hypothesis and also to extend it to other substances, but the alleged effect could not be reproduced, neither with alanine nor with any other substance (they tested, for example, thymine, thymosine, asparagine, glycine, uridine and aspartate). In none of these cases did the combination of the substance with ammonia yield more urea than ammonia alone.

The Ornithine Effect

However, Krebs did not give up – and the next experiment along the same lines (that is, a nitrogen donor tested alone as well as in combination with ammonia) brought ample rewards. On 15 November, Henseleit tested a broad range of metabolism-stimulating substances and their influence on urea formation; he also included, though, one additional vessel to which a new potential nitrogen donor was added: the diamino acid ornithine (see for the experiment's record the corresponding page of the notebook in Fig. 3; Fig. 5 shows the substance's structural formula). At that time, Henseleit had already been experimenting with duplicate runs;[36] hence, he set up two vessels each with only ammonia and ornithine, and another two in which the two substances were combined. The results were stunning: whereas in all the other cases since August none of the measurements had exceeded maximum values for a Q_H of between 3 and 4 at the very most (except for arginine), the two vessels of combined ornithine and ammonia yielded urea formation rates of 6.74 and 8.04

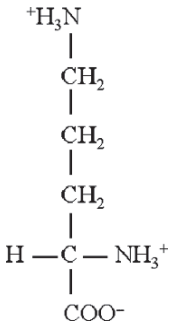


Fig. 5 The structural formula of ornithine

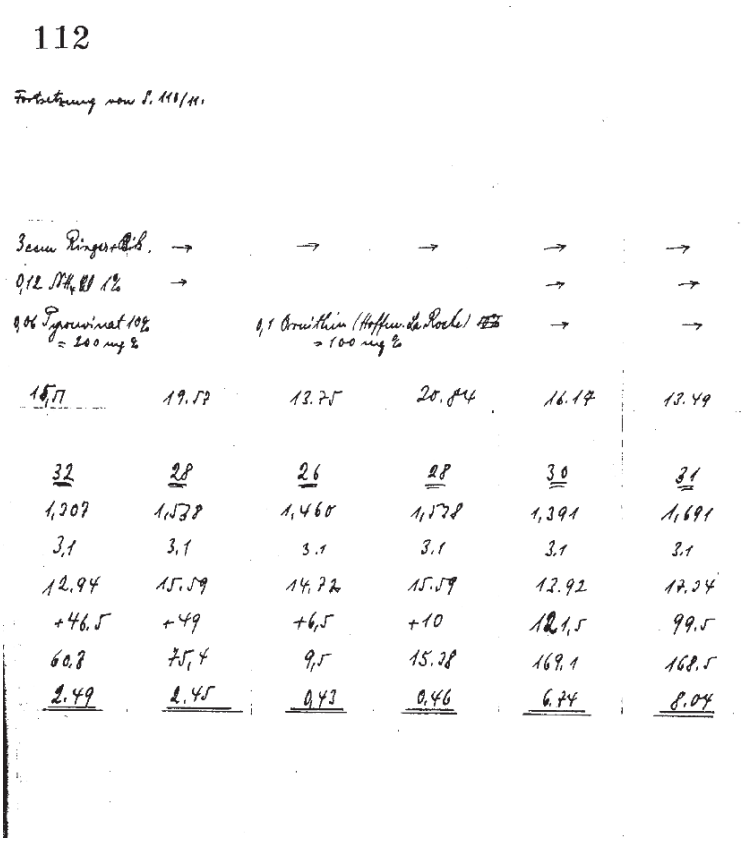


Fig. 6 The ornithine effect: note that the values of the two columns on the far right, numbered 30 and 31, which contain ornithine in combination with ammonia, obtained significantly higher values (6.74 and 8.04) than the others, whereas ornithine alone, in columns 26 and 28, yielded almost no urea at all (0.43 and 0.46)

(see the last two columns in Fig. 6)! Even more puzzling was the observation that the vessels with ornithine and ammonia combined were the only ones that showed this effect: ornithine alone showed almost no urea formation and ammonia alone showed the usual negligible value.

Exploring the Nature of the Effect

In the experiments carried out on 16 November, Krebs and Henseleit were eager to explore the nature of this effect (which, fortunately for them, was reproducible, unlike the first alleged combination effect of ammonia and alanine). To this end, Krebs tried to localize the stimulating aspect of ornithine that caused this effect in combination with ammonia. He first tried to reproduce the effect with valerianate,

which has the same carbon skeleton as ornithine but no nitrogen. Nothing happened; thus Krebs concluded that the effect was not evoked by a certain carbon structure, otherwise valerianate would have yielded results similar to those of ornithine.

During the following two weeks Krebs and Henseleit tested several other substances, namely methylamine, formamide, putrescine and cadaverine; all of them were tested alone as well as in combination with ammonia. At first glance, it is not at all clear what these substances had in common with ornithine and its effect. Holmes supposed, at least in the case of formamide, an approach to the problem that was not connected with ornithine,[37] which would be surprising considering that Krebs had just had his first real breakthrough. On closer investigation, however, each of these substances was either a well-established or at least a speculative decomposition product of ornithine in living animals, as reported in Neubauer's textbook or in the literature of the time.[38] Testing to see whether one of ornithine's decomposition products was able to reproduce the same effect was standard procedure if one wanted to elucidate the nature of a certain effect or reaction: if one of the decomposition products is able to reproduce the effect, one gains valuable clues as to (i) which intermediate reactions occur; and (ii) which part of the original substance is likely to be "efficient". Along the same rationale, Krebs and Henseleit also retested the diamino acid asparagine, presumably to investigate whether the effect was connected to the diamino property of ornithine. Neither the decomposition products nor the asparagine yielded, with or without ammonia, as much urea as ornithine: the effect remained specific and elusive.[39]

Textbooks of the time were unable to explain these findings; the standard hypothesis on urea formation had received a heavy blow, since it was incompatible with Krebs's experimental evidence. Krebs and Henseleit had tried to find out whether any of ornithine's degradation products could be established as an intermediate of the unfamiliar reaction – with no success. They had then experimented to see whether ornithine analogues, such as valerianate or asparagine, could replace the ornithine in the reaction – again with no success. They had considerably varied the experimental set-up, notably by introducing a newly composed saline solution. Again nothing had happened. The only remaining standard approach to find out more about the nature of the effect and the course of the reaction was to establish the quantities of the reaction. Which is what they did on 14 January 1932. To do so, they had to measure how much of the substance added at the start was still present in the reaction vessel after the synthesis of urea.

Quantifications

In 1932 the amount of ammonia present in a solution was usually measured in a Parnas-Heller steam-distillation apparatus. Additionally, the procedure involved determining the concentration of ammonia either colorimetrically or titrimetrically, depending on the quantities produced. In one of his interviews with Holmes, Krebs recalled that he had already intended to measure the ammonia consumption in the course of urea formation at an earlier stage of his research but that he had had to wait until the apparatus was ready.[40] This was the case by 16 December 1931: on this date,

Krebs recorded in his notebook a Parnas-Heller measurement of ammonia in the reaction vessel after having tested urea formation from formamide; this was connected with his suspicions, formulated two days earlier, on 14 December, that formamide decomposed to ammonia.[41] On 6 January 1932 Krebs determined, by the same means, the ammonia that was directly released from asparagine in the presence of urease.[42]

On 14 January Henseleit carried out his first Parnas-Heller measurement to determine how much of the ammonia that was added in combination with ornithine was still present in the solution after urea formation. He found a ratio of 2:1 of moles of consumed ammonia to moles of newly synthesized urea. Thus, for each new molecule of urea (which contains two nitrogen atoms), two molecules of ammonia (which contains one nitrogen atom) had disappeared. Henseleit repeated these measurements in the course of the next few days, and the ratio remained the same. This suggested that *all* of the nitrogen in urea originated from the ammonia and none from the amino acid ornithine – contrary to the standard hypothesis.

This possibility received additional support from the experiment's second outcome. In the same series of 14 January, Henseleit had also varied the concentration of the added ornithine. This, too, was standard procedure in exploring the effect of a substance being tested and had previously been applied by Henseleit to other substances.[43] In the first few weeks of their project, for example, Krebs and Henseleit had varied the concentrations of ammonia, and later they tried out the effect of various amounts of lactate, glucose, fructose, and so forth. Thymine, which Krebs at some point had suspected to be a promising candidate as a urea precursor, had likewise been tested in varying concentrations. Furthermore, testing the effect of different concentrations of ornithine was also a sensible way of at least gaining a rough impression of the extent to which ornithine contributed to the newly synthesized urea.

Henseleit found that "ornithine increases [urea formation] at each concentration but only slightly".[44] Thus, the stimulating effect of the ammonia-ornithine combination remained stable, even when only small amounts of ornithine were added to the vessel – amounts so small that the ornithine could not be acting as a nitrogen donor in the formation of urea. Krebs recalled this experiment in his autobiography as follows:

> In the first experiments which revealed the ornithine effect, the concentration of ornithine had been high because I had set out to explore whether ornithine acted as a nitrogen donor [as suggested by the standard hypothesis; KN & GG]. When I tested lower ornithine concentrations, the stimulating effect remained. The final result of this aspect of the work was the discovery that one molecule of ornithine could bring about the formation of more than twenty molecules of urea, provided that ammonia was present.[45]

A New Role for Ornithine

In contrast to the concentrated series of experiments on urea formation carried out thus far (Henseleit usually performed at least one experimental run every day), experimentation came to a complete stop for the two weeks following the quantification results. Krebs obviously needed time to make sense of the results and to reconcile them to a reaction model.

The following points were to be taken into account: first, the Parnas-Heller measurements of ammonia suggested that the nitrogen in urea originated solely from the ammonia, which would suggest the following mechanism:

$$2\ NH_3 + X \rightarrow \text{Urea [i.e. } CO(NH_2)_2] + Y$$

In this equation X and Y represent unknown factors that had to be included, since urea could not be formed from ammonia alone; as can be taken from its structural formula, at least the carbon and oxygen had to stem from some additional reactant, while two surplus hydrogen atoms from the ammonia had to be accommodated in an additional reaction product besides urea. The simplest way to fill the gaps would have been to complete this scheme by adding carbon dioxide and water in amounts that would balance the stochiometrical requirements, which would have corresponded to a common technique of the time.[46] That Krebs was well aware of this possibility is demonstrated by the fact that he explicitly mentioned the corresponding "summary equation" (*Bilanzgleichung*, as he phrased it) in his third paper on urea biosynthesis (see Fig. 7).

But the formation of urea from ammonia was only part of the story. Krebs's second finding showed that ornithine had to be present for the synthesis of urea from ammonia to occur, and that this amino acid could not be replaced, neither by some structurally related substance, nor by any of its decomposition products. Therefore, ornithine had to be included in the reaction model before it could be balanced stochiometrically:

$$2\ NH_3 + \text{Ornithine} + X \rightarrow \text{Urea [i.e. } CO(NH_2)_2] + Y$$

In order to avoid a process that involved the concurrence of four reactants at the same time – which was virtually impossible, as was known from reaction kinetics – some intermediate step had necessarily to be assumed:

$$2\ NH_3 + \text{Ornithine} + X \rightarrow Y$$

$$Y \rightarrow \text{Urea [i.e. } CO(NH_2)_2] + Z$$

Krebs could now have simply balanced this equation stochiometrically in order to provide at least a summary scheme for the occurring reactions similar to that in Fig. 7, taking into account the additional atoms of, for example, carbon that ornithine added

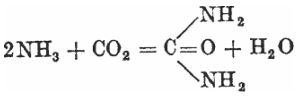


Fig. 7 The simplest summary equation (*Bilanzgleichung*) for urea production, with ammonia as the sole source of nitrogen. Taken from Krebs and Henseleit (1932c), p. 51

to the pool. However, the second of his findings on ornithine prevented him from doing so. Since ornithine stimulated the reaction in very small amounts, its function could not consist, in any substantial way, of contributing to the newly formed urea, not in terms of carbon, oxygen or nitrogen: as Krebs said, under optimal conditions more than twenty molecules of urea were produced per molecule of ornithine added to the ammonia solution (see the passage quoted above). This last observation gave Krebs the decisive clue: although the presence of ornithine was essential for the reaction to occur, very small quantities were sufficient to stimulate the complete conversion of ammonia. This fact could only be explained if one assumed that ornithine was *regenerated* during the course of the reaction, so that it could re-enter the reaction chain. But how could ornithine possibly be regenerated? One pathway that releases ornithine *and* at the same time urea was well known to Krebs and most other biochemists of the time: the arginase reaction. And this is exactly the substance to which Krebs turned his attention in his next experiment: on 26/27 January 1932 he tested different tissues for their level of arginase activity and established that arginase occurred only in liver tissue in considerable quantities. This was the evidence he needed.

Now, only one missing link remained before the circle could be closed: how were ornithine and ammonia actually converted to arginine? Krebs solved this problem by assuming that carbon dioxide and water were the additional reactants, and postulated the following summary equation:

$$\text{Ornithine} + 2\ NH_3 + CO_2 \rightarrow \text{Arginine} + 2\ H_2O$$

$$\text{Arginine} + H_2O \rightarrow \text{Ornithine} + \text{Urea}$$

That ornithine, ammonia and carbon dioxide reacted together to form arginine was at this stage only speculative; moreover, as mentioned before, since Krebs knew that it was highly unlikely that four molecules would react with each other, he had to find an intermediate that would connect the two reactions. As one can gather from the structural formulae of the involved substances, the sought-after substance was fairly well defined, and after an intense and well targeted search of the literature, Krebs found a newly discovered substance that exactly matched the required conditions: an amino acid called citrulline.

Through a series of experiments Krebs was able to demonstrate that citrulline was the circle's missing link. However, although this last building block was necessary in order to convince the scientific community, it played no major part in convincing Krebs: he had already published his concept of a cyclic path some months before finding the last piece of the jigsaw. It was the insight that ornithine acted *like a catalyst,* as Krebs and Henseleit had phrased it in their first paper on the subject, that proved decisive: ornithine's influence was specific and stood in no obvious quantitative proportion to the effect – very small quantities could evoke an enormous increase in the rate of urea formation. These observations firmly established Krebs's notion of a reaction cycle, which was the only solution that could accomodate all his experimental results together with earlier knowledge of the subject (see Fig. 9-11 for the main steps of the cycle and a structural representation of the compounds).

Where is the New Concept?

But is it justified to speak of Krebs's solution, which included discovering a reaction cycle and determining that a substance acted quasi-catalytically, as denoting a new concept? At first glance, neither the notion of a cycle nor that of a catalyst was new in 1930.

Cycles

The concept of cyclic metabolic processes had, in fact, been discussed in biochemistry before – not in connection with urea formation, though, but in relation to the oxidation of carbohydrates. Early experiments on respiration in mammals had established that, in this process, carbohydrates are completely oxidized to carbon dioxide and water. Between 1910 and 1930 it was found that several dicarboxylic acids oxidized as rapidly as glucose: Batelli and Stern, for example, reported in 1910 the oxidation of succinate to form malate.[47] Subsequently, fumarate was shown to be an intermediate of this reaction, while the oxidation of malate to form oxaloacetate was demonstrated, which in turn was shown to release first pyruvate and then acetate.

Already in the early 1920s, the Swedish biochemist Thorsten Thunberg proposed linking this chain of reactions to a cycle by condensing two molecules of acetate to one of succinate, although no evidence could be produced for this hypothesis.[48] In 1930 Toenniessen and Brinkmann proposed an alternative for "closing the circle", as they called it, but their proposition also quickly fizzled out under closer examination.[49] However, the notion of a cyclic reaction path for carbohydrates or at least carboxylic acid oxidation, at times called the Thunberg-Knoop-Wieland hypothesis, was being widely discussed by 1930, and Krebs was almost certainly familiar with it.[50] His achievement was not that he invented a reaction cycle but that he was the first to provide convincing evidence of the occurrence of a reaction cycle; furthermore, with his notion that one of the reactants had a catalytical influence, he defined for the first time the form that empirical evidence for this kind of reaction could take.

Catalysts

Similar to the notion of a cycle, the concept of a catalysis was nothing new and exciting in 1932 but had been proposed a hundred years earlier by the nineteenth-century chemist Jöns J. Berzelius. It was not immediately recognized as a rewarding research topic, though. In the standard biochemical textbook of 1908, the term "catalysis" is not even used, neither in connection with enzymes nor in any other way.[51] By the 1930s, however, catalysis and catalysts had become one of the most

frequently discussed topics of biochemistry, or rather physiological chemistry, as it was then called. Leafing through the volumes of the most important German periodicals in this field – that is, *Hoppe-Seyler's Zeitschrift für physiologische Chemie,* the *Biochemische Zeitschrift,* the *Zeitschrift für Biologie* and the *Ergebnisse der Physiologie* – it is clear that articles on catalysts and catalysis formed one of the discipline's most popular subjects. The Cologne physiologist Bruno Kisch, for example, published a long series of "communications" on catalysts in the context of presenting "o-chinones as model of ferments" (acting catalytically) in the *Biochemische Zeitschrift*, the fourth of which was received by the journal in March 1932, that is, shortly before Krebs and Henseleit published their findings.[52] Kisch had tested the catalytical effect of several organic substances, among others resorcine, hydrochinone and their derivatives, in oxidative amino acid deamination. And in May 1932, to name only one further example, Erik Hägglund and Torsten Johnson from Stockholm had sent an article to the *Biochemische Zeitschrift* – "On the Catalytic Effect of Solid Ligno-Sulphonic Acid" – [53] in which they presented their findings that "solid ligno-sulphonic acid can play an important role as a catalyst of hydrolysis."[54] What a catalyst was thought to be at this time is clearly expressed in a small monograph of 1931, entitled *Homogenous Catalysis*. The relevant definition reads: "A catalyst is a substance which changes the velocity of a reaction's striving for equilibrium, without itself being used up in the reaction."[55]

Several kinds of catalysis were discussed in this booklet. However, by far the most extensive treatment was given to the catalyzing influences of acids and alkalis; some non-hydrolytic cleavages were briefly mentioned as well as some oxido-reductions, such as the oxidation of hydrochinone to chinone. The catalytic influences of, for example, platinum, iron and other heavy metals were also touched upon, and for the topic of enzymatic catalysis the authors referred to another volume in the series. This range represented the basic spectrum of reactions and reactants involved in catalyses around 1930. On the reaction mechanism, it was generally believed that a catalyst formed an intermediate with its substrate – which could only rarely be determined, though – and by this means reduced the activation energy required for the reaction to occur.[56] However, it was not known how this happened in detail. It was assumed, for example, that, catalytically acting acids temporarily formed a salt with the substrate and accelerated the reaction in question by this means.[57] Whether enzymes acted in the same way was, however, unclear.

The kind of catalyst which Krebs was probably most familiar with was the one that Warburg had applied to his research into explaining cell respiration as the effect of "iron catalysis": Warburg had found that organic substances were not directly oxidized by molecular oxygen in the animal body but only catalytically mediated by iron, which changes from the +II to the +III state and transfers the temporarily fixated oxygen to organic substances (see Fig. 8).[58] In his Berlin days, Krebs had contributed to this research programme of Warburg's and had even corrected the proofs to Warburg's first comprehensive collection of papers on the subject.[59]

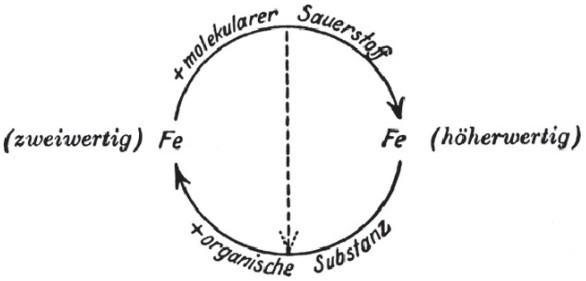


Fig. 8 Warburg's assumption of the different oxidation states of iron in organic compounds, on iron acting as a catalyst. Taken from Warburg (1924)

And Yet It Was New!

So in hindsight Krebs's proposal may not appear as very new and revolutionary; yet his contemporaries thought otherwise, as was made clear by the remarks of the Royal Society's President in 1932. It is one thing to speculate on the occurrence of a metabolic cycle, as Thunberg and Toenniessen had done; many people speculated on the most far-fetched biochemical possibilities, and few speculations could be ruled out due to methodological limits. It is quite another thing to provide conclusive evidence for the occurrence of a cyclic path, which is what Krebs's achievement comes down to. However, what was new was the discovery that an amino acid takes part in the reaction chain in the manner of a catalyst, the evidence of which proved so convincing that Krebs's proposal received few challenges after it had been published.[60]

Incidentally, ornithine's unusual role in the reaction was also the one element that Krebs stressed in his first presentation of the findings before the Freiburg Medical Society on 10 May 1932: Krebs did not dwell primarily on the fact that he had elucidated the first metabolic cycle but rather that ornithine was the first known case of an amino acid to have an additional function in the manner of a catalyst.[61] Likewise, in his Nobel Lecture of 1953 Krebs described the seminal role of finding substances that act along the same lines as ornithine for establishing the citric acid cycle.[62] The consequences of his 1932 discovery were far-reaching. No one had ever imagined that a well known amino acid could act in this way, and so been able to conclude the occurrence of a metabolic cycle – but many did so after him, not least Albert Szent-György and the two researchers, Stare and Baumann, who contributed essential findings to the establishment of the citric acid cycle.[63]

That Krebs was aware that he used the catalyst concept in a new and diverging way can be concluded from his formulation of the idea: Krebs never maintained that ornithine acted "as" a catalyst but merely "like" or "in the manner of" a catalyst. Although this may seem to be nit-picking at first glance, one can interpret this unusual formulation as evidence that Krebs had been focusing on a specific concept of catalysis and that the behaviour of ornithine did not match this concept in every aspect: similar to a catalyst, ornithine is regenerated during the course of the reaction, which enables it to promote the reaction even in very small quantities.

However, the molecules themselves nevertheless enter the actual reaction path – with the exception of the iron in Warburg's catalyst, which only transfers oxygen by being temporarily oxidized and is quickly reduced again. That Krebs's notion of a catalyst was a special one is furthermore reflected in his sharp differentiation between the way ornithine and citrulline act:

> The velocity of urea synthesis in the liver is greatly increased by ornithine. Ornithine acts thereby in the manner of a catalyst in that it is not used up in its action and even in small quantities can provoke high rates. This effect is peculiar to ornithine. None of the many other substances tested can replace ornithine. Citrulline accelerates the urea formation from ammonia but citrulline does not act in the manner of a catalyst. It is used up during the reaction in that it contributes one atom of nitrogen to the urea.[64]

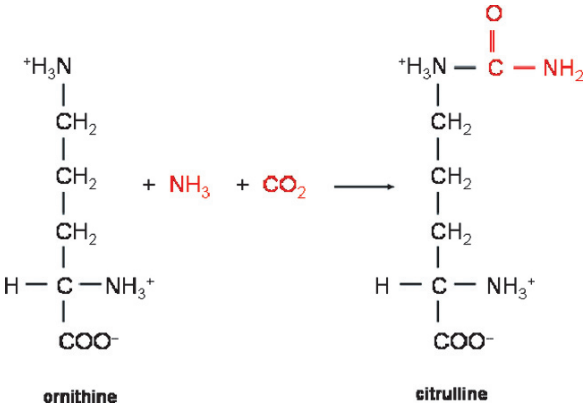


Fig. 9 The first step in the cycle: ornithine converts to citrulline on the absorption of ammonia and carbon dioxide. Note that no part of the structure of ornithine changes

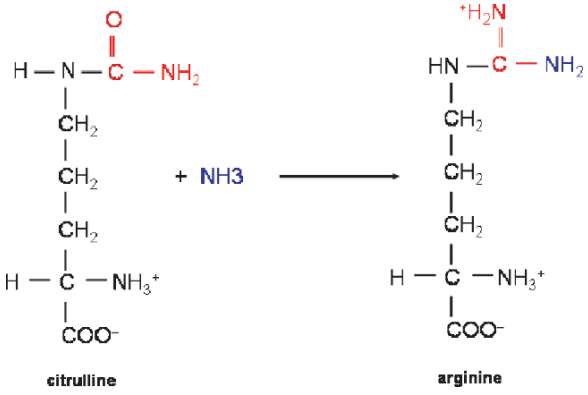


Fig. 10 Second step: citrulline converts to arginine by absorbing another molecule of ammonia. Note that the oxygen on top of the structure is replaced by an amino group

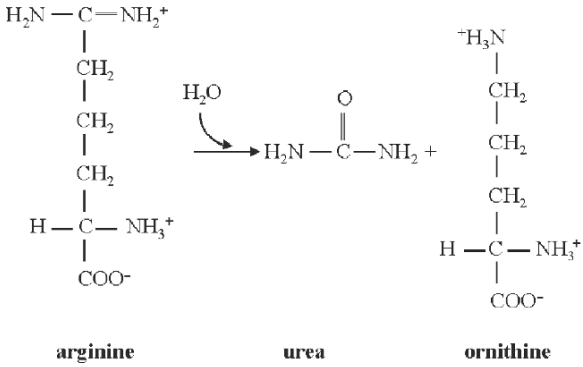


Fig. 11 Third step: arginine releases urea and regenerates ornithine, which involves a considerable change of structure

Concepts from the Bench

Thus Krebs's achievement was neither the invention of a cycle (not even a metabolic reaction cycle) nor the invention of a catalyst; rather, it lay in his discovery of an unusual combination of the two, applied to new phenomena, together with his ability to provide evidence for this new conception. Intuitively, we would have imagined the story differently, which is typical of a discovery such as this: a successful new concept often changes our view on the subject matter so much that, combined with a new terminology, it becomes extremely difficult to put oneself in the position of someone before the event. This becomes even more difficult if one takes into account that, in hindsight, one does not necessarily consider the findings to be as revolutionary as the historical players did. It was not so much the circle that Krebs's contemporaries found to be a novelty but rather his discovery of the catalytical role of ornithine.[65]

What deserves particular attention is how Krebs arrived at his hypothesis. By scrutinizing the notebooks, one can reconstruct, step by step, the path Krebs took – from the standard hypothesis to his unexpected solution. Seen from today's perspective, most of the time Krebs was going amiss, although this would not have been obvious to him at the time; from his point of view, his research strategy – that is, the following of standard procedures and assumptions for as long as possible – was good experimental practice. However, it led him nowhere. At a crucial stage in the project – in January 1932 – Krebs found himself caught amid confusing results, which completely contradicted his expectations. His great achievement was to arrive at a creative new solution from all this confusion– it almost seems as if he had to reach the apparent dead end in order to find a new way out. He had established that ammonia was the only source of nitrogen for urea and he had seen that ornithine stimulated the reaction at different concentration levels, and after two weeks of much thinking – in which Krebs ceased to carry out any further experiments – the penny dropped. Thereafter, Krebs only had to corroborate his hypothesis. He reconfirmed that arginase mainly occurred and acted in liver tissue, repeated the concentration experiments and explored some further aspects of the reactions; he

also established citrulline to be an intermediate. But these additional findings did not substantially influence his solution.

In a way it is surprising how apparently small Krebs's own contribution to the discovery of the urea cycle actually was. Almost all the building blocks lay readily at hand. The arginase reaction was well known, while the subjects of reaction cycles and catalysts were also already being discussed. Franz Knoop seems to have felt much the same when he wrote the following lines after having received Krebs's third paper for publication:

> Dear Doctor Krebs!
> I thank you very much for your beautiful paper. It is very convincing, how the way in which you come finally to the synthesis of urea by way of arginine and thereby to this significant role of ornithine, which follows entirely by itself out of the sequence of your investigations. At first I was extraordinarily surprised when Thannhauser reported these facts to me; but after the event, one can almost imagine that he could already have arrived at such a conception at the writing table.[66]

In a sense, Knoop was right. Anyone could have arrived at the cyclic solution which followed "entirely by itself" from Krebs's stringent investigation – although only Krebs did. His achievement consisted of recognizing the right elements for accommodating his experimental findings and putting them together in a new way. His concept of a reaction cycle was not invented by noticing an apple dropping in front of his eyes or by him having dreamed of substances going round in circles. It was a solution that was constructively shaped in the laboratory, indeed almost compelled by the outcome of a series of well-designed, methodologically sophisticated experiments, combined with a knowledgeable use of the literature and clever thinking. And, although for much of the time his experiments led him nowhere – indeed, maybe even because he found himself at a dead end – Krebs did finally get it right.

Notes

1. Hopkins (1933), p. 175.
2. For the most comprehensive description of Hans Krebs's life, see the seminal, two-volumed work by Frederic L. Holmes (1991, 1993). For Krebs's autobiography, see Krebs (1981).
3. "Untersuchungen über die Harnstoffbildung im Tierkörper". Krebs and Henseleit published this series of papers in the space of four months: Krebs and Henseleit (1932a, b, c). The third paper gives the most extensive account of their findings and method.
4. "Sie haben Grosses vollbracht!" Cited Holmes (1991), p. 354.
5. See Holmes (1991), p. 395. Neuberg's proposal was, however, unsuccessful. Two months later the Nazi Party came to power and academics of Jewish origin, such as Krebs, found themselves dismissed from their state posts. He left Germany in the summer of 1933 for Britain.
6. See Krebs (1981), p. 48.
7. Krebs (1981), p. 40. See also Krebs's first draft in German, transcribed in Werner (1991), doc. 46, p. 137f.
8. For a transcript of Warburg's letter to Krebs, see Werner (1991), doc. 105, p. 283f.
9. See Krebs (1981), p. 40.

10. Krebs and Henseleit (1932a), p. 759: "Nach diesen Versuchen wirkt Ornithin bei der Harnstoffsynthese in der lebenden Zelle wie ein Katalysator."
11. Krebs and Henseleit (1932c), p. 51.
12. Gözsy and Szent-Györgyi (1934), p. 9.
13. See, e.g., Szent-Györgyi et al. (1935), Szent-Györgyi et al. (1936), Szent-Györgyi (1937) and Stare and Baumann (1937).
14. See Krebs (1964).
15. Warburg's own Kaiser Wilhelm Institute for Cell Physiology was not founded until 1931. See, e.g., the relevant documents in Werner (1991), in particular p. 233ff.
16. See Krebs (1928). The standard manual for manometric methods of the 1960s, Kleinzeller (1965), still refers, besides Warburg, to Krebs (and his article) as one of the main figures in developing this methodological approach.
17. Four different reconstructions of various aspects of the story have so far been published: (i) the account by Hans Krebs himself, as in Krebs (1973, 1976, 1981); (ii) the reconstruction of Krebs's experimental pathway by Frederic L. Holmes in Holmes (1980, 1991); (iii) a computational model proposed by Deepak Kulkarni and Herbert Simon, see Kulkarni and Simon (1988); and (iv) the account developed by Gerd Graßhoff and various collaborators entailing an alternative computational model, see Graßhoff (1995), Graßhoff and May (1995) and Graßhoff et al. (2000).
18. The two respective notebooks were recently published in facsimile with a transcription and a synoptic outline of all the conducted experiments, their parameters and results in Graßhoff and Nickelsen (2001a, b). On the value of these laboratory notebooks for elucidating the urea cycle's pathway of discovery, see Graßhoff and May (2003).
19. For Krebs's own retrospective accounts, see the references cited above. The interviews with Holmes were an important source for his two-volumed monograph on the biochemist.
20. On the pitfalls of these historical fallacies, see Graßhoff (1994) and Graßhoff et al. (2000), p. 324ff.
21. Krebs (1981), p. 52.
22. For some of the most important papers, see, e.g., Schultzen and Nencki (1872), Knierim (1874) and Schroeder (1882). Historical surveys of the elucidation of the urea cycle can be found in, e.g., Holmes (1980), p. 216f. and Holmes (1991), p. 249ff., and in general histories of biochemistry, e.g., in Fruton (1999), pp. 352–357, Ord and Stocken (1995), pp. 105–109, etc.
23. Neubauer (1928), p. 808: "Durch die Festlegung der Aminosäuren als regelmässige Zwischenprodukte beim Eiweissabbau und den Nachweis ihrer Desaminierung ist das Problem der Harnstoffbildung insofern vereinfacht worden, als es jetzt in der Hauptsache so formuliert werden kann: *Wie entsteht der Harnstoff aus dem von den Aminosäuren abgespaltenen Ammoniak?"* (Emphasis inserted in the original German text.) Further discussions of potential pathways for urea can be found in the following pages of the same work.
24. This and the following equations are our reconstructions, which we arrived at by reformulating and re-interpreting the accounts of the historical players.
25. This rule had originally been proposed by Arthur Slator in relation to the study of the fermentation of alcohol Slator (1907), but was soon generally accepted as a guiding principle for elucidating metabolic processes. See also Holmes (1991), p. 10f.
26. Contrary to standard procedure, Krebs successfully adapted the method of manometrically measuring urea (by means of the enzyme, urease): rather than working in the usual acidic medium, he used an alkaline solution, which overcame the problem of carbon dioxide retention that had until then caused so many difficulties. For his own explanation of this phenomenon, see Krebs (1928), p. 1061.
27. See, e.g., the interpretation by Holmes (1980), p. 217.
28. In the course of their study of the urea cycle, Krebs considerably refined his technique. Notably he developed, e.g., a new saline medium for the tissue slices, which proved extremely successful and stayed in use for an exceptionally long time. It is probably due to the recipe of this solution that Krebs's paper on the urea cycle was, until very recently, so highly cited: it gained

more than 2,400 explicit citations between 1961 and 1981, which is remarkable considering its publication date of 1932. See, e.g., Garfield (1982).
29. Kossel and Dakin (1904).
30. See the discussion in Neubauer (1928) p. 809f.
31. See particularly Löffler (1917, 1920).
32. Krebs was still differentiating between the two mechanisms in Krebs and Henseleit (1932a), p. 758.
33. Cited Holmes (1991), p. 280.
34. Neubauer (1928), p. 776.
35. The unit of measurement Krebs and Henseleit used to determine the amount of urea produced was the "urea quotient" (*Harnstoffquotient*), which they abbreviated to Q_H.
36. This important change of the experimental set-up was introduced after Krebs had lost a month of valuable time unsuccessfully trying to reproduce another (alleged!) urea-producing effect of thymine, which was probably an artefact that had occurred as a result of a contaminated saline solution.
37. Holmes (1991), p. 288.
38. Detailed evidence for the hypothesis that Krebs knew the literature discussing the respective degradation paths is provided by Graßhoff and May (2003), p. 283ff.
39. Apart from testing the decomposition products of ornithine mentioned earlier, Krebs and Henseleit explored some circumstantial conditions of the ornithine effect: whether it occurred in other tissues, whether glucose had an additional influence, etc.; these tests resemble the experiments of the first phase of the research programme when they were exploring the conditions of the general experimental set-up. Only on 8 January 1932, when Krebs and Henseleit had already applied their whole repertoire of parameters with the ornithine effect without obtaining any instructive results, did they go back and include some new alternative nitrogen donors in their series of tests, namely histidine and oxyproline, tyrosine and leucine.
40. See Holmes (1980), p. 220.
41. See the corresponding record in Krebs's laboratory notebook of 14 December 1931: "It remains to be investigated whether formamide is decomposed to NH_3."
42. See Henseleit's laboratory notebook of 6 January 1932: "NH_3 cleavage from asparagine"; the day before he had concluded his asparagine experiments with the comment that one should investigate whether it decomposes in liver tissue.
43. Checking the literature of the time, in particular those articles that presented results obtained by manometric methods, investigating the influences of the following factors on the reaction under study were standard: temperature, medium, pH, alternative tissues, concentration of the reactants, oxygen percentage in the atmosphere, duration of experiment, prohibiting or stimulating effects of additional substances (e.g., the influence of hydrocyanic acid, which was known to inhibit cell respiration, was regularly tested). See, for some early examples of establishing the method, Battelli and Stern (1911a, b); presumably Krebs was particularly influenced by the similarly standard procedures applied by Warburg, as documented in the articles published in Warburg (1928).
44. "Ornithin steigert in jeder Konzentration, aber nur gering." For a facsimile of the respective page, see Graßhoff and Nickelsen (2001b), p. 162.
45. Krebs (1981), p. 56.
46. For an analysis of the function of chemical formulae in construction reaction models, see Klein (2003).
47. Batelli and Stern (1911).
48. Thunberg (1920).
49. Toenniessen and Brinkmann (1930).
50. See, e.g., the discussion of the (hypothetical) circle proposed by Thunberg in Neubauer (1928), p. 883.
51. See Röhmann (1908).
52. See Kisch (1932).
53. Hägglund and Johnson (1932).

54. Hägglund and Johnson (1932), p. 325.
55. Von Euler and Ölander (1931), p. 9: "Ein Katalysator ist ein Stoff, welcher, ohne selbst durch die Reaktion verbraucht zu werden, die Geschwindigkeit ändert, mit welcher eine Reaktion ihrem Gleichgewicht zustrebt."
56. See von Euler and Ölander (1931), p. 14ff.
57. See von Euler and Ölander (1931), p. 54.
58. See, as milestones among many other publications on similar topics, Warburg (1914, 1924). Warburg (1928) provides a collection of several papers by Warburg and his collaborators on the catalytic effects of living matter.
59. Warburg (1928); see Warburg's acknowledgment of Krebs's collaboration in the preface.
60. In 1934, a critical note was published by a Russian team of researchers, stating that the ornithine effect could not be reproduced in living animals. Krebs, however, quickly refuted this criticism on the grounds of methodological faults. See London et al. (1934) for the critique and Krebs (1934) for Krebs's reply.
61. See Holmes (1991), p. 351.
62. See Krebs (1964), p. 400f.
63. See, among others, Szent-Györgyi's collection of earlier articles (1937) and Stare and Baumann (1937).
64. Krebs and Henseleit (1932c), pp. 65–66: "Die Geschwindigkeit der Harnstoffsynthese in der Leber wird durch Ornithin stark gesteigert. Ornithin wirkt dabei nach Art eines Katalysators, indem es sich bei seiner Wirkung nicht verbraucht und schon in Spuren grosse Umsätze bewirken kann. Diese Wirkung ist nur dem Ornithin eigentümlich. Keiner von vielen anderen untersuchten Stoffen kann Ornithin ersetzen. Citrullin beschleunigt zwar die Harnstoffbildung aus Ammoniak, doch wirkt Citrullin dabei nicht nach Art eines Katalysators. Es verbraucht sich selbst bei seiner Wirkung, indem es ein Stickstoffatom für den Harnstoff liefert."
65. See the description of his discovery in contemporary reviews: e.g., the already mentioned speech by Hopkins (1933), the description in Schneller (1935) and the introduction of urea synthesis in Lehnartz (1938).
66. Cited Holmes (1991), p. 354.

References

Batelli, F. and Stern, L. S.: 1911, Oxydation der Bernsteinsäure durch Tiergewebe, *Biochemische Zeitschrift* **30,** 172–178.

Battelli, F. and Stern, L.: 1911a, Die Oxydation der Bernsteinsäure durch Tiergewebe, *Biochemische Zeitschrift* **30,** 172.

Battelli, F. and Stern, L.: 1911b, Die Oxydation der Citronen-, Apfel- und Fumarsäure durch Tiergewebe, *Biochemische Zeitschrift* **30,** 478–505.

Fruton, J. S.: 1999, *Proteins, Enzymes, Genes. The Interplay of Chemistry and Biology,* Yale University Press, New Haven/London.

Garfield, E.: 1982, Current Comments: To Remember Sir Hans Krebs: Nobelist, Friend, and Adviser, *Current Contents* **31,** 5–11.

Graßhoff, G.: 1994, The Historical Basis of Scientific Discovery Processes, *Behavioral and Brain Sciences* **17,** 545ff.

Graßhoff, G.: 1995, Die Kunst wissenschaftlichen Entdeckens. Grundzüge einer Theorie epistemischer Systeme, Habilitationsschrift, Universität Hamburg.

Graßhoff, G., Casties, R. and Nickelsen, K.: 2000, *Zur Theorie des Experiments. Untersuchungen am Beispiel der Entdeckung des Harnstoffzyklus* (Educational Materials; 2), Bern Studies in the History and Philosophy of Science, Bern.

Graßhoff, G. and May, M.: 1995, Methodische Analyse wissenschaftlichen Entdeckens, *Kognitionswissenschaft* **5,** 51–67.

Graßhoff, G. and May, M.: 2003, Hans Krebs's and Kurt Henseleit's Laboratory Notebooks and Their Discovery of the Urea Cycle: Reconstructed with Computer Models, *in* F. L. Holmes, J. Renn and H.-J. Rheinberger (eds.), *Reworking the Bench. Research Notebooks in the History of Science,* Kluwer, New York.

Graßhoff, G. and Nickelsen, K.: 2001a, *Dokumente zur Entdeckung des Harnstoffzyklus. Vol. 1: Laborbuch Hans Krebs und Erstpublikationen* (Sources; 1), Bern Studies in the History and Philosophy of Science, Bern.

Graßhoff, G. and Nickelsen, K.: 2001b, *Dokumente zur Entdeckung des Harnstoffzyklus. Vol. 2: Laborbuch Kurt Henseleit* (Sources; 2), Bern Studies in the History and Philosophy of Science, Bern.

Gözsy, B. and Szent-Györgyi, A.: 1934, Über den Mechanismus der Hauptatmung des Taubenbrustmuskels, *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **224,** 1–10.

Hägglund, E. and Johnson, T.: 1932, Über den katalytischen Effekt der festen Lignosulfonsäure, *Biochemische Zeitschrift* **250,** 321–325.

Holmes, F. L.: 1980, Hans Krebs and the Discovery of the Ornithine Cycle, *Aspects of the History of Biochemistry, Federation Proceedings* **39(2),** 216–224.

Holmes, F. L.: 1991, *Hans Krebs: The Formation of a Scientific Life,* Vol. 1, Oxford University Press, Oxford.

Holmes, F. L.: 1993, *Hans Krebs: Architect of Intermediary Metabolism,* Vol. 2, Oxford University Press, Oxford.

Hopkins, F. G.: 1933, Address of the President, Sir Frederick Gowland Hopkins, at the Anniversary Meeting, November 30, 1932, *Proceedings of the Royal Society of London. Series B.* **112,** 159–184.

Kisch, B.: 1932, o-Chinone als Fermentmodell. IV. Mitteilung: Bei pH 6 bis 8 unwirksame Katalysatoren, *Biochemische Zeitschrift* **249,** 63–71.

Klein, U.: 2003, *Experiments, Models, Paper Tools: Cultures of Organic Chemistry in the Nineteenth Century,* Stanford University Press, Stanford.

Kleinzeller, A.: 1965, *Manometrische Methoden und ihre Anwendung in der Biologie und Biochemie,* Gustav Fischer, Jena.

Knierim, W. v.: 1874, Beiträge zur Kenntnis der Bildung des Harnstoffs im Thierischen Organismus, *Zeitschrift für Biologie* **10,** 263–294.

Kossel, A. and Dakin, H. D.: 1904, Über die Arginase, *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **41,** 321–331.

Krebs, H.: 1928, Stoffwechsel der Zellen und Gewebe, *in* O. Arnbeck (ed.), *Methodik der wissenschaftlichen Biologie. Bd. II: Allgemeine Physiologie,* Julius Springer, Berlin, 1048–1084.

Krebs, H..: 1934, Bemerkungen zu einer Arbeit von E. S. London, A. K. Alexandry und S. W. Nedswedski: Über die Harnstoffbildung in der Leber', *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **230,** 278–279.

Krebs, H.: 1964, The Citric Acid Cycle. Nobel Lecture, December 11, 1953, *Nobel lectures. Including presentation speeches and laureates' biographies (Physiology or Medicine),* World Scientific, Singapore amongst others, 399–410.

Krebs, H.: 1973, The Discovery of the Ornithine Cycle of Urea Synthesis, *Biochemical Education* **1,** 19–23.

Krebs, H.: 1976, The Discovery of the Ornithine Cycle, *The Urea Cycle,* John Wiley, New York.

Krebs, H.: 1981, *Reminiscences and Reflections,* Clarendon Press, Oxford.

Krebs, H. and Henseleit, K.: 1932a, Untersuchungen über die Harnstoffbildung im Tierkörper I, *Klinische Wochenschrift* **11(18),** 757–759.

Krebs, H. and Henseleit, K.: 1932b, Untersuchungen über die Harnstoffbildung im Tierkörper II, *Klinische Wochenschrift* **11(27),** 1137–1139.

Krebs, H. and Henseleit, K.: 1932c, Untersuchungen über die Harnstoffbildung im Tierkörper, *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **210,** 33–66.

Kulkarni, D. and Simon, H. A.: 1988, The Process of Scientific Discovery: The Strategy of Experimentation, *Cognitive Science* **12,** 139–175.

Lehnartz, E.: 1938, *Einführung in die chemische Physiologie,* Springer, Berlin.

Löffler, W.: 1917, Desaminierung und Harnstoffbildung im Tierkörper, *Biochemische Zeitschrift* **85,** 230–294.
Löffler, W.: 1920, Zur Kenntnis der Leberfunktion unter experimentell pathologischen Bedingungen, *Biochemische Zeitschrift* **112,** 164–187.
London, E. S., Alexandry, A. K. and Nedswedski, S. W.: 1934, Die Frage der Beteiligung des Ornithins, des Citrullins und des Arginins am normalen Prozeß der Harnstoffbildung in der Leber unter Anwendung der Angiostomie, *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **227,** 233–241.
Neubauer, O.: 1928, *Intermediärer Eiweißstoffwechsel,* Julius Springer, Berlin, S. 797–825.
Ord, M. G. and Stocken, L. A.: 1995, *Early Adventures in Biochemistry,* Jai Press, London.
Röhmann, F.: 1908, *Biochemie. Ein Lehrbuch für Mediziner, Zoologen und Botaniker,* Springer, Berlin.
Schneller, H.: 1935, Physiologische Bedeutung des Ammoniaks im Organismus der Wirbeltiere, *Ergebnisse der ogie, Biologischen Chemie und Experimentellen Pharmakologie* **37,** 492–528.
Schroeder, W. v.: 1882, Über die Bildungsstätte des Harnstoffs, *Archiv für experimentelle Pathologie und Pharmakologie* **15,** 364–402.
Schultzen, O. and Nencki, M.: 1872, Die Vorstufen des Harnstoffs im thierischen Organismus, *Zeitschrift für Biologie* **8,** 124–146.
Slator, A.: 1907, Über Zwischenprodukte der alkoholischen Gärung, *Ber. chem. Ges.* **40,** 123.
Stare, F. J. and Baumann, C. A.: 1937, The effect of fumarate on respiration, *Proceedings of the Royal Society of London. Series B.* **121,** 338–357.
Szent-Györgyi, A.: 1937, *Studies in Biological Oxidation and Some of Its Catalysts,* Eggenberg & Barth, Budapest.
Szent-Györgyi, A., Annau, E., Banga, I., Blazsó, A., Bruckner, V., Laki, K. and Straub, F. B.: 1936, Über die Bedeutung der Fumarsäure für die tierische Gewebsatmung. III. Mitteilung, *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **244,** 105–152.
Szent-Györgyi, A., Annau, E., Banga, I., Gözsy, B., Huszák, S., Laki, K. and Straub, B.: 1935, Über die Bedeutung der Fumarsäure für die tierische Gewebsatmung, *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **236,** 1–68.
Thunberg, T.: 1920, Zur Kenntnis des intermediären Stoffwechsel und dabei wirksame Enzyme, *Skandinavisches Archiv für Physiologie* **40,** 1–91.
Toenniessen, E. and Brinkmann, E.: 1930, Über den oxydativen Abbau der Kohlenhydrate im Säugetiermuskel, insbesondere über die Bildung von Bernsteinsäure und Brenztraubensäure, *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **187,** 137–159.
Von Euler, H. and Ölander, A.: 1931, *Homogene Katalyse. Teil I: Nicht-enzymatische Katalysen. Teil II: Enzymatische Katalysen,* (Sammlung Göschen; 1037 u. 1038), de Gruyter, Berlin.
Warburg, O.: 1914, Über die Rolle des Eisens in der Atmung des Seeigeleis nebst Bemerkungen über einige durch Eisen beschleunigte Oxydationen, *Zeitschrift für physiologische Chemie* **92,** 231–256.
Warburg, O.: 1924, Über Eisen, den sauerstoffübertragenden Bestandteil des Atmungsferments, *Biochemische Zeitschrift* **152,** 479–494.
Warburg, O.: 1928, *Über die katalytischen Wirkungen der lebendigen Substanz. Arbeiten aus dem Kaiser-Wilhelm-Institut für Biologie, Berlin-Dahlem,* Springer, Berlin.

How Experiments Make Concepts Fail: Faraday and Magnetic Curves

Friedrich Steinle

Can concepts "go amiss" or "fail"? To put the question differently (and perhaps more appropriately): what could it possibly mean for concepts to fail? And what significance could the possible failure of concepts bear for the practice of research? It is those questions on which my paper focuses. I shall approach them both from general considerations and from a specific case of experimental research.

Concepts and Failure

The talk of concepts has been notoriously unsharp and disputed. But despite ongoing debates about the exact notion of concepts,[1] there are some more or less undisputed core aspects, i.e. a shared basic understanding that is sufficient for my purpose. Concepts are understood as *elements* of our thoughts, rather than extended *systems* as theories are. Moreover, and this is a sort of corollary, it is crucial that concepts are usually just used and taken for granted, but not discussed and reflected upon. They are normally implicit, whereas theories are normally explicit. Of course, it is exactly in periods of concept formation that concepts become explicitly discussed. But these processes, I claim, have a different character from those of discussion of theories.

In discussing scientific concepts in particular, one is directed to an epistemic level that is more fundamental than the level of theories. It is concepts and conceptual frameworks by which we structure our world in things, entities, categories, facts and so on; a structure that we necessarily presuppose and usually just take for granted when debating about theories and, highly significant, even when formulating individual empirical statements and experimental outcomes. In experimental research, we have to take whole sets of concepts and larger conceptual structures

F. Steinle (✉)
Fachbereich A, Wissenschafts- und Technikge-Schichte, Gaussstr. 20, 42119 Wuppertal; University of Wuppertal, Germany

I thank all participants of the workshop for discussion, and Jutta Schickore and Giora Hon for careful reading of earlier drafts and intriguing questions.

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_8,

as granted in order to frame our questions, design experiments, express observations, and draw our lessons from the outcome of experiments. For illustration, one may think of basic concepts such as chemical reaction, catalysis, mechanical force, gravitation, evolution, struggle for existence, cells or metabolism.

History tells us that concepts have been created, often enough reshaped, and in some cases even gone away – think of "action at a distance," "phlogiston," "absolute space" "biological species", "microsome", or "particle." Concepts shift in time, and some are now regarded as clear failures. However, while it seems to be obvious that there is something like failure on conceptual level, it is difficult to grasp what that exactly means – difficult both for researchers and for philosophers and historians. Processes of conceptual change have not been well understood so far (at least in the history of the physical sciences), and this holds even more for cases in which changes of such a type took place in the course of experimental activity. In those cases we find an epistemically precarious situation: concepts necessarily direct our experimental acting, our framing of questions, and attempting at answers. How then is it possible, in the course of this action, to revise its very conceptual basis, i.e. to open it to instability and change?

Concepts and categories cannot be said to be wrong or right, as propositions or theories can. We can argue whether such a thing as "species" exists, whether it's better defined by one than another criterion, whether "cross-species breeding" is a contradiction in itself etc, and propositions in these matters can be said wrong or right. But can we say the concept of species is wrong? Similar questions come up when talking about particles, genes, infection disease or, to switch to physics, force, velocity, polarity, and many others. The basic idea I shall elaborate in this paper is this: I suggest that rather from talking of concepts being wrong or right, we should regard them as being appropriate or inappropriate. This clearly and explicitly brings the question of goals into the horizon: "appropriate" always means: "appropriate for a certain end." Scientific concepts are purposefully introduced to order things, to classify them, to express experimental outcomes, to stabilize and speak about empirical "fact" and to theorize about them. But every order has its purpose, and the concepts introduced are framed such as to serve this purpose. A pharmacist will classify flowers and plants by concepts that are different from those used by a botanist, a chemist, a druggist, or an apothecary. A mechanic will conceptualize mechanical action different from a mathematical physicist. Anthropologists may use different concepts for classifying colours than physiologists, physicists, colour metricians, painters or dyers. In all those cases, the specific concepts chosen will serve their purpose better or worse. While they cannot be said being right or false, it is well possible to call them appropriate or not, i.e. find them serving better or worse the purpose they were intended to fulfil. This can go so far that some concepts may be regarded as "failed", i.e. as useless for the purpose they were intended to serve. Whether we keep or drop certain concepts may essentially depend on our goals, and we may even keep a specific concept for one goal while dropping it for another – a situation not unobserved in science (as well as in everyday life), think again of the concept of species or particle.

Focusing specifically to experimental research, one may ask what the ends of experimental research are? What work do we want our concepts to do in experimental science? Experimental research may pursue various different purposes: creating new effects, formulating empirical regularities, framing a mathematical expression for the field in question, fitting the field into an existing theoretical and explanatory context, rendering the field accessible to technical manipulation, and probably even other goals. While these purposes may (and often do) partially overlap, they are still different, and often make different concepts appear preferable.

It is worth to note, however, that the criteria in this field are usually neither clearcut nor explicit. This well may have a systematic reason: it is less easy to take the reflective distance to our concepts, to our language, than to explicit propositions and theories. This might be one reason why there are only few explicit discussions of the topic, neither in science nor in history and philosophy of science. In order to come closer to what conceptual failure can mean, we have to take a look on how it occurs and is treated in the practice of science. In the next sections, I shall discuss in some detail a specific case of experimental research, and afterwards come back to my more general questions.

Faraday, Electromagnetic Induction, and Magnetic Curves

My case deals with Michael Faraday's analysis of electromagnetic induction, i.e. an action of magnetism onto electricity. Such an effect had been sought ever since the Danish academic Hans-Christian Oersted, in 1820, had published his discovery of an action of electricity onto a magnetic needle. The question of an inverse action had been treated again and again, but without success.[2] Faraday, initially assistant, later on director of the laboratory of the Royal Institution of London, had followed these attempts with interest, but only in 1831 he was free to really choose his own research agenda.[3] He took up electromagnetism and in particular the search for the induction effect.

The Unexpected Features of the Induction Effect

His experimental arrangement of August 1831 (Fig. 1) consisted of an iron ring with two separate sets of coils (A and B) wound on it. One set was connected with a wire running over a magnetic needle, the other with a battery. At the moment when the connection to the battery was made, or broken, the needle was affected. Since the two coils were not connected, this was clearly an effect of electromagnetic induction.

I shall not deal with the pathway that had led Faraday to designing his particular arrangement – indeed we do not know too much about this point – but rather with how he evaluated and tried to make sense of the experimental result. While an induction effect in general had well been expected, the experimental result Faraday had obtained now provided some unexpected features, indeed a

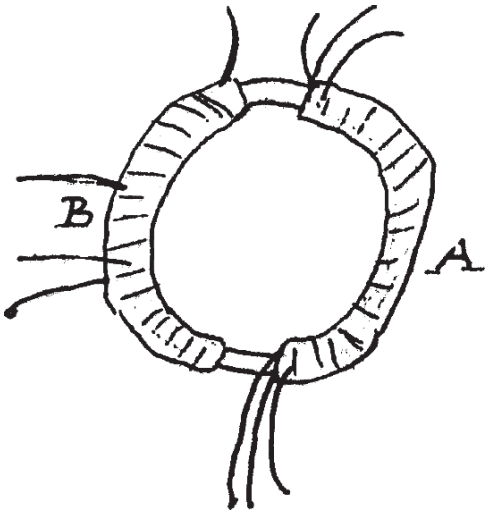


Fig. 1 Faraday's sketch of his induction ring (Diary, D2)

striking deviation from basic features that everyone would have expected: The most puzzling feature was that the effect occurred only at the moment the inducing current was switched on or off. By contrast, and in the "common condition" of the wire, as Faraday put it,[4] he would have expected the effect occurring as long as its cause, i.e. the inducing current, was present. The experimental outcome stood at odds with the traditional and fundamental notion of cause and effect. One might well describe the situation as an experimental anomaly, a deviation that could no longer easily be reconciled with previous thinking in the field in question.

It is striking to observe how Faraday's took his next steps. While he knew well that the publication of this effect would be extremely spectacular, he decided not to publish his result immediately. By contrast, he kept it secret and spent all the time that his numerous other duties left him to researching the effect – this may well be an indicator of how puzzled and indeed troubled he was. He worked essentially alone, without communicating his work to anyone. At the same time, he had a favourable working context. Being director of one of the best equipped laboratories of Europe, he had free access to ample resources. While the laboratory was mainly chemical, the earlier work of Humphry Davy on electromagnetism had led to bringing substantial resources even for that domain into the laboratory, and Faraday well took advantage of this setting.

Following his research in its entirety is impossible within the framework of one paper: he pursued a multitude questions in parallel, jumped back and forth, and had both immensely tight research days and long breaks. What I shall do, with regard to my initial question, is to pick out one thread of research, a thread that finally resulted in forming new concepts out of "failure" of existing ones. Though such an approach does not reflect the full complexity of experimental research, it might help to illuminate the general questions put above.[5]

Separating Effects and Pathways of Research

The induction ring was a complicated arrangement, with electric and magnetic effects closely and inseparably interwoven. Faraday thus first attempted to "simplify" the arrangement, i.e. to separate electric from magnetic effects.

First, he attempted to exclude magnetism: he changed the arrangement in such a way as to give the coils nonmagnetic cores, of wood or of air, for example. He found that the induction effect was still present, though considerably weakened. What he had obtained thus was remarkable indeed: an induction effect just between currents: "Hence there is an inducing effect without the presence of iron", he noted in his laboratory notebook (D39).[6]

But he also worked in the opposite direction: he tried to exclude electricity as possible source of induction, i.e. to use common magnets as a source. Here the transient character of the effect posed special problems: common magnets cannot just be switched on or off. Faraday found an escape of that difficulty by using a soft iron bar that went magnetic, as long it was in contact with a magnet, and lost its magnetism as soon the external magnet was removed (Fig. 2). Indeed, and surprisingly, he obtained an induction effect, and again he was fully aware of what he had achieved: "conversion of magnetism into electricity", he noted (D33).

As a result, he had now *two* significantly different types of induction, each of which seemed to form a consistent field of effects:

- induction of currents by currents, without any magnetism involved,
- induction of currents by magnetism, without any inducing currents involved.

Fig. 2 'Making and unmaking a magnet' (Diary, D33)

In order to account for this new complexity, he introduced two new, and more specific concepts of induction: in the first case, he spoke of "Volta-electric induction", in the second case of "magneto-electric induction".

It is worth noting the specific conceptual constellation. Faraday had started with a rather unspecified notion of (electromagnetic) induction, understood as generation of electricity by magnetism, including electromagnetism. As a result of his experimental activity, he was able to specify that concept, but forced at the same time to split it up into two – with the new problem arising that their relation was totally unclear. The background of this change was, on the one hand, Faraday's desire to get insight into necessary and sufficient factors for producing induction effects and, on the other hand, the experimental results he obtained while pursuing that goal.

The new conceptual situation, i.e. the separation of two types of induction, had direct consequences for Faraday's experimental activity. For the time being, he treated the two types separately, as specific domains of effects. His main goal in each of these domains was to formulate correlations and regularities, i.e. to find out what were the conditions under which the effects occurred.

Volta-Electric Induction

With the first type – *volta-electric induction* – he rather quickly succeeded in formulating a regularity. In varying and simplifying the arrangement – from coils to straight wires – he found that, when switching on the inducing current, the induced current was always in the contrary direction.[7] In order to formulate this regularity, he could directly use concepts that had earlier been used in Ampère's electrodynamics: the concept of *interacting* currents, in *parallel* or antiparallel direction.

Still puzzling, however, was the transient character of the effect: induction occurred only when switching on or off the inducing current, not when it steadily persisted. In order to account for this discrepancy between expectation and observed effects, Faraday introduced a hypothesis: he speculated that the wire under induction was in a particular "electrotonic" state, supposed to counteract the persistent effect of the steady induction current. While Faraday thus saved the traditional notion of physical causation, he explicitly acknowledged that he had not been able to detect any other effects of that hypothetical state, such as electrical effects or attractive or repulsive powers[8] – a result that left his proposal in a problematic state: the electrotonic state remained a postulate that had solely been made to maintain the traditional notion of cause and effect even for the volta-electric induction effect, but found no support from somewhere else.

Magneto-Electric Induction: A First Regularity

The case of magneto-electric induction turned out to be even more complicated. The typical arrangements were combinations of magnets and wires, and the effects depended in a complex way on the geometric constellation. In particular, the relative

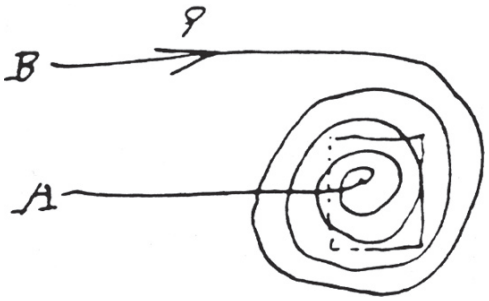


Fig. 3 Moving a spiral in front of a magnet (Diary, D201)

motion between magnet and wire came out to be an essential factor. But it was unclear to which reference frame the arrangements and motions could be referred in order to give a coherent regularity.

The first and simplest idea, given that physical processes were commonly considered in terms of attraction and repulsion, was to consider just movements of approaching or distancing from the poles. But this was immediately revealed as much too coarse: approaching the magnetic pole with the wire rendered vastly different results, depending on the specific constellation and direction of motion. Faraday's second attempt to formulate a law stemmed from his knowledge of electromagnets. Every magnet could be assigned a direction of (imagined) currents around it, whereby it could be made an electromagnet, and Faraday tried whether the direction of those currents could serve as a frame of reference. The approach was reinforced by that this direction was the same as the one of the hypothetical circular currents that Ampère famously had speculated to exist in all magnets, and indeed forming the cause of magnetism. Accordingly, Faraday examined experimentally circular arrangements in which a spiral of wire was moved towards, and withdrawn from, a bar magnet (Fig. 3: magnet perpendicular to drawing plane, front side depicted as square). And indeed, he was thus able to formulate the results into a regularity: a magnet approaching a circular wire induced a current in it, with a direction opposite to the imagined circular current around the magnet. That regularity had the nice feature, moreover, to have a strong analogy to the Volta-electric case, a point that Faraday noted explicitly: "Hence again inducing and induced current (first) in opposite directions" (D204).

A Mistake and a Failed Attempt

But there arose problems. First, Faraday discovered that he had committed a straightforward mistake in determining the direction of currents from the polarity of the battery and the specific constellations. These determinations were difficult tasks involving a sequence of mental operations, and what he realized now was that he had done one of these operations incorrectly. As a result of correcting this mistake, the direction of the induced current came out to be exactly opposite to what he had formulated in the above regularity. This change destroyed the previous fit between the laws of Volta-electric and magneto-electric induction.

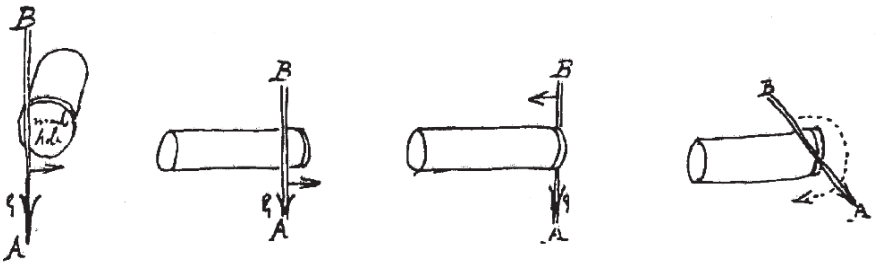


Fig. 4 Moving a straight wire along a magnet (Diary, D209 – D211)

What became even more serious, however, was another point. Triggered by the discovery of his mistake, he carefully re-examined his former experiments, and he realized that the dependency on the direction of motion was more complicated than he had thought before. In experiments with straight wires, moved along a bar magnet (Fig. 4), the experimental result depended on whether the wire was moved to the right or to the left, straight away from the magnet or so on, and those specificities could no longer be grasped by referring to the imagined circular currents (D211). The above regularity for magneto-electric induction fell down, i.e. proved as instable. And this cast serious doubts on whether the concept of circular (Amperian) currents did really work as appropriate reference frame. As a result, Faraday found himself in a similarly unclear situation as at the very beginning: it was again open what would be an appropriate reference frame to enable the formulation of empirical regularities, or laws, of magneto-electric induction. But compared with his starting situation, there were two differences: He had considerably strengthened experimental stability on the level of individual effects, and he had obtained much knowledge about what reference frames did *not* work, i.e. not allow to formulate regularities of the experimental results.

A New Concept as Solution

That Faraday indeed questioned the appropriateness of the reference frame is indicated by that he took a third attempt, with again a different concept. He tried out *magnetic curves*, i.e. the pattern of lines made by iron filings round a magnet. This meant strongly deviating from established physical reasoning: Those curves were well-known (see, e.g. Fig. 5, taken from the Encyclopedia Britannica of 1823), but were usually taken as not more than a curiosity. Nobody had ever thought of attributing physical meaning to them, rather there had been attempts to calculate their form as being a result of overlapping central forces. The most recent of these attempts had been undertaken right ‘before Faraday’s eyes’: A few months earlier, and in the house journal of the Royal Institution, of which Faraday was co-editor, Peter Mark Roget (like Faraday fellow of the Royal Society) had published a paper exactly on the mathematical properties of these curves.[9] Magnetic curves were well present at the period, but bore no physical meaning.

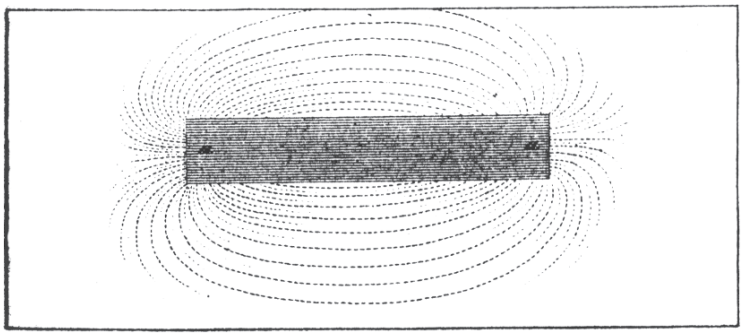


Fig. 5 Magnetic curves in the Encyclopedia Britannica, 1823

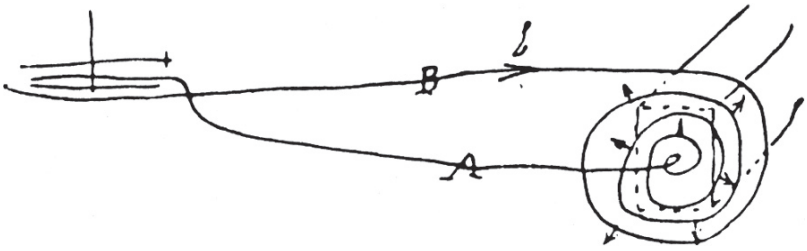


Fig. 6 Moving a spiral in front of a magnet, now referring to magnetic curves, indicated by small arrows (Diary, D214)

Faraday was well aware of how unconventional a step he took in attributing these curves a central role: he was indeed very tentative. But his attempt led to immediate success: he could quickly account for most of his experiments, including the problematic cases, i.e. he could formulate a regularity. In repeating the experiment with spirals, for example, he now saw how to account for the formerly irregular results with the new concept of magnetic curves (Fig. 6). Still it was difficult to verbalize the regularity, however, and Faraday took several attempts, without a really satisfactory formulation. Finally he resorted to a pictorial explication (Fig. 7): He drew a magnet AB, with the pattern of magnetic curves around it, and the cross section of a knife NP, vertical to the plane of the paper, with its edge at N. The law was then that a current was induced in the knife exactly when it was moved in such a way that it literally "cut" the curves.[10] And the effect would last as long as the motion went on.

After his numerous failed attempts, this law of magneto-electric induction meant a major success for Faraday. Indeed he had reached his goal: to formulate a law (or, more correctly, two laws) of induction that comprised the already large number of experimental results. It is indicative that he now, and only now, started to compile

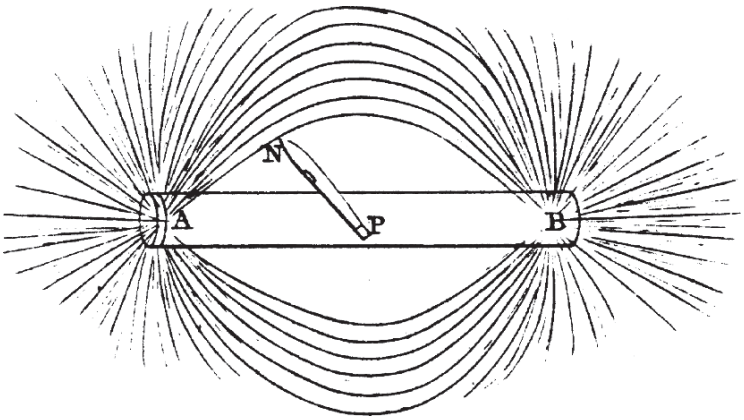


Fig. 7 Faraday's illustration of his induction law (Faraday 1832, par. 116)

his first publication on induction and submitted it to the Royal Society shortly later. He had reached this success only by several failed attempts, and by finally introducing a new concept or at least drastically changing its former meaning: the concept of magnetic curves.

It should be noted that by this step the separation of induction in two different domains was considerably solidified. Not only did the effects of volta-electric and magneto-electric induction look quite different, but there were now different concepts, a different language used here and there. Magnetic curves applied only to magneto-electric effects, while the electrotonic state applied only to volta-electric induction. The transient character of the effect was only mysterious with volta-electric effects, but no longer with magneto-electric ones, where motion was built in as essential factor into the very law. How fundamental this dichotomy was for Faraday became visible in his first publication: The paper was structured along this dichotomy, and the two types of induction were treated in separate chapters.[11]

With respect to the status of his new concept, it should be noted that Faraday explicitly refrained from talking about the physical "reality" of magnetic curves. He took them as an effective means to formulate the law of induction, and thus as a "mere expression of arranged magnetic forces" round a magnet. His law could work without any further commitments as to the physical existence of these curves.[12] And indeed he would keep such an agnostic approach for the next twenty years, while all the way expanding successfully the domain of application of these curves, soon to be renamed "lines of force".

Mixing up Effects and Categories

But the state of satisfaction did not last long. Around this time, one of Faraday's various other strands of his experimental activity interfered. He had investigated the

so-called Arago effect, a mysterious "rotation-magnetism" of nonmagnetic materials such as copper. The effect had bothered many since 1825, and Faraday saw now a way to explain it by his new induction effect, i.e. by currents induced in the rotating disc.[13] In this context, he conducted a striking experiment: If a metal plate revolved below a current-carrying wire, a current was induced in it (D347)! Only successively Faraday got aware of the subversive power of this result, when analyzed from the perspective of induction. The experimental device was a purely Volta-electric arrangement – there was no magnetism involved – but at the same time there was the alien factor of motion. What is more, Faraday immediately saw (and even explicated in his private notes) how that effect could be accounted for by *magnetic curves*: If only such curves were supposed also *around wires*, the effect was a clear case of his law of magneto-electric induction: the wheel, when rotating, would cut those curves, and a current would be induced![14] Faraday was deeply puzzled. The basic distinction of volta- and magneto-electric effects that he had used so successfully did not work in this experiment: the effect belonged to both and none of them at the same time. What is more, it was clear that the concept of magnetic curves could be expanded here with great success. But if one did so, the dichotomy between volta- and magneto-electric induction broke down even stronger: a concept belonging to the second domain would infer to the first! This was a serious challenge on a conceptual level and put into question the use of the basic conceptual distinction. The trouble was even stronger since the dichotomic concepts worked so well in other experiments.

Faraday found himself in a peculiar situation now. His paper to the Royal Society was basically structured along the dichotomic conceptualization that he now saw in trouble. What is more, he had the final manuscript in hand and wanted it to be printed soon, for reasons of already upcoming priority disputes. In that uneasy tension between research and publication interests, he decided to go for publication. He gave the manuscripts to be typeset, but not without cutting out the report of the problematic experiment that he had already inserted.[15]

Induction Effects: General Regularity and New Concept

Not surprisingly, the problem did not leave his mind. We do not know about his activities of the following weeks, but there was a four weeks period in which there are no laboratory entries. If he found time at all to do research in these weeks, he obviously did it without experiments. Indeed, the challenge he was facing was one of concepts, not primarily of further experimental results.[16] Though we have no records of how his considerations developed, the result makes one presume that he worked intensely on the problem. Virtually in the last moment (probably he had the page proofs on his desk…) he came up with a proposal that would solve the problem – again a very unconventional proposal: He proposed to drive the concept of magnetic curves even further, by considering them as moveable, and indeed moving in those moments, when an electric current was switched on or off. With

such a concept, a coherent law of induction could be formulated with the central condition that magnetic curves had to cut the wire. And that condition could be fulfilled by either the wire, or the magnet, *or the magnetic curves* moving: They had to be considered as spreading out in expanding concentric circles from the wire at the moment when the current was switched on. Likewise, these circles had to be considered as "shrinking" again towards the wire when the current was switched off.[17] All cases of Volta-electric induction could easily be covered by this conception, while the cases of magneto-electric induction could be attributed to the visible motion between magnet and wire.

This insight had drastic consequences. The dichotomy between Volta- and magneto-electric lost its use completely. In other words, it was now definitely revealed as a failure, even as misleading, since it veiled the result that both cases were covered by one and the same law of induction! Likewise, the concept of the electrotonic state lost its meaning. The transient character of the effect (that had been the only reason for introducing that state) was no longer a mystery, but rather an inherent, essential feature, a necessary consequence of the law. That was a most deep reconceptualization of the field indeed, concerning both the dichotomic concept of induction and the electrotonic state.

This reconceptualization was framed at a moment when Faraday had the page-proofs of his paper already laying on his desk – of a paper that was structured along the dichotomic approach, and in which he had devoted a whole chapter to the electrotonic state, a paper, moreover, that he wanted to be printed as soon as possible. Thus, while his concepts indeed had undergone fundamental revision, he neither was inclined nor probably had the possibility to revise the text substantially. In this awkward situation, the only thing he did was to add a footnote, stating that all induction effects could "be fully explained without admitting the electrotonic state".[18] In other words, he declared the electrotonic state as useless, and the chapter devoted to it for more or less meaningless – with a single footnote at its beginning. For all questions the surprised reader would perhaps raise, Faraday referred him or her to a second paper to be published soon! Thus when his first publication on the new induction effect appeared in print some weeks later, it bore a rather odd feature – a feature that strikingly points to the tortuous pathway of its generation, and exhibited even to his contemporaries at least one case of visible failure of a concept, the concept of electrotonic state.

Concepts Going Amiss

There are various ways in which things went amiss in Faraday's work. There was at least one straightforward mistake in the sense of wrong operation: his "miscalculation" of the direction of the induced current. Indeed, here we find the only case in which he spoke of "confusion" that he wanted to remove by "making precise observations." As I have discussed, this re-examination had led to most serious consequences. What is even more present in the episode, however, is change, revision, and

eventually failure on conceptual level. The list of concepts that became reshaped, and eventually re-stabilized, is impressive indeed:

(Electromagnetic) **Induction:**

- When Faraday started, he had just a general, though somewhat vague notion of processes in which *electricity was produced by magnetism (including electromagnetism)*. This was still clear enough to enable Faraday to recognize his very first experiment as an induction effect.
- The concept was then specified and broken up at the same time in two separated, and only loosely correlated concepts: *Volta-electric induction* as production of electricity by electricity, *magneto-electric induction* as production of electricity by magnetism.
- At the end of the whole series, finally, a unified notion emerged again, though in significantly different shape than initially: induction as the *production of electricity when a wire cut (or was cut by) magnetic curves*!

Motion and its reference frames nearby a magnet:

- Faraday initially conceptualized motion – like everyone in the period – as *approaching to or withdrawing from* specific points (magnetic poles),
- he then attempted to refer it to a certain *sense of rotation* connected with the magnet,
- and finally arrived at formulating motion in terms of *cutting magnetic curves*

Magnetic curves:

- Faraday started with the usual concept of magnetic curves as curious patterns around a magnet, as an *epiphenomenon* of attractive and repulsive forces.
- He then turned it into first a means to express the *distribution of force* around a magnet,
- ensuingly into a more general means to express the *distribution of magnetic force* even around a wire,
- and finally into a means to express the *distribution and change of magnetic force* around wires and magnets.

It should be noted, however, that one concept had well been challenged, but nevertheless been kept stable throughout the sequence: the concept of physical causation, implying that an agent caused an effect exactly as long as it itself was present. Compared with the others, this concept was much more general and fundamental, and Faraday did not open it for revision. Rather, he was ready to speculate and to invent *ad hoc* the hypothetical electrotonic state in order to save the specific concept of physical causation.

Experiment and the Failure of Concepts

What was the driving force of those changes? Why was Faraday not happy with the former concepts, and what was it that made the new ones appear "better" or more appropriate? The key issue is that those concepts served a specific purpose: they had to enable

ever more general formulations of the experimental outcomes in the form of regularities and laws that comprised many of those outcomes. In cases when this could not be done, Faraday was ready to ask whether the concepts were perhaps inappropriate, and he tried to readjust them (as in case of "induction"), or to find even totally new ones (as in the case of reference frames). The function of those concepts was most closely related to experiments and to the specific goal of formulating regularities or laws.

This goal, and the specific epistemic problem involved here can best explicated be by portraying a related, but still contrasting approach. John Stuart Mill was one of the few to discuss the functions of experiment in terms of searching for regularities or "causal connexions" as he named it.[19] Among the four experimental methods he proposed, the so-called "method of differences" is central (and the three others can even be reduced to it). First, it looks rather simple:

- Make an experiment with certain conditions A, B, C, …, and observe the outcome x.
- Then redo the experiment, but with one experimental parameter changed (say A → A'), while all others remain unchanged ("caeteris paribus").
- If the outcome changes, then there is a "causal connexion" between A and x, otherwise not.

In detail, the argument has some intricate problems. For example, it could be that changing A means at the same time changing another parameter D that has gone unrecognized so far. In that case, the causal connexion might not be between A and x, but between D and x. Mill discusses those cases and proposes ways to remedy that type of problem (and so do many philosophers of causation up to this day).[20] So, there are serious sources of error here – of erroneous logical inference, to be more specific. Indeed, as Giora Hon has pointed out, the only type of error or "fallacy" Mill discusses is "logical fallacy" in various forms.[21]

Another point has scarcely been noted, however. In Mill's argument it is of crucial importance that the categories A, B, C, and so on, are well-defined already *before* the whole procedure. Mill well saw the point and emphasized that those categories, i.e. the conceptual structure in which a field was handled, had to be established beforehand. He suggested that this was to be achieved not by experimenting, but rather a sort of contemplative process that he did not explain in much detail. Only with such a conceptualization ready-made, experimental conditions and outcomes could be formulated, and the experimental search for regular correlations could take off. There was no idea in Mill, however, that such a conceptualization itself could possibly be mistaken, inadequate or even fallacious, and should be readjusted, let alone in experimental context.

There is a type of experimental activity, rarely taken into view, that shares central features of Mill's "method", and that I have called "exploratory" elsewhere[22] – the above episode of Faraday's research is but one example. The experimental procedure is like Mill proposes: to vary parameters of the arrangement, not many at once, but only one by one, in order to see the effect of that change onto the experimental outcome – exactly as in the "caeteris paribus" condition. But the above case illustrates (and so do many others) that this procedure does not always lead to success, i.e. to stable correlations and laws, even when the realm of experiments has been taken really wide and

thorough. In such a situation of "failure" the researcher is left with several options, one of them being to drop the belief that such regularities exist (hard to take for a scientist), another one just to drop the subject, to leave that "dead end", and to switch to other topics (an option that does not occur too rarely, see the paper of Hans-Jörg Rheinberger in this volume). There is also the option, however, to "blame" the concepts that were used, i.e. to regard the situation as a failure of concepts, and consequently to put in question and open to revision the existing concepts and categories by which the experimental procedure was directed. This is what we see in the above Faraday episode, and in many others. But this is exactly a move not envisaged by Mill: what is at stake here is failure on the level conceptual level, i.e. the possibility of inadequate or mistaken concepts. As I have already indicated, and can specify now, situations of such a type are epistemically precarious in that it is not only the regular correlations between experimental elements that are investigated, but also the very categories in which they are formulated. Stability can only be achieved simultaneously, for concepts and regularities at the same time. The very fit of these two is the central indicator and criterion of stability.

To come back to my initial question, I hope both to have illustrated that there are indeed situations in experimental research in which concepts can fail, and to have spelled out some criteria for success and failure of concepts. It is essential to highlight that such talk makes only sense with regard to the specific goal or purpose involved in the research episodes under scrutiny. In Faraday's case, the goal was to formulate laws. There may be other goals in experimental situations, and a concept that is found to be appropriate in the context of one of these goals is not necessarily appropriate in other ones, such as application, mathematization, production, optimization, …While the talk of inappropriate or even failed concepts thus always is relative to the specific purpose of the research in question, it might nevertheless be useful – not only as analytical tool for the historian/ philosopher of science, but well to grasp the historical actor's situation, as Faraday's repeated attempts at new or revised concepts illustrate.

As a final remark, a note on reference may be in place. Talking of success or failure of concepts does not necessarily have any implication on their reference to some entities, or to "physical reality." With respect to magnetic curves, (mechanical) force, action at a distance, phlogiston, electrons, genes etc. there are not only the options to conceive them either as physically existing entities or as mere inventions of the mind without physical referent. They can also be conceived just as appropriate means to order things while being ignorant about physical existence with all the intricacies of the question what "physical existence" could mean in such cases. It is the latter position that is highlighted by the talk of concepts as elements of language that serve certain purposes. And it appears to me that this is the way these concepts are very often pragmatically used in research practice, without too much concern about their specific ontological status. The talk of concepts and their success or failure does not focus on ontology, but rather on the actual practice of introducing and using new talk in a specific realm. Even if many concepts, once formed in exploratory context, have later been turned into "facts" and into "physical reality" (such as Dufay's two electricities, Kekulé's benzene ring, the electron, perhaps the gene, …), the process of their formation might well (and perhaps better) be understood without having questions of reality in mind.

This observation may render the study of conceptual appropriateness of failure easier for historians and philosophers of science: it takes away, at least for a moment, the burden of ontological questions. And we might well need more studies of those processes: After all, the Faraday episode illustrates that situations of conceptual uneasiness or conceptual failure might be of outstanding importance in the development of science, and that they perhaps deserve closer attention, both with respect to the specific experimental approaches pursued and the question of going amiss or failure of concepts.

Notes

1. (Smith and Medin 1981), (Prinz 2002), (Gurova 2003).
2. (Ross 1965), (Devons 1978), (Teichmann 1990).
3. For an excellent brief introduction to Faraday, see (Cantor et al. 1991), for a more extended treatment, see (Williams 1965).
4. (Faraday 1832), par. 60.
5. A fuller account, though not focussed on the present questions, is found in (Romo and Doncel 1994) and (Steinle 1996).
6. Faraday's laboratory notebook is fully edited: (Martin 1932–1936). I refer to the paragraphs of vol. I by giving their number after a D . . . .
7. (Faraday 1832), par. 26.
8. (Faraday 1832), par. 61–62.
9. (Roget 1831).
10. (Faraday 1832), par. 114–116.
11. (Faraday 1832), § 1 and 2, respectively.
12. (Faraday 1832), note to par. 114. This remark is pertaining when it comes to discuss whether the appropriateness of concepts depends on the existence of any physical 'referent,' see the last section of this paper.
13. Newenduobe: (Steinle 1994).
14. Cf. Faraday's manuscript version to par. 241 of (Faraday 1832): RS MS PT 20.5. For arguments about the date and context of that passage, see (Steinle 1996), 178–180 and 211–212.
15. (Steinle 1996), 182.
16. There is a striking similarity here to the research pathway of Hans Krebs, in its last phase, when he had an abundance of experimental results, but was still lacking a coherent picture: Krebs also had a phase of several weeks without any laboratory research, at the end of which he came up with a fundamentally new conceptualization of the domain, see Graßhoff and Nickelsen, this volume.
17. (Faraday 1832), par. 238.
18. (Faraday 1832), 139.
19. (Mill 1843), bk. III, chs. 7 & 8.
20. (Graßhoff et al. 2000), as an intriguing example.
21. (Hon 1991).
22. (Steinle 2002, 2003, or 2005).

References

Cantor, G. N. et al. (1991), *Faraday*. London: Macmillan.

Devons, S. (1978), "The search for electromagnetic induction", *The Physics Teacher* (December): 625–631.

Faraday, M. (1832), "Experimental researches in electricity. §1. On the induction of electric currents. §2. On the evolution of electricity from Magnetism. §3. On a new electrical condition of matter. §4. On arago's magnetic phenomena", *Philosophical Transactions* 122: 125–162.

Graßhoff, G. et al. (2000), *Zur Theorie des Experimentes. Untersuchungen am Beispiel der Entdeckung des Harnstoffzyklus*. Bern: Bern Studies for the History and Philosophy of Science.

Gurova, L. (2003), "Philosophy of science meets cognitive science: The categorization debate", in D. Ginev (ed.), *Bulgarian studies in the philosophy of science*. Boston studies in the philosophy of science, 236. Dordrecht: Kluwer, 141–162.

Hon, G. (1991), "A critical note on J.S. Mill's classification of fallacies", *British Journal for the Philosophy of Science* 42: 263–268.

Martin, T. (ed.) (1932–1936), *Faraday's Diary. Being the various philosophical notes of experimental investigation made by Michael Faraday, DCL, FRS, during the years 1820–1862 and bequeathed by him to the Royal Institution of Great Britain.* London: G.Bell & Sons.

Mill, J. S. (1843), *A System of Logic Ratiocinative and Inductive. Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation.* London: Parker.

Prinz, J. J. (2002), *Furnishing the Mind: Concepts and Their Perceptual Basis*. Representation and Mind. Cambridge, MA: MIT Press.

Roget, P. M. (1831), "On the geometric properties of the magnetic curve, with an account of an instrument for its mechanical description", *Journal of the Royal Institution of Great Britain* 1 (February): 311–319.

Romo, J. and Doncel, M. G. (1994), "Faraday's initial mistake concerning the direction of induced currents, and the manuscript of Series I of his researches", *Archive for History of Exact Sciences* 47: 291–385.

Ross, S. (1965), "The search for electromagnetic induction 1820–1831", *Notes and Records of the Royal Society of London* 20: 184–219.

Smith, E. E. and Medin, D. L. (1981), *Categories and Concepts*. Cognitive Science Series 4. Cambridge, Mass: Harvard University Press.

Steinle, F. (1994), "Experiment, speculation and law: Faraday's analysis of Arago's Wheel," in David Hull, Michael Forbes & Richard M. Burian (eds.), *PSA 1994. Proceedings of the 1994 biennial meeting of the philosophy of science association.* East Lansing: Philosophy of science association vol. 1, 293–303.

Steinle, F. (1996), "Work, finish, publish? The formation of the second series of Faraday's 'Experimental Researches in Electricity'", *Physis* 33: 141–220.

Steinle, F. (2002), "Challenging established concepts: Ampère and exploratory experimentation", *Theoria: revista de teoria, historia y fundamentos de la ciencia* 17: 291–316.

Steinle, F. (2003), "Experiments in history and philosophy of science", *Perspectives on Science* 10(4): 408–432.

Steinle, F. (2005), *Explorative Experimente. Ampère, Faraday und die Ursprünge der Elektrodynamik*. Boethius 50. Stuttgart: Franz Steiner Verlag.

Teichmann, J. (1990), "Der Entdeckungsweg der Induktion 1820–1831", *NiU Physik* 1: 44–49.

Williams, L. P. (1965), *Michael Faraday: A Biography*. London: Chapman & Hall.

A Pioneer Who Never Got It Right: James Dewar and the Elusive Phenomena of Cold

Kostas Gavroglu

To discuss errors, especially in the work of those scientists whose activities were predominantly experimental, is somewhat uninteresting: errors happen all the time, and many experimenters either (eventually) discover them or they are pointed out to them. Errors, are in a way, an integral part of experimentation, it is important to record them, they have been discussed by the experimenters themselves in articles they wrote about their experiments, and they have, even, been commented upon in historical writings. Often, they have been part of the account of priority disputes. Nevertheless, such errors appear to be devoid of much historiographic interest, since, they are, in a way, of a technical character. Almost always their sources have been uncovered, the reasons for their creeping up have been well understood and, in most cases, they were not repeated by subsequent experimenters. Even though they are part and parcel of the experimenters' way of life, errors of this sort do not have any appreciable effect on the overall practice of the scientists involved, they do not seem to affect macroscopically what they do.

Alternatively, there is a state of affairs which is often confused with errors and which presents immense historiographic interest: going amiss. The study of cases of going amiss is a distressful process full of predicaments. On the one hand, such a study needs a healthy dose of anachronistic readings and, on the other, it is necessary to rise above the indignities inflicted by such indulgence with anachronisms. What is *not* meant by going amiss is getting nowhere. When one is going amiss, one does, in fact, get results – experimental, mathematical, theoretical, interpretational. We know someone was going amiss, because where he got did not turn out to be "correct" according to later developments. The study concerning going amiss of programs and persons working within research programs, is not a discussion about the false consciousness of the researchers, nor is it an attempt to discuss the psychology of discovery.

The discussion concerning going amiss is rather challenging for the historian of science, who is obliged to understand as correct the results that later on we know they were either wrong or indifferent to later developments. Going amiss is not an

K. Gavroglu (✉)
Department of History and Philosophy of Science, Athens University, Athens, Greece

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_9,

accusation, but rather a characterization: it is an *a posteriori* characterization of the overall framework within which theoretical or experimental practices are defined. It is conditioned by methodological choices, philosophical preferences, ontological commitments, beliefs which turn out to be prejudicial attitudes, self-evident assumptions which turn out to be unnecessarily constraining conceptions. Such are the ingredients of research programs and theoretical frameworks which, in a way, oblige the researchers, to insist that they unswervingly persevere in what it is they were doing. Going amiss is about such constraining frameworks and the ways the scientists' practices are accommodated within these frameworks. And understanding the character of constraints unfolds the reasons for the excessive rapport and the perceived affinity scientists have to specific theoretical schemata. This is something related to individual scientists, but it, also, displays social and cultural dimensions. Hence the discussion about going amiss, is a discussion about motives, and most significantly, about understanding practices – experimental or theoretical. Examining such practices help us clarify modes of obtaining experimental knowledge, the conceptualisation of prejudices, the extent of commitments to theories or theoretical schemata and the ensuing deadlocks of the programs. One thinks of Einstein's decades long attempts to unify gravitational and electromagnetic forces – with a lot of results, yet not being able to show any tangible progress towards unification. Research on unified theories during the 1970s and 1980s has shown that Einstein's perception of these forces, their assumed "similarity" and, hence, his insistence to unify electromagnetic and gravitational forces, was a case of going amiss.

In what has been said above there is a clear cut distinction: errors are of a technical character, going amiss can only be assessed after the event, ex post and anachronistically. Hence, in questions of what we call "wrong interpretations" errors do not creep in. Of course, there are cases when wrong interpretations are, in fact, interpretations full of errors. But what characterises wrong interpretations is that they are cases of going amiss, since most often they are formulated within conceptual frameworks conducive to going amiss. A striking case is superconductivity. In 1911 Kamerlingh Onnes observed that at about 4° absolute the electrical resistance of mercury dropped abruptly to an immeasurably low value. This experimental result was interpreted as being a case of purely zero electrical resistance and for over 20 years the physicists' exclusive emphasis in their attempts to understand the phenomenon of superconductivity, was the understanding of *infinite conductivity* at very low temperatures. In 1933 the experimental testing of an "expected" magnetic behavior of metals at such low temperatures, provided "unexpected" data and this changed the interpretation of the original experimental results, changing the whole viewpoint about the problem of superconductivity and transforming it from a problem in electricity to a problem in magnetism. It was only then that it became possible to build the first successful theory. Thus one can have correct results but wrong interpretations. And in some cases the correct results, exactly because the experiments had been thought, designed and their data interpreted within a framework with well defined ontological commitments, induce misleading interpretations. Furthermore, the question arises as to the possibilities to assess going amiss from the actors' viewpoint. It is, indeed, possible to have such insights. Disputes

sometimes express this. Sometimes there are explicit statements by scientists as to why they prefer particular theoretical approaches which later on turn out to be complete deadlocks (e.g., S-matrix theory) as there are cases when misplaced ontological commitments are dominant in the formulation of the theory (e.g., Ostwald's version of energetics).

A discussion, then, about going amiss, is all about understanding the insistence of the researchers to continue along – what is for us presently – the wrong paths, for – what was for them in the past – the correct reasons, while doing everything – what for us and them – in the right manner.

This paper will analyze two incidents of going amiss, related to the multifaceted work of James Dewar.

The two incidents are the following:

- Dewar was convinced that what Lord Rayleigh and William Ramsay had discovered in 1895 was not the inert element argon, but an allotropic form of nitrogen. He himself performed a number of experiments, but the results were interpreted within the framework of what he termed "chemical thinking" and which deterred him from accommodating the notion of a chemical element which did not react with any other.
- Encouraged by his liquefaction of hydrogen, Dewar thought that he had also liquefied helium. But the liquefaction of helium proved rather resistant to a laboratory practice of "brute force" for the lowering of temperature and it was liquefied at Leiden by H. Kamerlingh Onnes in 1908 within a radically different laboratory culture than that of Dewar's.

The study of going amiss provides us with the possibility to make comparisons with those cases which got it right. A comparative approach will accentuate the methodological differences in the relative research programs, but will also help to reveal underlying assumptions and commitments to different theoretical schemata. In both the cases we shall examine, an attempt will be made to compare research programs so that to articulate the differences of laboratory cultures and experimental practices.

James Dewar (1842–1923) was an experimenter of amazing versatility who worked in problems related to physical chemistry, spectroscopy and molecular physics. His researches led to a large number of extremely significant discoveries and improvements in instrumentation and experimental techniques (Dewar flask, improvements in spectrometers, thermometry and cryogenic apparatus, achievement of high vacua). Apart from hydrogen, he was also able to liquefy fluoride, develop smokeless cordite, propose the initial idea related to the benzene ring and study the chemical and physical properties of a large number of phenomena (constants of hydrogenium, chemical reactions at the temperatures of the electric arc, conditions affecting the excitation of spectra, absorptive power of charcoal, electrical and magnetic properties of metals at low temperatures). He was born in Scotland, was educated in Edinburgh and worked briefly there. For a short time he worked with Kekule at Ghent. In 1875 he was elected the Jacksonian Professor of Experimental Philosophy at the University of Cambridge and 2 years later he became the Fullerian

Professor of Chemistry at the Royal Institution. He held both posts until his death and from 1896 to 1923 he was the Director of the Davy-Faraday Research Laboratory of the Royal Institution. He held a number of public offices and was knighted in 1904. He was a member of the Government's Committee on Explosives from 1888 to 1891, he delivered the Christmas lectures for children at the Royal Institution (1878–1879), he served as President of the Society for Chemical Industry (1897), the Chemical Society of London (1897–1899) and of the British Association for the Advancement of Science (1902). He delivered numerous public lectures and wrote many letters on a wide range of subjects to the *Times*. James Dewar belonged to a generation who was well anchored in the culture of classical physics, who lived for many years through the most decisive period which established quantum theory and, yet, he, as many others in Britain, was totally impervious to the directives of the new view point. Dewar had a remarkably wide range of scientific preoccupations, he derived significant results in almost all the special topics towards which he directed his researches, he proceeded to technical inventions and improvements which had lasting effects and a substantial portion of his work could be grasped by an wide audience.

Discovering Argon: Can a Chemical Element Not Combine with Other Chemical Elements?

Lord Rayleigh's measurements for the exact determination of the densities of gases had started in 1882 while he was the Cavendish Professor of Experimental Physics at the University of Cambridge. He continued them in 1888, having left Cambridge and having been appointed Professor of Natural Philosophy at the Royal Institution in London. It was a program aimed to test Prout's hypothesis by finding the atomic weights of gases and observing the extent to which they were multiples of the atomic weight of hydrogen. By 1892 Rayleigh found a curious discrepancy. In a letter to *Nature* he noted that the density of nitrogen depended on the method used to isolate the gas. The nitrogen he derived by the two different methods he called "physical" nitrogen and "chemical" nitrogen. The former was isolated by removing the oxygen, moisture and carbon dioxide from samples of atmospheric air. Chemical nitrogen was prepared from ammonia. It was found that physical nitrogen was heavier than chemical nitrogen by about 1/1000. Rayleigh, a few years later, recalled that the next step was to find ways to exaggerate this difference

> One's instinct at first is to try to get rid of a discrepancy, but I believe that experience shows such an endeavour, to be a mistake. What one ought to do is to magnify a small discrepancy with a view of finding out an explanation.[1]

Further improvements showed that chemical nitrogen was about 0.5% lighter than physical nitrogen. The first alternative Rayleigh entertained was that atmospheric nitrogen was too heavy because of the imperfect removal of oxygen from the atmospheric air or chemical nitrogen was too light because when it was removed from ammonia it was contaminated with gases which were lighter than nitrogen.

Further experiments by Rayleigh excluded both possibilities. There was also the possibility that the discrepancy might be due to the dissociation of the nitrogen molecules and their subsequent formation into N_3 much like the situation with the production of ozone from oxygen by silent discharge. Rayleigh ruled out this possibility too, by showing that electrification and sparking had no appreciable effect in altering the densities of the two kinds of nitrogen. By the beginning of 1894 Rayleigh was convinced that the atmosphere contained a new hitherto unknown constituent.

Rayleigh's experiments to isolate the new constituent were, in effect, very similar to the experiments performed by Cavendish in 1785. Rayleigh tried first to remove the oxygen from the atmospheric air then the nitrogen and then the carbon dioxide and other similar gases. The difficulty, of course, was in the removal of nitrogen since it chemically combines only with certain elements and under specific conditions. A characteristic run is found in the very first page of Rayleigh's *Notebooks*.[2] He started with 50cc of air and continuously added oxygen and with the help of the sparks he could have a union of oxygen with nitrogen. The addition of oxygen continued until there was no noticeable contraction of the volume of the gas inside the test tube after sparking for 1 hr. What remained was transferred to another tube and found to be 1cc. This was passed over alkaline pyrogallate and the final product was 0.32cc. This substance could not have been nitrogen since it did not decrease after continuous sparking nor could it have come from somewhere else since repeated measurements had shown that it was proportional to the mass of the original intake of atmospheric air. Rayleigh called it the "residue".

In the meantime, William Ramsay, Professor of Chemistry at University College London, had proposed to Rayleigh that there may be a more efficient way for the study of the problem. His method consisted of a series of tubes connected between them which contained magnesium, copper oxide which could unite with India rubber, preheated soda, limesoda that it would not contain water vapour and phosphoric anhydride. The heated magnesium absorbed the nitrogen and repeating the process by recirculating the gas collected at the end of the previous run, Ramsay, starting with 1094cc of nitrogen he was left with a residue of 50cc which, nevertheless was still not very pure. Up until the beginning of August 1894 Rayleigh and Ramsay were working independently and at that time they decided to join forces and plan towards a joint publication. They were both convinced that atmospheric air contained either a new element or a new compound.

The results were first presented during the meeting of the British Association for the Advancement of Science at Oxford in August 13, 1894. In a brief announcement read by Rayleigh it was reported that atmospheric nitrogen when purified from all the other known constituents of air was found to contain another gas to the extend of about 1% which was even more inert than nitrogen. The density of this gas was found to be between 18.9 and 20 and preliminary observations of its spectrum had found a characteristic line. Right after the BAAS meeting at Oxford, James Dewar, wrote two letters to *The Times* claiming that what was found by Rayleigh and Ramsay was the triatomic form of nitrogen. Dewar suggested that the allotropic form of nitrogen could produce spectra which were distinct from nitrogen and in the case of

Rayleigh and Ramsay "the new substance is being manufactured by the respective experiments, and not separated from ordinary air."[3]

Dewar's *Laboratory Notebooks*,[4] are particularly revealing in understanding the kinds of experiments he was performing at the time with nitrogen. It appears that Dewar started to study intensively the low temperature behaviour of chemical nitrogen and atmospheric nitrogen, right before the meeting of the BAAS. He was desperately trying to establish that Rayleigh and Ramsay had mistaken the new gas with N_3, since he had found that both kinds of nitrogen liquefied at the same temperature.[5] He drew his inferences from experiments involving liquefaction of air and the white deposits he always found in the otherwise transparent liquid. The conclusions he reached were through a characteristic chemical thinking: He suggested that the theoretical density of the new nitrogen compared to hydrogen, should be 21, while the experimental numbers are between 19 and 20. He surmised that for such a body, "chemists would infer", that it ought to be characterised by great inertness, because phosphorus, the element most nearly allied to nitrogen, easily passes into an allotropic form known as red-phosphorus, which, relative to the yellow phosphorus, was an inert body. If, therefore, such an active body as phosphorus could become, in condensed form, far less active chemically, then, «by analogy, nitrogen, so inert to start with, must in the new form, become exceedingly active (WRONG)!»

On December 6, 1894 Dewar presented to the Chemical Society his experiments concerning the liquefaction of nitrogen. The meeting took place less than a week after Lord Kelvin in his Presidential Address at the Royal Society had referred to the discovery of the new constituent as the greatest scientific event of the year. In his talk Dewar claimed that chemical and physical nitrogen liquefied at the same temperature and boiled off at the same rate.[6] From this he inferred that the assumed new substance present at the atmospheric nitrogen does not condense at temperatures when all other gases condense or that it behaves in exactly the same manner as nitrogen. In an unsigned piece the next day reporting the meeting at *The Times*, it was remarked that "Chemists will appreciate the extreme singularity of a substance with the assigned density which fulfils either condition."[7]

Both Rayleigh and Ramsay did not attend the meeting. Dewar's announcement gave the opportunity to Henry Armstrong, to underline the case of his fellow chemists. He ventured to say that Lord Rayleigh and Prof. Ramsay now could not hope to keep so remarkable a discovery to themselves much longer. He was adamant that "chemists could not be expected to remain…under the imputation that they had been eyeless during a whole century." And he concluded by talking about the "unquestionable rights of the chemists" to exercise entire freedom of judgement, and to critically examine the statements which had been made.[8] Apart from wishing to be absolutely certain before fully committing themselves to the suggested discovery, the other reason why Rayleigh and Ramsay were quite secretive about the details of their experiments, was that they were planning to claim the Smithsonian Hodgkins Prize awarded to discoveries related to the atmosphere. This they received in 1895 after the final announcement of their discovery.

The final announcement was made at a meeting of the Royal Society at the Theatre of University College London on January 31, 1895. The paper was presented

by William Ramsay and Lord Kelvin was chairing the session. This time it was Dewar who was absent. Ramsay described all the different methods used to isolate atmospheric nitrogen and chemical nitrogen and the difference of less that 1% in the measured densities of the two kinds of nitrogen. Then he presented the methods for removing the nitrogen and the different methods to induce chemical combinations with nitrogen. There was always a remaining residue which could not be gotten rid of. Ramsay, then, discussed a number of ways to isolate the new gas and to obtain it in relatively large quantities. Having achieved that, William Crookes and Arthur Schuster examined its spectrum and found that it did contain certain lines which were not contained in the nitrogen spectrum. This was one piece of convincing evidence that what was found was not N_3. The other was the extreme inertness of argon whereas most of the chemical evidence implied that it would be almost explosive. Ramsay continued describing the solubility of argon in water and its liquefaction and a more detailed account was presented at the same meeting by Olszewski. By measuring the velocity of sound in argon, Rayleigh and Ramsay managed to find the ratio of specific heats. It was found to be 1.66. This implied that argon was monatomic and, hence, quite impossible to be accommodated in the periodic table as that table was structured at the time.

After the formal announcement of the discovery of argon *Nature* carried a detailed report of the meeting with various comments most probably written by Arthur Rucker, professor of physics at the Royal College in London. The report remarked that

> All that is known of argon was told to all.…As has been well said, the result is "the triumph of the last place decimals", that is, of work done so well that the worker knew he could not be wrong…[and concerning the disagreements about the monatomicity of argon it was added that] The courts of science are always open and every litigant has an unrestricted right of moving for a writ of error.[9]

Accommodating and Legitimating the New Element

It is rather interesting that the official reporting of the meeting in *Nature* expressed an explicit historiogpahic preference: The argon story had been projected as a paradigmatic case of a culture of "next decimal pace" measurements. At the end of 1894 Lord Kelvin in his Presidential address at the Royal Society, reminded his audience of the comments he had made 23 years earlier. Commenting on the "still anonymous fifth constituent of our atmosphere" as "the greatest scientific event of the past year", he continued.

> Accurate and minute measurement seems to the non-scientific imagination a less lofty and dignified work than looking for something new. But nearly all the greatest discoveries of science have been but the rewards of accurate measurement and patient long-continue labour in the minute sifting of numerical results.[10]

This view which has been since shared by many scientists, is a greatly misplaced assessment. By considering the discovery of the argon in such a context a crucial

element of this discovery is lost. The argon discovery, is hardly a discovery of the "next-decimal-place." As Rayleigh, was the first to point out, Cavendish, nearly a century earlier, in his *Researches on Air* while attempting to remove all the nitrogen from a jar, had noticed that there was a residue of less than one hundredth left and which he could not remove. Hence, it is at least difficult to justify that it was the culture associated with the specific type of measurements at the end of the nineteenth century which led to the discovery of argon. Furthermore, Rayleigh and Ramsay in the beginning of their paper put a quote from De Morgan's *A Budget of Paradoxes* as if to counteract any such attitude about "next-decimal-place".

> Modern discoveries have not been made by large collections of facts, with subsequent discussion, separation and resulting deduction of a truth thus rendered perceptible. A few facts have suggested an *hypothesis*, which means a *supposition*, proper to explain them. The necessary results of this supposition are worked out, and then, and not until then, other facts are examined to see if their ulterior results are found in nature.[11]

The deep significance of the argon story is slighted unless it is considered as a story involving a bitter public dispute concerning the legitimacy of a new chemical element whose most important characteristic was that by being chemically inert, it was negating the very notion of a chemical element. Argon forced chemists to re-appraise some of the constitutive notions of their discipline. Similarly physicists were obliged to re-think the boundaries between physics and chemistry and start coming to terms with the notion that chemistry, after all, might not be all reducible to physics. Argon was discovered during a time when physical chemistry was articulating its own autonomous language with respect to both physics and chemistry, when it was charting its own theoretical agenda and formulating its own theoretical framework and it appears to have had very little to do with next-decimal place measurements.

Ramsay, in fact, had discussed many similar issues related to physical chemistry with Ostwald and FitzGerald. Ostwald had written to Ramsay that he would gladly publish his paper in the *Zeitschrift fur Physikalishe Chemie*. "The fact is that I do not care very much for the new elements. But one so unexpected and almost impossible as that which you have found is something totally different from the trivial discoveries amongst the rare earths."[12] FitzGerald proposed that Ramsay make a determination of the specific heat at constant volume and a calculation of it from the value of and the P,V,T relation and thus decide whether it obeys the Dulong Petit law. Ramsay was seeking FitzGerald's opinion about the peculiarities of the ratio of specific heats he had found for argon.[13] The latter was convinced that such a calculation would lead to the atomic weight of 40. "This is certainly very mysterious". FitzGerald suggested that this may imply that the two atoms may have little or no independent motion and so the molecule behaved like a single atom. "I make this in the interests of chemistry because physically there can be no objection to an atomic weight of 40."[14] Ramsay had suggested to FitzGerald the possibility of a system of elements with zero atomicity and the latter, though very enthusiastic about the suggestion, warned Ramsay that the "Chemists will never believe in an element with no chemical affinity."[15] And Ramsay felt no scruples in telling Smithells that the implications of argon were such that "the whole fabric of chemistry is going to receive a shake."[16]

The discovery of argon was far from being a joyous affair for chemists. It deeply insulted many and distinguished British chemists. The two most vocal critics among the chemists were Henry Armstrong, President of the Chemical Society and James Dewar. It is often the case that the extreme dislike both entertained against Ramsay is given as the reason for this conflict. From all the evidence which has been coming to surface and to use chemical language, it appears that Dewar had a rather strong affinity in disliking people, generally, whereas Armstrong's tendency was an incomprehensible hero worship of Dewar. Though personal factors cannot be denied in trying to understand the reactions of the chemists to the discovery of argon they are neither sufficient nor can they be a substitute for understanding such conflicts and public disputes in the context of the dramatic developments which were taking place during the end of the nineteenth century in both chemistry and physics.

The disagreements were suggestive of the way each experimental tradition chose to articulate its own agenda within the newly emerging sub-discipline of physical chemistry. The disagreements so aggressively expressed by many chemists about the new element, had mainly to do with the threatening emergence of physical chemistry as a distinct new subdiscipline rather than personal enmities. To many chemists the "invasion of mathematics to chemistry",[17] and the increasing openness of many other chemists to adopt techniques and concepts from physics, were unsettling indications that the very character of chemistry – the(ir) laboratory science *par excellence* – was being negotiated. And it was exactly against this new framework that strong phobias were expressed by many chemists. At the end, after the dust had settled, it appeared that argon «belonged» to those physicists who for a moment felt like chemists and to those chemists who started realising that physical chemistry was not simply a way of enriching chemistry with techniques borrowed from physics.

Liquefying Hydrogen

It was in 1894 – 5 years before the announcement of its actual liquefaction – that an article about the problems involved in the liquefaction of hydrogen appeared in *The Times* on September 1, based on details provided by Dewar.

> The strenuous efforts now being made by physicists to approach the zero of temperature are in some ways analogous to the numerous attempts that have been made, or are now being made, to reach the North Pole.…In both cases success may be said to depend upon equipment, persistency, and the selection of the right road.

Dewar's eventual liquefaction of hydrogen in 1898 was above all, a triumph of an ingeniously performed combination of instrumentation and manipulation. From his very first attempts, Dewar had decided to have a mixture of hydrogen and nitrogen at high pressure and at temperature of –200 degrees and by expanding it to have a lower temperature for the starting point of the cycle which will deal with the lowering of the temperature of hydrogen. This mixture could be seen to

form a jelly of solid nitrogen and produced hydrogen, since the gas coming off burned with an intense flame. And even when hydrogen contained some 2–5% air, a white solid matter along with a "clear liquid of low density, which is so exceedingly volatile that no known device for collecting has been successful."[18] It should be noted that when Dewar started his experiments, there was no consensus that hydrogen could be liquefied.[19] And in 1891 the situation was summed up by William Ramsey: Hydrogen has never been "condensed to the solid or liquid states. Cailletet and Pictet who claim to have condensed it by cooling it to a very low temperature, and at the same time very strongly compressing it, had in their hands impure gas."[20]

In his announcement at the Chemical Society,[21] Dewar reported in a vague manner about his experiments. Hydrogen was cooled to –200 degrees and passed through a fine nozzle under 140 atm, but no liquid was seen. If hydrogen contained a few percent oxygen, the latter was seen to liquefy and hydrogen was given off. If, however, hydrogen was cooled by a bath of boiling air, and allowed to expand at 200 atm over a regenerative coil previously cooled to the temperature of –200 degrees, a liquid jet was observed after the circulation continued for some minutes along with a liquid which was "in rapid rotation in the lower part of the vacuum vessel." It was not possible to accumulate the liquid both because of its low specific gravity as well as because of the rapid current of gas. "These difficulties will be overcome by the use of a differently shaped vacuum vessel, and by better isolation. That liquid hydrogen can be collected and manipulated in vacuum vessels of proper construction cannot be doubted. The liquid jet can be used in the meantime (until special apparatus is completed for its collection) as a cooling agent, like the spray of liquid air is obtained under similar circumstances, and this being practicable, the only difficulty is one of expense."[22]

James Dewar's experiments on the liquefaction of hydrogen started on February 25, 1898. It was the first run with the new apparatus he had just constructed. For 3 hr he accumulated liquid air in the inner vessel. The hydrogen was kept under pressure during the cooling with a "slight leak through the regenerator coil."[23] All apparently went well until after the liquid air was put under exhaustion "when in attempting to open the H[ydrogen] valve nothing would come." There was an obstruction in the tubes and it could not be easily removed. Dewar thought that the obstruction was solid air, and the stoppage remained for hours afterwards. The air valve was opened and the hydrogen which was under pressure rushed into the liquid air coil. Such sudden adiabatic expansion gave a mist of solid air and solid impurities along with liquid drops which were seen on the glass for a moment. The expansion took place through a vacuum V vessel that was open at the bottom. During the gaseous discharge, two vivid electric discharges were observed in the hydrogen which was in the inside of the vacuum vessel. He realised, however, that the sparking was most probably due to solid particles of solder carried by the hydrogen from the coils.

The next set of experiments was performed on March 14, having been interrupted since February 25 because of "repeated failures in getting a silvered vacuum tube to fit the regenerator" and "also for some reason in getting this to work." Starting

from room temperature and air compressed at 200 atmospheres, he could get only 5% yield of liquid air. The use of V test tubes with outlet in the bottom for the collection of liquid air could not be properly regulated and there was not even a yield of liquid air. When the experiment was repeated hydrogen gas was running at about 12 cubic feet a minute, and after a few minutes, a minute amount of liquid collected in the mirror of the two vacuum test tubes round which the hydrogen was entering from the regenerator.

> I looked in a few minutes later, but found all the liquid gone. Can this have been liquid hydrogen?? Unfortunately, the H ran only a short time, yet the pressure was reduced 20 atm…After shutting some hours, the regenerator remained plugged. Looks as if mechanical stoppage due as before to solder carried forward, not due to solid air. In the experiment it was a white solid, this was carried down into the V-tube. Can the H[ydrogen] be solid?

The new experiments were, again, seriously hampered by the clogging of tubes with the frozen impurities of the hydrogen gas. In one experiment after 2.5 h of hydrogen circulation, the circulation stopped because of clogging.

> No visible liquid appeared in the vacuum tube. Noticed, however, the H[ydrogen] escaping from the spiral of the vacuum vessel was so cold that as the H[ydrogen] discharged was arrested, the air was sucked in the liquefied in the spiral. The H[ydrogen] stopped just when the real cooling began.[24]

The small pieces of solid continued to be a source of annoyance.[25] Finally, on May 10, 1898 Dewar was able to liquefy hydrogen. Liquid air had already been prepared and the hydrogen started running from 175 atm at about 15 cubic feet per minute. His *Laboratory Notebook* entry for the day reads:

> Shortly after starting the nozzle plugged but it got free by good luck and almost immediately drops of liquid began to fall…and soon accumulated 20cc.. The H[ydrogen] was a clear transparent liquid well defined meniscus (even better seen than liquid air) showing no absorption spectrum and as long as the surroundings of the vessel were cool seemed to evaporate very slowly.

When Dewar immersed a sealed tube containing helium into the liquid hydrogen he "could see a liquid formed. This tube gave nothing when placed in liquid air."[26] He repeated the experiment 2 days later and found liquid hydrogen appeared "more brilliant than liquid air."[27] He tried to perform some experiments with liquid hydrogen and attempted unsuccessfully to measure its density. On the same day, May 12, 1898, Dewar presented his results at the Royal Society and asserted that he was able to liquefy both hydrogen *and* helium. At the same meeting, the President Lord Lister announced that Dewar had send him, on May 10, a preliminary communication *On the liquefaction of hydrogen and helium* where he mentioned that he had been able to liquefy hydrogen in quantity and that by means of liquid hydrogen he had managed to liquefy helium as well. The contents of Dewar's formal presentation at the Royal Society are similar to part of what was said in the Friday Lecture at the Royal Institution on December 15, except for his claim that he had liquefied helium as well.[28] Dewar in his experiments with the Bath gas had surmised that the volatility of hydrogen and helium would be close together. He had helium in a sealed bulb with a narrow tube attached and when the tube was immersed in liquid

hydrogen, then a distinct liquid was seen to condense. Such was not the case when liquid air instead liquid hydrogen was used. Of course, what he had not realised at the time was that he did not have pure helium and what was liquefied was hydrogen. As in all the papers dealing with the liquefaction of hydrogen Dewar, was at best, vague about the details of his apparatus.[29]

Liquid Helium

Kamerlingh Onnes' early researches between receiving his doctorate in 1879 and his appointment to the Chair of Experimental Physics at the University of Leiden in 1882, defined, to a large extent, his program in low temperature physics. His doctorate titled "New proofs for the axial rotation of the earth" was the first exact mathematical treatment of Foucault's pendulum. In 1881 he published his "General Theory of the Fluid State." It was a purely theoretical work continuing from where van der Waals left in 1873 when in his doctorate he had proposed his celebrated equation of state. In this work, Kamerlingh Onnes among other things, would reformulate the law of corresponding states which, in 1880 van der Waals had derived from his equation of state.[30]

Right after his appointment, Kamerlingh Onnes embarked on a program to determine the various parameters of the equation of state and test some of the implications of both the work of van der Waals and himself for as many substances as possible in as wide a temperature range as possible. High temperatures were not suitable for the study of the equation of state, since in those ranges there were chemical changes, and, so, he turned towards the low temperatures.

The first indications that Dewar was starting to seriously consider the liquefaction of helium were in 1901. He was experimenting with the permanent gases and was trying to devise ways to separate them in large quantities through liquefaction techniques. On January 31, 1901 he wrote in his *Laboratory Notebooks* that it "looks as if little condensation of Neon took place." By the beginning of August 1901 he had completed the preliminary experiments to determine the minimum amount of hydrogen and air required for liquefaction by continuous circulation. He realised that he had a long way to go and noted that "altogether the experiment looks promising for the Helium although there are enormous difficulties to be overcome before this liquid can be obtained."[31] He had to improve the yield of his liquid hydrogen *and* isolate sufficient helium. In 1903 he felt he had enough of helium and decided to have a go. He failed and wrote to Kamerlingh Onnes "It is however a very complicated and risky business as you well know. I have already lost 1 cubic foot of helium by the breaking of vacuum vessels during the course of its circulation at liquid air temperatures and I dread any repetition of the disaster." He was also feeling very weak and could not concentrate on his reasearch. Uncharacteristically, bad health brought about a confession "I only wish that I had again the gift of growth so that I might begin my scientific career after a training in your Dutch school of science."[32]

Kamerlingh Onnes wrote back of how sorry he was about the "fearful disaster with helium". Very diplomatically, he wondered whether the disaster could be avoided if Dewar had used a compressor developed in 1900[33] at Leiden for pure and costly gases. He wished he could be of some help for Dewar's "splendid attack on helium, the boldest attack that can be dreamt of in low temperatures."[34]

By the summer of 1904, it appeared that Dewar was giving up.

> I had hoped by this time to have settled the helium question but bad health has been a great obstacle. In my work I have never been able to do anything unless substantially with my own hands. In pioneering work assistants are a waste.

And, then, the most unexpected of Dewar's otherwise predictable responses. "If my health breaks down then I would hand the helium to you. This is the best thing I could do for Science."[35]

One year later, Kamerlingh Onnes made an unwelcome demand. He had written to the authorities of the Bath springs inquiring about the possibility of providing him with helium. They referred him to Dewar and Kamerlingh Onnes asked him whether he could have large quantities of the impure gas out of which he hoped, after an estimated 2 years of preparatory work, to have everything arranged in an "unobjectionable way."

At this point Kamerlingh Onnes needed the gas to start his measurements of its isotherms. Van der Waals' theory and his own experience had convinced him that the determination of the isotherms was the most reliable way of estimating the critical point.

> I am sure you will sympathize with my attacking the problem of the isotherms. The determination of the isotherms is the rational way to get the data for calculating the critical points… and exact determinations of isotherms is just in my line of accurate measuring work.[36]

Dewar told Kamerlingh Onnes that the helium supply of the Bath gas had been overly exaggerated. "I have in my own way been engaged on this subject for years and after many misfortunes, and with no little expenditure I have been unable to accomplish my specific object. We both want the same material in quantity from the same place at the same time and the supply is not sufficient to meet our great demands." He promised to give some to Kamerlingh Onnes as soon as he was able to secure some for himself. But "things are in a sad way with me. For the last 4 months I have been seriously ill and quite unable to do any work, so that all research has been halted."[37]

In March 1908, Kamerlingh Onnes, thought he had solidified helium and it appeared that there was no liquid state. He had made a mistake. What he thought was helium had, first, turned, into a white syropy liquid and then solidified. He was confident that the difficulties were caused by hydrogen impurities. The tube had broken and Kamerlingh Onnes could not say with certainty anything about the nature of the cloud. In his apologetic letter to Dewar he wrote that "the question of condensing helium remains an open one." But he was now more confident than ever that these experiments taught him "how careful one has to be in reaching conclusions from the appearing or not appearing of a cloud by expansion."[38]

Dewar, in the meantime had dutifully written to *The Times* about Kamerlingh Onnes' success, so that the world would be informed "that you had succeeded where I had failed." Now he tried to console Kamerlingh Onnes by telling him that everyone was aware of how difficult such an experiment was and "we can all be mislead." Upon receiving Kamerlingh Onnes' letter, Dewar sent a letter to *The Times* explaining the cause of the presence of solid matter in the experiments, and that put the matter all right with the public. "I wish all scientific men were as magnanimous as yourself in making immediate correction of faulty inferences from experimental data they had reason to believe at that time was correct." By now Kamerlingh Onnes' experiments suggested that the critical point of helium would be around 5 degrees absolute, and if this were correct then Dewar felt that they were still a long way from having static helium in a liquid or solid form. It would have been nearly impossible to continue experiments at the Royal Institution which had no money to pursue such very expensive experiments, since "it had no endowments to draw upon. My health is improving but at my age one must anticipate a gradual or sudden sunset."[39]

In less than 3 months after discovering his blunder, Kamerlingh Onnes would liquefy helium. The first letter he wrote was to Dewar

> All the time the helium apparatus remained perfectly clear! Nobody but you and myself know what this means…].It was a good thing to have trusted the theory of Van der Waals and my isotherms, which would only be obtained after many years of preliminary work, but have proved efficient.[40]

Immediately afterwards he sent Dewar a handwritten note about the details of the liquefaction.[41]

> Notes on the work leading to the liquefaction of helium
>
> Group A leads to
> I. the course of refrigerators giving liquid air.
> II. the cycle for continuous liquefying of hydrogen
> Group B leads to the isotherms of helium.

> Group A I. The object was to arrange permanent baths for accurate measurements by the cycle method. From different reasons high purity of the gas in the cycle giving the bath of liquefied gas was necessary. There was special attention paid to make only a minimum of additional gas circulate in the cycle, and to ensure that the gas would not become contaminated in prolonging the work. The liquefied gas of one cycle being <u>also</u> available for cooling the compressed gas in a cycle with a less cercible gas than was obtained in a cascadfe. Regeneration makes that there is obtained a very great economical effect even at the lowest temperature.
>
> From the renegeneration cascade there was in work in 1892 the chloromethyl and ethylene cycle, the oxygen cycle only so far that succeeded in pouring off some liquid oxygen at ordinary pressure. In 1894 this cycle was in good working order and the

counters of the permanent liquid oxygen bath brought to more than 1/4 of a liter. This without Dewar's glasses. The third cascade was ameliorated by the by,… and accordingly by a fourth (open) cycle for liquefying and evaporating air was added. According to the great economical effect mentioned above this cycle gives 9 liters of liquid air per hour, so the 75 liters used in the Helium experiment could be obtained without difficulty. {Last sentence underlined by Dewar}

II. As soon as the oxygen cycle was completed the hydrogen cycle was taken to hand. The work was done (as published in 1896) according to the same theorem that has now been laid at the foundations of the method of liquefying helium. Then extreme purity was still more required if continuous work should be arrived at, and continuous work was of the utmost importance for the liquefaction of helium. The cycle worked well in 1906. It gives 4 liters of liquid hydrogen per hour. The apparatus has been provided since 1906 with an arrangement to easily prepare a store of exceedingly pure hydrogen gas. This arrangement is not yet described but it is very simple in principleas it depends on freezing out the impurities by evaporating pure hydrogen in circulation. this make it possible to take off 20 liters of liquid hydrogen as well, used in the helium experiment, though the apparatus is relatively small. {Last two sentences underlined by Dewar}.

Group B. All the importance of having at disposal such an efficient cooling at –259° appeared from the determination of the isothermals as they put the Boyle point at –259°. Long before the helium was known there was worked out isotherm determinations at very low temperatures. Object was then hydrogen. There were arranged cryostats, manometers, piezometers expressly for this and an elaborate series of investigations of low temperature determinations was [one word could not be read] taken. For only by very accurate work in comparing(???) the deviations of the law of Boyle Charles for this nearly permanent gas can be derived its critical temperature.

Research Traditions and Laboratory Practices

Let me raise a number of issues with respect to the liquefaction of helium.

Ever since Francis Bacon and Robert Boyle investigated the nature of cold, the instrumentation and the techniques developed were of paramount importance in producing cold, and much more importantly, in *sustaining* it for a long time. Bacon's complaint that cold cannot be produced as easily as heat, highlighted the

innate asymmetry between heat and cold: "we cannot obtain [cold] in any great degree, for furnaces of fire are far hotter than a summer sun, but vaults and hills are not much colder than a winter's frost." Then it was found that evaporation induced a drop in temperature and by 1844 Faraday had liquefied all the gases except oxygen, nitrogen and hydrogen (and the three compound gases carbonic oxide, marsh gas and nitric oxide). The Joule-Thomson effect, the development of thermodynamics, Dewar's vacuum flask, the improvements by Linde and Hampson to the cold producing machines by incorporating the regenerative process, the researches in thermometry and constant calibration had greatly facilitated the lowering of temperatures. By 1898, Pictet in Geneva, Cailletet in Paris, Olszewski and Wroblewski in Cracow and Dewar in London had managed to liquefy all the gases, including the newly discovered argon – all the gases, that is, except helium. And it was thought that with helium, it was, *in the last analysis* again a question of starting from very low temperatures, expanding it under pressure, regenerating it and doing the whole cycle again. It was believed that a large amount of liquid hydrogen, a well insulated apparatus and sufficiently large amount of helium gas would do the job.

Dewar's main directions of his research during the period preceding the liquefaction of helium was the attainment of high vacua by cooled charcoal, the measurement of absorption rates of various gases – including the inert gases – by charcoal and using these techniques to separate helium. He did perfect the construction of metallic vacuum vessels and of thermometers for the reliable measurements of the very low temperatures. Even as late as 1907, in his lecture on high vacua and helium at low temperatures, he talked about his difficulties in isolating helium. He could not produce liquid hydrogen in any particularly large quantities, he was uncertain as to the critical point of helium and referred to the "*probability* of helium being liquefiable" in responding to Olszewski's view that helium had such a low critical point that it may not be liquefiable.[42] And in his talk a year later, in 1908, on the "nadir of temperatures and allied problems" there was not much more to report than some additional data about absorption.

Kamerlingh Onnes had other priorities in mind and was quite explicit on the character of his experimental strategy. In the preface of his thesis, he quoted parts of Helmholtz's memorial lecture on Gusav Magnus

> It seems to me that nowadays the conviction gains ground that in the present advanced stage of scientific investigation only that man can experiment with success who has a wide knowledge of theory and knows how to apply it; on the other hand, only that man can theorize with success who has a great experience in practical laboratory work.

He turned all his energy to realize such a program. During the same 10 years, that is since 1898, Kamerlingh Onnes had developed a multi-pronged research activity. By 1908 there were 36 articles for the two directions in instrumentation and thermometry. He had completed the exhaustive measurements on the isotherms of diatomic gases and their binary mixtures. They involved the study of hydrogen which he was able to liquefy in 1906, producing about 4 l/h. He had also done

systematic measurements of the isotherms for the monatomic gases together with a host of other papers on the determination of the Ψ-surface of van der Waals.

Though the despotic manner by which Kamerlingh Onnes ran his Laboratory could have been envied even by Dewar himself, he was fully conscious that he was not particularly innovative with instrumentation and that able technicians would be indispensable for his program. The first person to be appointed to the new post of *Instrumentmaker* at the Physical Laboratory in 1899 was G. Flim and in 1901 Kamerlingh Onnes established the School of Instrument-makers and Glass-blowers to be run by the Physical Laboratory of the University of Leiden. It was to play an absolutely decisive role for all future work in low temperatures for a long time even after Kamerlingh Onnes' death in 1926. His results were published by the *Communications from the Physical Laboratory of the University of Leiden* which he had founded in 1886 and where all the articles were mainly in English and very few in French, and were characterised by the analytic descriptions of all the experiments performed and the detailed drawings of the apparatus. He was not involved in any controversies except in a minor dispute with de Heen concerning the method for determining critical points.

Dewar did not strive for any moves to bring forth any institutional changes necessary for pursuing work in low temperatures. Perhaps Dewar's greatest achievement in Cambridge was to convince, George Liveing, the Professor of Chemistry, that there were other things a Professor of Chemistry should be doing except teaching and chairing committees. They started their collaboration in spectroscopy in 1877, 2 years after Dewar's appointment to the Jacksonian Chair and it lasted until 1904 a few years before Liveing's retirement. But even this work did not involve too many institutional changes and none was explicitly asked in the report signed by both Liveing and Dewar concerning the arrangements in the new Chemical Laboratory being built in the early 1890s. It did involve buying a spectroscope – being in fact the only instrument Liveing did not leave to the Chemical Laboratory upon his retirement in 1908, but asked to have it at home for "private researches" – which he never did. Dewar's monastic character found a refuge at the Royal Institution. His successes at low temperatures were due more on his remarkable dexterity and the almost saintly forbearance of his staff than because any institutional changes he had brought about for the overall support of his work. Everything done at his Laboratory was closely guarded and, it is quite amazing, that in none of his papers is there a detailed description of the hydrogen liquefier and at his demonstrations the whole apparatus was covered except the part where liquid hydrogen was seen to be accumulated.[43]

The issue of the liquefaction of helium should not be dealt within a *problematique* related to issues of priority, but rather by examining the actual laboratory practice of the two protagonists. Kamerlingh Onnes strongly articulated the need for experiments in low temperatures as a *necessary* outcome of the study of a series of theoretical issues of the work of van der Waals. All the experiments were planned as part of a long term program to extend the implications of the equation of state proposed by van der Waals which together with the law of corresponding states formed

the theoretical framework that, in effect, determined the characteristic themes and trends of low temperature physics at Leiden. It was this framework which provided the possibility for developing his agenda for a program in low temperature physics. Through his own theoretical work, Kamerlingh Onnes (re)assessed the role of the equation of state proposed by van der Waals. He regarded the equation of state as providing an underlying organising principle for a framework within which it would become possible to classify, compare and study substances, thus, achieving a taxonomy of substances much in the spirit of the periodic table, but in this case with respect to their *physical rather than chemical behaviour.* To facilitate these researches in 1901, he proceeded to develop in series the original equation of van der Waals. The equation introduced the virial coefficients which depended on temperature and an expression for each coefficient was given. On the whole, the proposed equation had 25 parameters, with which it became possible to provide even better descriptions of the experimentally determined values than it was the case with van der Waals' parameters.[44]

Dewar had not adopted an analogous framework and the success of his program depended almost exclusively on his admittedly amazing manipulative ability of the apparatus. He never embarked on the measurement of isotherms, despite the large number of papers reporting the values of various physical parameters at low temperatures. Though Dewar in his 1902 Presidential Address to the British Association for the Advancement of Science showed to be fully knowledgeable about van der Waals' work and its significance, considering it in fact as important an advance as Carnot's cycle, there is no evidence that Dewar incorporated in his researches van der Waals's work *or any other theoretical schema.* It appears that the liquefaction of helium was, for Dewar, a challenge to be met solely by trying to devise *techniques* for lowering the temperature and purifying gases. He could not realise that the liquefaction of hydrogen showed the success *and at the same time the limitations* of such a "brute force" approach. It failed with the more tricky question of helium. And the lack of isotherm measurements meant that during his quest for ever lower temperatures, *he was not becoming as knowledgeable about the behaviour of gases at these low temperatures*. It was the measurement of isotherms in Leiden which was so decisive in comprehending such a behaviour. Dewar's lack of interest in the various issues of kinetic theory and thermodynamics being debated at the time is evidenced not only by the absence of a single theoretical paper of his, but, more significantly, by the way his papers were written where any reference to theoretical results was cursory.[45] In fact, his papers did not even provide any clues about his position on the theoretical issues being debated at the time. Even his extensive references to Tait were exclusively about the latter's experimental work. One cannot help wondering whether mentioning Wroblewski's isothermal measurements whenever he refers to the significance of van der Waals's equation for determining critical points, was more strongly motivated by Dewar's intense dislike of Olszewski, rather than by a wish to acknowledge Wroblewski's measurements – especially since these very experiments were performed by Wroblewski after the breakdown of his collaboration with Olszewski.[46]

Conclusion

Perhaps the most interesting historical issue raised by both the argon case and the liquefaction of helium is the difference in outlook among the different groups of physicists and chemists, concerning the intermediate region of physical chemistry which was being established during the end of the nineteenth century and the beginning of the 20th. Dewar never transcended the view about physical chemistry as a way of adopting physical techniques for chemistry. This, at the beginning, appeared, a convenient and promising approach. Eventually, however, an insistence on this outlook, led Dewar's laboratory practice in problems related to this new hazy area of physical chemistry, to a deadlock. Dewar's researches concerning argon and the liquefaction of hydrogen and helium took place during a time when physical chemistry was articulating its own autonomous language with respect to both physics and chemistry, when it was charting its own theoretical agenda and formulating its own theoretical framework. It was the time when the disciplinary boundaries were drawn and re-drawn. The differences between the laboratory practices of Rayleigh, Ramsay, Kamerlingh Onnes and Dewar were suggestive of the way each chose to map the undefined and undelineated middle ground of physical chemistry. Questions were being asked as in whose domain physical chemistry was. Was it an activity for physicists or chemists? How would the boundaries be drawn, what would be the methodological priorities, the ontological commitments and, above all, what would be the character and extent of the practitioners' allegiances to physics and chemistry? These issues which bore an immediate relation to the whole question of the disciplinary status of physical chemistry would be discussed and disputed well into the interwar years, even after the successes of quantum mechanics in chemical problems.

The discovery of argon by Rayleigh and Ramsay and the liquefaction of helium by Kamerlingh Onnes, were not simply a triumph of a superior technique and improved instrumentation. Though these discoveries express the ability of the protagonists to assimilate such techniques and instrumentation in the emerging disciplinary framework, the simultaneous deadlocks of Dewar displayed all the characteristics of going amiss – misplaced priorities in the design of experiments which were conditioned by the overall framework within which Dewar functioned. What prevented Dewar from realizing the existence of an inert element in the atmosphere and liquefying helium, were not the various technical errors he made. Nor were the researches of Rayleigh, Ramsay and Kamerlingh Onnes free of errors. Dewar's deadlocks precipitated because of his refusal to appropriate the new practices and to adopt the new theoretical culture which were being articulated by the emergence of physical chemistry. And despite the interesting results he derived in his experimental researches with nitrogen, hydrogen and, to a certain extent, helium, Dewar, as it was realized a short while after these developments, was going amiss.

Acknowledgements I wish to thank the editors of the volume, and especially, Giora Hon for their helpful suggestions in writing the paper. I also thank Libraries that permitted me to use material

from their archival collections: The Library of Hanscomb Air Base, at Bedford Massachusetts, where Lord Rayleigh's Papers are kept, the Library of University College, London, where the William Ramsay Papers are kept, the Library of the Royal Institution where the James Dewar Papers are kept, and the Library of the Boerhaave Museum, Leiden where the H. Kamerlingh Onnes Papers are kept.

Notes

1. Rayleigh, "Argon" Proceedings of the Royal Institution, 1895, 14, 524–538.
2. Rayleigh's Notebooks are among his Papers kept at the Hanscomb Air Base, Bedford, Massachusetts, USA.
3. James Dewar, The Times, London, August 18, 1894.
4. James Dewar's Laboratory Notebooks are in the Dewar Archives at the Royal Institution, London.
5. Dewar's Laboratory Notebooks. Entries for August 9; August 14; November 21; November 27; November 29; December 3; December 14; December 20, 1894.
6. James Dewar, "The relative behaviours of chemically prepared and of atmospheric nitrogen in liquid state" Chemical News, December 6, 1894, 222–225. See also Dewar's Laboratory Notebooks, entries throughout November 1894.
7. The Times, December 7, 1894. The summary of the discussion to Dewar's paper was most probably written by Dewar's most fanatically, Henry Armstrong.
8. Henry Armstrong, Chemical News, December 6, 1894 as reported to the Chemical Society, London.
9. A. Rucker, Nature, February 7, 1895, on p. 337.
10. Lord Kelvin, Chemical News, December 14, 1894, pp. 291–292.
11. Rayleigh, W. Ramsay "Argon, a new constituent of the atmosphere" Philosophical Transactions, 1895, 186A, 187–241.
12. W. Ostwald to W. Ramsay December 24, 1894. William Ramsay Papers University College Library, London.
13. George FitzGerald to William Ramsay December 14 and December 20, 1894.
14. FitzGerald to Ramsay December 28, 1894.
15. FitzGerald to Ramsay January 8, 1895.
16. Ramsay to A. Smithells March 11, 1895.
17. A. Smithells, British Association for the Advancement of Science, 1907, on p. 477.
18. J. Dewar, "Experiments in the liquefaction of hydrogen," The Times, September 1, 1894.
19. Wroblewski had in January 1884 found that hydrogen cooled to the boiling point of oxygen and expanded quickly from 100 atm to 1 atm showed the same appearance of sudden ebullition as Cailletet had seen in his own experiments. Soon afterwards Olszewski expanded hydrogen from 190 atm previously cooled with oxygen and nitrogen boiling in vacuo. Olszewski noted in 1884 that when the gas was expanded to 40 atmospheres there were colourless drops running down the tube. Wroblewski could not confirm these results and he called his hydrogen a ''liquide dynamique.'' He thought that the only way he could get static liquid hydrogen was only if he could use hydrogen gas as a cooling agent. For details see K. Gavroglu, Y. Goudaroulis, Methodological Aspects of Low Temperature Physics: Concepts out of Contexts, (Dordrecht: Kluwer Academic Publishers, 1989).
20. William Ramsay, System of Inorganic Chemistry, (London: Unwin, 1904), p. 28. Measurements by Wroblewski of the isotherms of hydrogen had led him to the conclusion that hydrogen should have a critical temperature of –240 degrees, critical pressure of 13.3 atm, critical density of 0.027, and the boiling point was determined to be around –250 degrees. It should be remembered that Dewar wanted to have liquid hydrogen at its boiling point in an open

vacuum vessel. This implied a host of problems and, it was, of course, a much more difficult task than observing liquid hydrogen in a tube under pressure –and, of course, at a higher temperature.

21. James Dewar, "New Researches on Liquid Air", Proceedings of the Royal Institution, 1896, xv, 133–146, on p. 144.
22. Ibid.
23. Dewar, Laboratory Notebooks, entry for February 25, 1898. See note 5.
24. Ibid., entry for March 25, 1898.
25. Ibid., entry for April 25, 1898.
26. Ibid., entry for May 10, 1898. On May 12 he repeated the same experiment and until May 27 he had a succession of failures due to the breaking of the vacuum vessels.
27. Ibid., entry May 12, 1898.
28. J. Dewar, "Liquid Hydrogen", Proceedings of the Royal Institution, 1899, xvi, 1–14 and 212–217.
29. For details of the apparatus see K. Gavroglu, "The myths of low temperature physics", European Journal of Physics, 1993, 21, 171–190.
30. K. Gavroglu, "The reaction of the British Physicists and Chemists to van der Waals' Early Work and the Law of Corresponding States" Historical Studies in the Physical and Biological Sciences, 1990, 20, 199–237.
31. Dewar, Laboratory Notebooks, entry for August 7, 1901.
32. Dewar to Kamerlingh Onnes May 7, 1903. Heike Kamerlingh Onnes Papers, Boerhaave Museum, Leiden, the Netherlands.
33. H. Kamerlingh Onnes, "Methods and apparatus used in the cryogenic laboratory. II. Mercury pump for compressing pure and costly gases under high pressures" Communications from the Physical Laboratory of the University of Leiden, No. 54, 1900.
34. Kamerlingh Onnes to Dewar December 29, 1903. Dewar Papers, London.
35. Dewar to Kamerlingh Onnes January 5, 1904. Kamerlingh Onnes Papers, Leiden.
36. Kamerlingh Onnes to Dewar June 8, 1905. Dewar Papers, London.
37. Dewar to Kamerlingh Onnes July 15, 1905. Kamerlingh Onnes Papers, Leiden.
38. Kamerlingh Onnes to Dewar April 7, 1908. Dewar Papers, London.
39. Dewar to Kamerlingh Onnes April 15, 1908. Kamerlingh Onnes Papers, Leiden.
40. Kamerlingh Onnes to Dewar July 20, 1908. Dewar Papers, London.
41. This report is found in the Dewar Papers at the Royal Institution, and there is no copy of it in the Kamerlingh Onnes Papers at Boerhaave Museum.
42. J. Dewar, "Studies in high vacua and helium at low temperatures" Proceedings of the Royal Institution, 1907, XVIII, 747–756, quote on p.748.
43. K. Mendelssohn in The Quest for Absolute Zero, (London: Taylor and Francis, 1977) also makes this observation. Nevertheless, the apparatus used to liquefy hydrogen can, in fact, be reconstructed by the published reports and, especially, by Dewar's descriptions in his Notebooks, (March 14, 1898 – January 15, 1899). See note 29.
44. H. Kamerlingh Onnes, "Expression of the equation of state of gases and liquids by means of a series", Communications from the Physical Laboratory of the University of Leiden, No. 71, 1901.
45. It is interesting to note in this respect the extensive correspondence between Dewar and D.W. Dickson. The latter appears to be an alter ego of Dewar for all the theoretical calculations, and relevant discussions. See especially letters to Dewar from Dickson dated March 24, 1895; January 13, 1900; April 24, 1902; April 29, 1902; June 24, 1902; January 6, 1903 and January 8, 1903. Dewar Papers, London.
46. This is even more so, if we realize that Dewar was much more at ease – and had used successfully – Trouton's rule for determining critical points.

Part IV
Instrumental Artifacts

Distinguishing Real Results from Instrumental Artifacts: The Case of the Missing Rain

Wendy Parker

Introduction

A striking feature of nineteenth century science is its unprecedented concern with systematic and precise observation of the natural world (Cannon 1978; Miller 1986; Daston 1995). A fine illustration of this can be found in the field of meteorology, where the nineteenth century brought not just a dramatic increase in the number of stations at which regular observations of weather conditions were made, but also the organization of those stations into coordinated national observing networks. As these networks were established and expanded, scientists contemplated which instruments to use, where to locate them, how often to make observations, and how to record and communicate those observations.

When it came to measuring rainfall, an especially difficult question concerned the appropriate height at which to place a rain gauge. It was well known that gauges located high above the ground often caught significantly less rain than gauges placed near the ground. But the cause of this height-catch difference[1] remained a vexing mystery. Do elevated rain gauges give inaccurate readings? Or is there simply less rain present at elevated locations? If there is less rain, can elevated gauge readings be adjusted to ground-level readings—the readings that matter for practical purposes, such as agriculture? During the nineteenth century, the answers to these questions were hotly debated by amateurs and eminent scientists alike. Sir John Herschel himself remarked that "no more interesting problem can fix the attention of the meteorologist" than that of explaining the height-catch difference (Herschel 1861, 105).

This paper outlines scientists' investigation of the height-catch problem from its discovery in the middle of the eighteenth century to its resolution in the late nineteenth century. Ultimately, scientists came to agree that the height-catch difference reflected the "imperfect collection" (Symons 1882) of rain by elevated gauges. Both the gauge itself and the building on which it was located would disrupt the flow of the wind, and the resulting wind patterns would carry past the gauge some

W. Parker(✉)
Department of Philosophy, Ellis Hall 202, Ohio University, Athens, Ohio, 45701, USA
e-mail: parkerw@ohio.edu

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_10,

of the raindrops that otherwise would have landed in it. In today's terminology, the height-catch difference was found to be an artifact of the design and placement of the measuring instrument; readings from elevated rain gauges contained systematic errors due to the action of the wind. Yet it took more that a century for scientists to settle on this conclusion. In the interim, it was widely believed that the height-catch difference instead reflected a *real* difference between the amounts of rain present at elevated and near-ground locations.

A closer look at the story of the height-catch problem will bring to the fore a number of interesting points about the nature and significance of instrumental artifacts. First, it will show how difficult it can be to determine whether results are artifactual, even when the measuring instrument is as simple as an ordinary rain gauge. In addition, it will illustrate that sometimes a surprising observational result will be classified as an instrumental artifact only after its *cause* has been discovered; in the case of the height-catch difference, scientists would categorize elevated gauge readings as erroneous only after they agreed that wind effects were responsible. Most importantly, it will demonstrate that the recognition of an instrumental artifact can have both practical and theoretical significance: resolving the height-catch problem would change not only what counted as an accurate rainfall measurement but also how scientists conceived of the rainfall process itself.

The second section of the paper will discuss the discovery of the height-catch problem and the subsequent proliferation of hypotheses about its cause. According to the leading hypothesis, the height-catch difference was a natural consequence of the growth of raindrops by condensation during their descent. The third section of the paper will describe the influential observational study that cemented support for this condensation hypothesis in the 1830s. The fourth section will then focus on several important attacks on the condensation hypothesis, some of which simultaneously drew attention to the possibility that winds might be responsible. Yet even when the condensation hypothesis was abandoned in the 1860s, investigators did not immediately conclude that the height-catch difference reflected some sort of error. The fifth section of the paper will discuss the gradual accumulation of evidence that eventually did persuade scientists that elevated rain gauges exposed to the wind are especially poor collectors of rain. The significance of scientists' recognition of wind effects on gauge catch will be reviewed in the concluding section of the paper.

Why Do Elevated Rain Gauges Catch Less Rain?

William Heberden, a well-known physician and Fellow of the Royal Society, is usually credited with the discovery of the height-catch difference. Around 1765, Heberden noticed that one London rain gauge almost always caught more rain than another (Heberden 1769, 359). After satisfying himself that these two gauges were well-constructed and ruling out "every probable cause" of interference with the measurements, he concluded that the difference "did not appear to be owing

to any mistake, but to the constant effect of some circumstance, which not being supposed to be of any moment, had never been attended to" (ibid.). Noticing that the gauge that caught less rain was located very high above the ground, he conjectured that height might be the "circumstance" somehow responsible for the difference.

Heberden decided to test this idea. He set up two rain gauges in the vicinity of a house, placing one above the highest chimney and the other on the ground in the garden. Again, he found that the gauge placed high above the ground usually caught less rain than the one at the ground. Would the difference in catch increase if a gauge were placed at a height much greater than previously tried? A third gauge was situated, this time on the roof of Westminster Abbey (about 200 feet above the ground), and the monthly rain totals measured by all three gauges were recorded for 1 year, from 7 July 1766 to 7 July 1767. Steps were taken to ensure that the gauges gave readings that were as accurate as possible—a special apparatus limited evaporation, and gauge placement was chosen in a way that avoided shadowing by nearby buildings (ibid. 360).

The rainfall amounts received by the three gauges confirmed Heberden's previous finding—over the course of the year, the gauge on Westminster Abbey caught less rain than the gauge on the house, which in turn caught less than the gauge near the ground. The differences were not small, either: the highest gauge caught only 12.1 inches of rain, just over half of the 22.6 inches measured by the ground gauge (see Fig. 1). Convinced that elevated gauges consistently caught less rain than gauges placed near the ground, Heberden noted that it was important to bring attention to this difference "in order to prevent that error, which would frequently be committed in comparing the rain of two locations without attending

		Below the top of a houſe.	Upon the top of a houſe.	Upon Weſtminſter Abbey.
		inch.	inch.	inch.
1766 from the 7th of July to the end		3,591	3,210	2,311
	Auguſt	0,558	0,479	0,508
	September	0,421	0,344	
	October	2,364	2,061	1,416
	November	1,079	0,842	0,632
	December	1,612	1,258	0,994
1767	January	2,071	1,455	1,035
	February	2,864	2,494	1,335
	March	1,807	1,303	0,587
	April	1,437	1,213	0,994
	May	2,432	1,745	1,142
	June	1,977	1,426	1,145
from the 1ſt of July to the 7th		0,395	0,309	
		22,608	18,139	12,099

Fig. 1 Rainfall observations made by William Heberden (after Heberden 1769, 361)

to this circumstance" (ibid.). While admitting that he was unable to identify the cause of the "extraordinary difference" in catch that he had discovered, Heberden closed his report with some speculation on the matter. He suggested that "some hitherto unknown property of electricity" was probably at work, since "this power has undoubtedly a great share in the descent of rain" (ibid. 362).

Electricity was a popular explanatory catch-all at the time, suspected to be involved in the production of a host of meteorological phenomena, including the growth of raindrops.[2] In addition, at least two other processes by which falling raindrops might grow had been proposed: the "gathering up" of water from the surrounding air and the collection of smaller raindrops, which fall more slowly than larger ones (see Middleton 1965, 97).[3] The former process is similar to what would later be known as condensation and will be referred to here using that terminology. (The latter is similar to what would later be known as coalescence but will not be important in this discussion.) Under the influence of either of these two processes, a falling raindrop should grow larger the further it falls, assuming that there is both negligible evaporation from the drop and continual availability of water in the form of vapor or droplets.

It did not take long to find in the condensation process a rough-and-ready explanation of Heberden's rainfall observations: if cold raindrops collect water vapor from the air throughout their fall, then drops reaching the lower gauge will have grown larger (since they have fallen farther) than drops collected by the elevated gauge. Perhaps the earliest suggestion that this process might contribute to the observed height-catch difference was made by Benjamin Franklin, in 1771.[4] However, Franklin also raised an immediate concern: Wouldn't falling drops quickly exhaust the available dissolved water in the atmosphere (see Franklin 1771/1974, 156–157)?

Indeed, at the time of Franklin's speculation, experiments that cast doubt on the condensation hypothesis had already been conducted. After attending the December 1769 meeting of the Royal Society of London at which Heberden's report had been read, the lawyer and naturalist Daines Barrington decided to undertake an experiment of his own. He arranged for the construction of two similar gauges and for their placement at the same height above the ground at two different elevations above sea level, one near the top of a small mountain and the other about a half-mile distant, in a lower-lying area. With this placement, assuming that the two gauges received rain originating from the same height above the ground, rain reaching the lower gauge would have to travel through a much greater depth of atmosphere than rain collected in the upper gauge. Again, care was taken to avoid well-known sources of error; as Barrington put it, "precautions were also taken, that neither cattle, nor any other accident, should interfere with the experiment" (ibid. 295).

In late 1770, Barrington reported his findings to Heberden in a letter that was read the following June before the Royal Society. Barrington had found *no consistent difference* between the rainfall amounts measured by the two gauges. He concluded that the amount of rain caught in a gauge depends upon how close the gauge is to the ground (Barrington 1771, 296). As Heberden emphasized in a note appended to the published version of Barrington's letter, these observations spoke against all attempts to explain the height-catch difference in terms of the greater depth of

atmosphere through which drops reaching the lower gauge had fallen (Heberden 1771, 297).

As time passed, other possible explanations of the height-catch difference were also introduced. Some investigators suspected that the height-catch difference might be due to the interference of the wind. By 1812, Luke Howard, a chemist and amateur meteorologist, had noticed that when a gauge was placed on the windward side of a building's roof, some of the rain was blown over the top of the gauge and deposited beyond it (Howard 1812, 423).[5] Likewise, in 1819, H. Meikle implicated rain gauges themselves in the production of the height-catch difference, when he described an acceleration and vertical movement of air encountering a rain gauge:

> I can hardly pretend to give a complete solution of this well-known paradox [that elevated gauges catch less rain], but am disposed to think it is in some way owing to the obstruction which the gauge itself offers to the wind. Perhaps the winds being made to rush with greater rapidity, and a little upward in beginning to pass over the mouth of gauge, prevents the rain from falling into that part of it which is next the wind.
>
> (as quoted in Jevons 1861, 427)

In 1822, geologist Henry Boase also identified wind as a factor somehow connected to the height-catch difference. After measuring both wind speed and rainfall, he found that the difference between the amounts of rain received by his upper and lower gauges was "for some reason or other, proportioned to the velocity of the wind" (Boase 1822, 20). He conjectured that the observed difference in catch was due "chiefly to the whirl or eddy occasioned by the recoil of the gusts of wind striking on the sides of the building—an effect very visible in the disturbance of smoke issuing from chimneys during a high wind" (ibid.). According to these investigators, the height-catch difference might not be a real difference after all; perhaps elevated rain gauges simply fail to catch some of the falling rain, because eddying winds carry some raindrops past the gauges.

A different explanation, sometimes first attributed to M. Flaugergues (1819), focused on the extent to which falling raindrops followed non-vertical paths. According to this inclination hypothesis, raindrop paths are more slanted at the height of elevated gauges than at the ground, and raindrops following more slanted paths are further apart (in horizontal distance) than drops that fall more vertically; as a consequence, fewer drops fall into the mouths of elevated gauges than into those of ground-level gauges. While many supporters of this hypothesis insisted that it followed from simple geometric considerations, other investigators denied that it made any sense at all (see e.g. the extended discussion in *British Rainfall 1871*), and it is not clear what the inclination hypothesis was generally believed to entail about the actual amount of rain present at elevated locations. However, some investigators inferred from it that a given volume of rain would be spread over a greater horizontal area at an elevated level than at ground level. If so, then the depth of rain falling per unit area at an elevated location would be less than that at the ground, which is just what the gauges reported.

Thus, at least four possible explanations of the height-catch difference had been offered by 1830: the electrical hypothesis, the condensation hypothesis, the wind

hypothesis, and the inclination hypothesis. With no well-established theory of rain formation available at the time, nor any independent means of estimating rainfall at elevated locations, each hypothesis was judged by some investigators to give a plausible explanation of the height-catch difference.

Over time, however, some hypotheses garnered more support than others. The influence of electrical effects on rainfall intensity was rarely mentioned in later discussions of the height-catch problem (for an exception see Ingram 1871), and the electrical hypothesis seems to have undergone little development. Likewise, by the middle of the nineteenth century, several investigators had dismissed the inclination hypothesis as fallacious (Phillips 1833, 404; Herschel 1861, 104; Jevons 1861, 428). Nevertheless, because the inclination of falling rain is dependent upon wind speed, the inclination hypothesis would become entangled with the wind hypothesis and would survive in various forms to the end of the nineteenth century. Support for the wind hypothesis itself, while not negligible, was rather limited. The condensation hypothesis, despite its apparent incompatibility with Barrington's results, would remain the favored explanation of the height-catch difference until the middle of the nineteenth century (Jevons 1861; Herschel 1861; Pennant 1871; Middleton 1965). According to this hypothesis, the height-catch difference was a real difference, reflecting the presence of more rain at the ground than at elevated locations.

The Height-Catch Difference is a Real Difference

The height-catch problem was one of the first to receive attention and funding from the British Association for the Advancement of Science at York after the group's inauguration in 1831 (see Symons 1882, 41). The Association awarded a grant to William Gray and John Phillips, later professor of geology at King's College and at Oxford, who proposed to make a 3-year study of the problem.

Gray and Phillips acquired three nearly identical rain gauges and placed them at heights similar to those chosen by Heberden more than 60 years earlier. The mouth of the lowest gauge was only 2 inches above the ground, in a garden. The middle gauge was located on a nearby roof, approximately 44 feet above the ground, and the highest gauge was placed on the tower of York Minster at a height of about 213 feet (Phillips 1833).

In 1833, Phillips published results from their first year of measurements, along with some conclusions. He rejected the inclination hypothesis on the grounds that "the number of drops of rain which fall, under the joint influence of gravitation and ordinary wind, upon horizontal surfaces, will be, *cæteris paribus,* exactly the same at all elevations below the point from which the rain descends" (ibid. 404). He admitted that wind effects might be significant in some extreme cases but considered it "evident that in the majority of cases the effect of the eddying wind is quite unimportant" (ibid.). Instead, he favored the condensation hypothesis:

> ... the whole difference in the quantity of rain, at different heights above the surface of the neighbouring ground, is caused by the continual augmentation of each drop of rain from the commencement to the end of its descent, as it traverses successively the humid strata of air at a temperature so much lower than that of the surrounding medium as to cause the deposition of moisture upon its surface.
>
> (Phillips 1833, 410)

Phillips found what he considered to be substantial evidence for the condensation hypothesis. First, the hypothesis could be used to explain a number of patterns in the data. For example, an examination of monthly rain totals revealed that the height-catch difference was largest during the coldest months of the year. Why would this be? According to Phillips, the coldest months were also the most humid months.[6] More humidity would mean more water vapor available for condensation on the falling drop, and more condensation would increase the growth of drops during their descent between the upper and lower gauges. Thus, it was not surprising that the greatest height-catch difference occurred in the coldest months of the year. Several other patterns were similarly explained (see ibid. 410–411).

Even more impressive were the results of Phillips' mathematical analysis. Examining the data collected during the first year of his and Gray's study, he determined that the average height-catch difference over a season could be accounted for, to a first approximation, using a simple formula (ibid. 406):

$$d = m\sqrt{h}$$

where d is the percent reduction in gauge catch relative to ground catch, h is the height of the gauge above the ground, and m is a coefficient of proportionality that varies by season and locale. Eager to demonstrate that the formula had broader applicability, Phillips examined other published data as well, including Heberden's original height-catch data. He showed that the same (approximate) dependence on the square root of height was present and that, as expected, the coefficient m varied with the temperature of the season, taking its largest values during the coldest months (ibid. 407–408).

As new data became available, Phillips found that his simple formula was too simple. Unfazed, he eventually arrived at a more complicated formula that expressed the average height-catch difference over a period as a function of three factors: gauge height, mean annual temperature for the year in which the observations occurred, and mean temperature for the period for which the difference was to be calculated (see Phillips 1835). The values given by this formula matched reasonably well with the monthly averages calculated from Gray and Phillips' data, which led Phillips to conclude that the formula expressed "at least the nature of the proximate influential causes" of the height-catch difference (ibid. 176). Because Phillips believed that temperature could serve as a proxy for humidity (see Note 6), the condensation hypothesis seemed to him to be vindicated once again.

Support for the condensation hypothesis was further galvanized by Phillips' findings. At the same time, accepting the hypothesis meant accepting a particular conception of how rain develops during its descent – raindrops must grow quite

rapidly during the last few hundred feet of their fall. Phillips believed that this growth followed "some settled laws" (1833, 411) whose form he had begun to uncover through his mathematical analysis. Such laws, if found, might have had important practical implications for rainfall measurement. For instance, they might have been used to convert elevated gauge readings to ground-level readings, allowing observers to place rain gauges at whatever heights they found most convenient. Crucially, however, such conversions would not have been *corrections* to elevated gauge readings; according to the condensation hypothesis, there was actually less rain present at elevated locations than at near-ground locations.

Trouble for the Condensation Hypothesis

Not everyone was fully convinced by Phillips' analysis of the height-catch problem. Some investigators, while not challenging the condensation hypothesis directly, doubted that *only* condensation was contributing to observed height-catch differences. For example, Alexander Bache, professor at the University of Pennsylvania and great-grandson of Benjamin Franklin, found that "the effect of eddy winds … was by no means a secondary one" when it came to the collection of rain (Bache 1838, 25).

Bache began making height-catch observations in 1833, at the suggestion of a colleague who had heard Phillips' first report to the British Association (see ibid.). He discovered that a rain gauge placed on the windward side of a tower roof consistently caught less rain than a gauge placed on the leeward side, i.e. that even gauges placed at the *same* height above the ground differed in their catch when they were not equally exposed to the wind. Bache suggested that "currents of air deflected by the tower" on which the gauges were located might be responsible for this difference (ibid. 26).

Yet Bache apparently did *not* conclude that wind effects were responsible for the height-catch difference. Rather, he seemed to view the wind as something that *obscured* the true cause of the height-catch difference. He insisted that he "could not hope to deduce a law, nor to throw any light on the nature of the phenomena, until this disturbing action [of the wind] was got rid of" (ibid. 25). It appears that he, too, hoped to discover some law governing the growth of raindrops by condensation during their descent, if only he could eliminate the wind's interference.[7] Decades later, however, Bache's results would be seen as compelling evidence for the wind hypothesis.

Other investigators challenged the condensation hypothesis directly. The most devastating of these challenges came in the form of theory-based calculations, and the most influential of these was carried out by John Herschel, in 1857. He argued that condensation was "totally inadequate" to account for the substantial height-catch differences actually observed (Herschel 1861, 104). Making what he considered generous assumptions about the process of condensation, he estimated that it could produce less than 1/17th of the typical difference to be accounted for (ibid.).

It has been said that Herschel's calculations "demolished" the condensation hypothesis (Symons 1882, 43). While this may be an overstatement, it is clear that they did immediately become a central piece of evidence against the hypothesis.

Less influential at the time, but tremendously important in the long run, was a paper that appeared just a few years later by William Stanley Jevons, now known primarily for his work in logic and economics. In his 1861 paper, Jevons offered a comprehensive and creative assault on the condensation hypothesis and a forceful defense of the wind hypothesis. According to Jevons, the height-catch difference reflected not a real difference in the amounts of rain present at elevated and near-ground locations but rather the "erroneous nature of the rain gauge" and the "very unsuitable" placement of rain gauges on tall buildings, where they are exposed to high winds (Jevons 1861, 424).

Jevons offered several lines of evidence for his views. Like Herschel, he attempted to show by calculation the implausibility of the condensation hypothesis, but he claimed that Herschel had made overly generous assumptions; his own more realistic calculations revealed the hypothesis to be even less credible (Jevons 1861, 429–430). In fact, his analysis led him to conclude that "under no possible conditions will the increase [in the size of drops] within the last few hundred feet of descent be more than almost infinitesimal" (ibid. 430). The condensation hypothesis could be true only if "some of the best established facts of physical science" were overturned (ibid. 432).

Jevons' second line of attack focused on observations rather than theory. He argued that if condensation were responsible for the height-catch difference, then there should be a significant difference between the temperature of rain in the upper gauge and the temperature of rain in the lower gauge (due to latent heat release during condensation), yet no such difference was ever observed (ibid. 431). Likewise, if raindrops were growing substantially by condensation during their fall, then rain should appear to increase in intensity as it falls, but naked-eye observations only seem to reveal the opposite, i.e. that falling rain sometimes lessens in intensity, even to the point of ceasing (ibid. 432). Furthermore, Jevons noted, the condensation hypothesis was unable to explain certain features of height-catch data that had already collected, including Heberden's data for March 1767. Those data showed an especially large height-catch difference—the lowest gauge caught 1.8 inches of rain, more than three times the 0.58 inches caught by the highest gauge (see Fig. 1). Phillips, who did not find March to be an especially humid month, had recognized Heberden's March data as "very anomalous" (Phillips 1833, 408) but could offer no explanation. Jevons, by contrast, insisted that the result was easily explained using the wind hypothesis: " … March is in Europe the month in which strong, dry north-east winds and equinoctial gales most occur, the very circumstances under which we should expect the [upper gauge] results to be most erroneous" (Jevons 1861, 427).

Jevons derived his most ingenious evidence for the wind hypothesis from a set of "small experiments" that might be described as early wind tunnel experiments.[8] Using a small vessel into which he inserted objects representing rain gauges or houses, he studied how air moved when it encountered an obstacle. He described the set-up as follows:

> One end of the vessel communicates through a pipe with a chimney or an aspirator, so that a regulated current of air may be drawn through it, to represent on a small scale a section of the wind moving over the surface of the earth. The curves described by the currents of air are shown very distinctly and beautifully by simply holding a piece of smoking brown paper in the draught of air which is about to enter the glass vessel.
>
> (Jevons 1861, 424)

Jevons found from his experiments that air encountering an obstacle was accelerated over the top of the obstacle. He argued that " … rain-drops falling through such wind upon the windward part of the obstacle will be further apart, in horizontal distance, than where the wind is undisturbed and of ordinary velocity" (ibid. 433). From this, he concluded that some rain would be carried to the lee side of the obstacle and even beyond it, so that less rain fell onto the top of the obstacle. He speculated that the obstacles of most significance in creating such an effect were the houses and towers on which the gauges had been placed, but he also considered the rain gauge itself to be an obstacle sufficient for bringing about some deficiency in catch, especially at elevated locations, where wind speeds are usually higher than wind speeds near the ground (see ibid. 427–428, 432). Wanting to distance himself from the inclination hypothesis, Jevons emphasized that it was "the divergence of the rain-drops, owing to the varying velocity of the wind" that he proposed to be the cause of the height-catch difference (ibid. 428).

As further corroboration for his hypothesis about the effects of wind on gauge catch, Jevons noted that previous investigators, including Howard, Meikle, and Boase, had each come via different experiments to conclusions rather similar to his own (as discussed above).[9]

What were the implications for rainfall measurement? Jevons concluded that "all observations by rain-gauges elevated or exposed to wind must be rejected as fallacious and worse than useless" (ibid. 432). If the deficiency of rain in elevated gauges depends upon the speed of the wind, which can vary greatly over short periods of time, then it will be extremely difficult to correct the erroneous readings given by elevated gauges. Jevons instead advocated placing all gauges close to the ground, where wind speeds are relatively low. He estimated that the error due to wind effects would be small if gauges were placed with their mouths just 1 or 2 feet above the ground. But despite his forceful arguments, it would be at least two decades before Jevons' analysis of the height-catch problem – and his recommendations for the placement of rain gauges – would become widely accepted.

Rain Gauges are Imperfect Collectors

In 1860, meteorologist George J. Symons began publishing *British Rainfall*, an annual volume of meteorological observations taken at sites throughout the British Isles. Convinced of the practical and scientific value of knowing how much rain fell where and when, he would spend the next 40 years developing a network of mostly volunteer rainfall observers and trying to ensure that the rainfall measurements

made were as accurate as possible (see Symons 1865, 199; Mill 1901; Pedgley 2002). From the start, the height-catch problem was a source of great vexation to Symons, who had found that "gauges throughout the kingdom were at all sorts of elevations, from 90 feet downwards" (Symons 1865, 200). Wanting to compare rainfall amounts at different locations, Symons considered it "evident that some means must be devised" for adjusting the catch of elevated gauges "to what they would be if made at one uniform level" (ibid.). For that purpose, he helped to organize a number of observational studies investigating how rainfall varies with height.

One of the most frequently cited studies was conducted at Rotherham, Yorkshire, from 1865 to 1872. As usual, rain gauges were placed at a number of different heights above the ground. But measurements at Rotherham were made using several experimental rain gauges specially designed for investigating the relationship between wind, the angle of falling rain, and the height-catch difference. These experimental gauges included a tipping gauge, a five-funneled gauge, and other gauges with inclined mouths. The tipping gauge was designed to catch as much rain as possible by always both pointing into the wind and keeping its mouth perpendicular to the plane of the falling rain (see Mill 1901, 30). The five-funnel gauge consisted of one horizontal-mouthed funnel and four vertical-mouthed funnels, each emptying into individual measurement tubes; it was designed to allow estimation of both the angle of inclination of the falling rain and the compass direction from which it came (see Fig. 2).

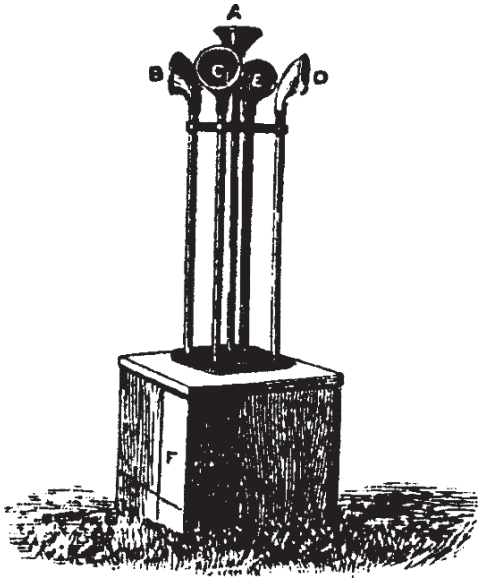


Fig. 2 Five-funnelled rain gauge (after Mill 1901, 30)

Preliminary measurements from the Rotherham experimental gauges seemed "to show clearly the influence of the wind's velocity, not only on the angle of the falling rain, but also on the ratio of the fall in elevated gauges to that on the ground" (Symons 1867, 31). When 4 years' worth of observations had been taken at Rotherham, Symons published a more detailed analysis. Using tables, graphs, and rather creative diagrams, he demonstrated that wind velocity, the angle of inclination at which rain falls, and the decrease in gauge catch with height were correlated in the data for daily, monthly, and annual time periods (e.g. see Fig. 3). While Symons' report was silent on what these data might indicate about the cause of the height-catch difference, the silence would not last for long.

The data collected at Rotherham and elsewhere appeared to be consistent with both the wind and inclination hypotheses, and a spirited debate over which was the true cause of the height-catch difference soon began. This debate played out primarily within the pages of another publication edited by Symons, his *Monthly Meteorological Magazine*, and was especially vigorous in 1871. During that year, there was such a volume of material submitted to the *Magazine* on the topic of the height-catch problem that Symons resigned himself to temporarily increasing the length of the publication in order to avoid encroaching on space usually devoted to other topics. Dozens of contributors weighed in on the matter. Despite its dismissal by Herschel and others, the inclination hypothesis seemed still to have many

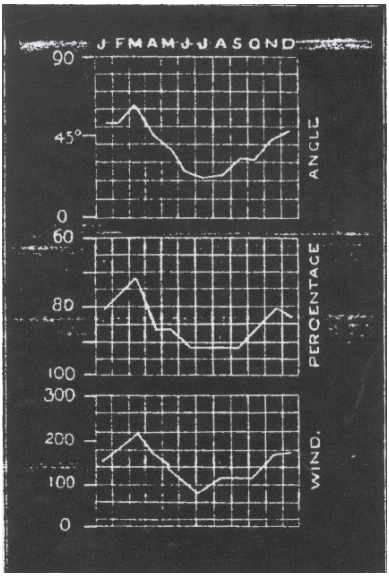


Fig. 3 Observed correlation among: angle of incidence of falling rain, upper gauge catch as a percentage of lower gauge catch, and wind speed in miles per day (after Symons 1870, 22)

faithful supporters, no doubt buoyed by the Rotherham results. Indeed, Symons himself apparently favored this hypothesis in early 1871 (see Symons 1871). Other contributors failed to see how anyone could take the inclination hypothesis seriously. As one put it: " … I could as soon believe that 2 and 2 could make 5, as believe that 'when rain is deflected by wind it is spread over a somewhat larger area than when it falls vertically'…" (Du Port 1871, 182). Despite all of the attention, after more than a year of discussion, it did not appear that much progress had been made in selecting between the two hypotheses. Some contributors began to feel that argument alone would not settle the matter: "Such a war of mere hypotheses may last for ever; had we not better wait for some more definite result from the experiments?" (Stow 1871, 183).

And the experiments and observational studies did continue. Perhaps the most decisive were conducted during 1876 and 1877 by George Dines, who considered the unsolved height-catch problem to be "a standing disgrace to meteorologists" (Dines 1878, 17). Although Dines apparently was unaware of Bache's experiments from the 1830s, his own studies with multiple gauges on a high tower were remarkably similar to those conducted by Bache. Dines determined that the amount of rain caught by a gauge was significantly impacted by "the position it occupied on the tower with reference to the direction of the wind" (ibid. 20). Like Bache, Dines had found that when gauges were placed at different locations on the top of a tower, the windward gauges consistently caught less rain that the leeward gauges. Furthermore, such differences were detectable even when gauges differed only very slightly in their positions relative to the wind: when four small-mouthed gauges were placed within a single large-mouthed gauge, "the [small] one on that side farthest removed from the wind collected the greatest quantity of rain" (ibid.).

Dines' results would make little sense if the inclination hypothesis were correct. After all, if the inclination of falling rain is what determines gauge catch, and inclination is a function of height, then why would gauges located *adjacent* to one another at the *same* height above the ground give different readings? Dines himself chose to avoid discussion of the inclination hypothesis in his published report, but his remarks indicate that he interpreted his results to support the wind hypothesis: "… can there be any reasonable doubt remaining that the rainfall, at 50 feet. above the ground, is *nearly* the same in amount as that which reaches the ground, but owing to the influence of the wind, we are unable to collect, and so measure it correctly?" (ibid. 23, italics in original).

The wind hypothesis soon did become the favored explanation of the height-catch difference. In 1878, Symons reviewed the vast experimental work that had been conducted and described the matter as "nearly settled," attributing "the greater part" of the height-catch difference to "the eddies produced by the rain gauge funnels, and by the buildings on which they are placed" (Symons 1879, 30). By 1881, Symons would appear to be completely convinced of the wind hypothesis. In a paper read before the British Association at York, he praised Jevons for providing "the most important theoretical contribution" to the investigation of the height-catch problem and highlighted Dines' results as a key piece of experimental

evidence (Symons 1882, 44–45). In the published version of Symons' paper, he included at the end a paragraph describing the reaction of Sir William Thomson (later Lord Kelvin), then president of the British Association, upon hearing the paper read. Reportedly, Thomson found it "exceedingly satisfactory" that the height-catch problem was finally settled, the conclusion being that the deficiency of rain in an elevated rain gauge is only an apparent one, resulting from "imperfect collection by the gauge" (ibid. 45).

While confusion between the wind and inclination hypotheses would occasionally resurface to the end of the nineteenth century, for practical purposes the height-catch problem seems to have been resolved by the early 1880s. It was clear that, other things equal, a gauge placed near the ground would give more accurate readings than one placed on the edge of a rooftop, and Symons was advocating the placement of all gauges so that their mouths stood just 1 foot above the ground. However, as Jevons had noted, some wind loss could be expected to occur even when gauges were located close to the ground, since the gauge itself presents an obstacle to the wind. By 1885, Symons had heard of at least two methods for controlling the wind in the vicinity of a near-ground gauge. F.E. Nipher had invented a funnel-like shield that fit near the top of a rain gauge and was designed to prevent

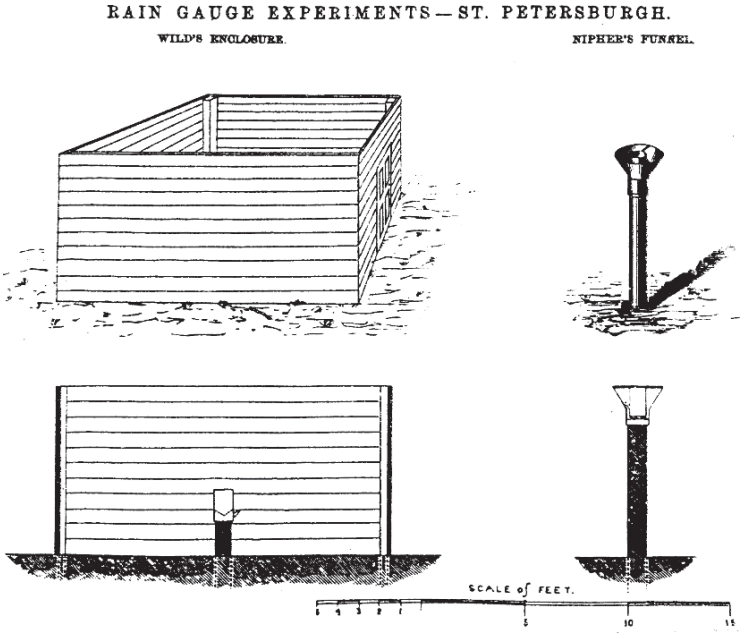


Fig. 4 Two attempts to control wind in the vicinity of the gauge: Wild's Enclosure and Nipher's Shield (after Symons 1886, 23)

upward deflection of air currents encountering the gauge, and the German meteorologist and physicist Heinrich Wild had proposed the use of a fence-like enclosure to reduce wind speeds in the vicinity of the gauge (see Fig. 4). Effort had shifted from defending the wind hypothesis to determining how to minimize the effects of wind on gauge catch.

Concluding Remarks

The height-catch story is, on one level, about the discovery of an instrumental artifact. As the preceding discussion shows, however, the story is about much more than the functioning of an instrument. In their quest to explain the height-catch difference, scientists were forced to address fundamental questions about the physics of rainfall and to reevaluate their assumptions about where and how rainfall could be accurately measured.

For much of the nineteenth century, scientists understood the height-catch difference to be a real difference, accurately reflecting the fact that raindrops grow continually by condensation during their descent. To accept this explanation of the height-catch difference was to accept, as Jevons so nicely put it, that "the larger part of the rain which falls upon the surface of earth does not proceed from the clouds, as we should naturally suppose, but is derived from the lower strata of the atmosphere, within 200 or 300 feet of the surface" (1861, 421–422). Working under this assumption, an important research goal was to discover the natural laws that governed the rapid growth of falling raindrops, so that rainfall measurements made at any height above the surface could be converted to ground level.

A very different picture emerged with the acceptance of the wind hypothesis. Perhaps most significantly, rain was once again understood to come almost entirely from the clouds. The height-catch difference itself was believed to reflect not a real difference but rather the failure of elevated gauges to accurately measure rainfall when exposed to the wind. And the relatively strong yet highly-variable winds present at elevated levels left little hope that elevated gauge readings could be routinely converted to ground-level readings. Instead, to accurately determine the amount of ground-level rainfall, rain gauges should be located near the ground and should be shielded from the wind.

Recognizing the effects of wind on rain gauge catch thus had significant practical and theoretical implications: it changed not only what counted as an accurate rainfall measurement but also how scientists understood the origin and development of rain itself.

Acknowledgements Sincere thanks to James Fleming, Francis Longworth, John Norton, and the editors and anonymous reviewers of this volume for helpful suggestions on earlier drafts of this paper. I would also like to acknowledge the National Meteorological Archive, Met Office, Exeter, UK, which holds the works from which Figs. 2, 3 and 4 are taken.

Notes

1. The problem of explaining the observed deficiency of rain in elevated gauges will be referred to in this paper as the *height-catch problem*. The decrease in gauge catch with height will be referred to as the *height-catch difference*. For ease of discussion, proposed explanations of the height-catch difference have also been given names in this paper.
2. A century later, George Dines would lament the continued use of electricity as an explanatory catch-all in meteorology: " … electricity, so ready to the hand of the meteorologist, who thus endeavors to explain the obscure by that which is still more obscure" (1878, 24).
3. These growth processes had been proposed prior to the 18th eighteenth century. Middleton (1965, 97–99) suggests that they were treated as competitors, rather than as processes that work together in increasing the size of raindrops.
4. Franklin knew Heberden already in the 1750s, if not before, and may have been invited to dinner by Heberden twice in early 1767 (see Heberden 1767/1970, 141).
5. Thanks to John Norton and Uljana Feest for their help with translation from the German. Luke Howard is famous in the meteorological world for his cloud classification system, which is still used today.
6. Lacking any direct measurements of humidity, Phillips decided to use diurnal temperature range (i.e. the difference between the daily maximum and minimum temperatures) as a proxy for humidity—the smaller the temperature range, the greater the humidity. Since the diurnal temperature range was smallest in the winter months, these months were considered the most humid.
7. In a brief discussion of the height-catch problem, Middleton (1965, 169) remarks that Bache obviously disagreed with the conclusions reached by Phillips. Certainly Bache disagreed with Phillips' conclusion that the effects of the wind were unimportant, but it does not appear that he rejected the condensation hypothesis.
8. According to a recent history of wind tunnels (Lee 1998), Frank H. Wenham is generally thought to have built and tested the first wind tunnel in 1871. It is unclear what place, if any, Jevons' work should be given in this history.
9. One wonders whether Jevons' experiments with the smoking paper were inspired by the comments made by Boase (1822), quoted earlier, about the visible eddying of smoke when it is windy.

References

A.D. Bache, 'Note on the Effect of Deflected Currents of Air on the Quantity of Rain Collected by a Rain Gauge', *British Association for the Advancement of Science, Transaction of Section* 8 (1838), 25–27.

D. Barrington, 'Letter of December 24, 1770, to William Heberden,' *Philosophical Transactions of the Royal Society* 61 (1771), 294–297.

H. Boase, 'On the Differences in the Annual Statements of the Quantity of Rain Falling in Adjacent Places', *Annals of Philosophy* (*London*) 4 (1822), 18–21.

S. Cannon, *Science in Culture: The Early Victorian Period* (New York: Science History, 1978).

L. Daston, 'The Moral Economy of Science', *Osiris* 10 (1995), 3–24.

G. Dines, 'Difference of Rainfall with Elevation', *British Rainfall 1877* (1878), 15–25.

J.M. Du Port, 'To the Editor of the Meteorological Magazine', *Symons's Monthly Meteorological Magazine* (1871), 181–183.

M. Flaugergues, 'Memoir on the Quantity of Rain and the Number of Days of Rain, Snow and Drizzle at Viviers, During 40 Years', *Annals of Philosophy* 14 (1819), 108–116.

B. Franklin, 'Letter to Dr. Thomas Percival', in W. Willcox (Ed.) *The Papers of Benjamin Franklin*, vol. 18 (New Haven: Yale University Press, 1771/1974), 154–157.
W. Heberden, 'To Dr. Franklin', in L. Labaree (Ed.) *The Papers of Benjamin Franklin*, vol. 14 (New Haven: Yale University Press, 1767/1970), 141.
W. Heberden, 'Of the Different Quantities of Rain, Which Appear to Fall, at Different Heights, Over the Same Spot of Ground', *Philosophical Transactions of the Royal Society* 59 (1769), 359–362.
W. Heberden, 'Note', *Philosophical Transactions of the Royal Society* 61 (1771), 297.
J. Herschel, *Meteorology* (Edinburgh: Adam and Charles Black, 1861).
L. Howard, 'Beobachtungen über den Regen und über die Regenmesser', *Annalen der Physik* 1:41 (1812), 417–425.
H. Ingram, 'To the Editor of the Meteorological Magazine', *Symons's Monthly Meteorological Magazine* (1871), 209–210.
W.S. Jevons, 'On the Deficiency of Rain in an Elevated Rain-Gauge, as Caused by Wind', *London, Edinburgh and Dublin Philosophical Magazine* 4:22 (1861), 421–433.
J.L. Lee, 'The Origin of the Wind Tunnel in Europe, 1871–1900', *Air Power History* 45:2 (1998), 4–15.
W.E.K. Middleton, *A History of Theories of Rain and Other Forms of Precipitation* (New York: Franklin Watts, 1965).
H.R. Mill, 'The Development of Rainfall Measurement in the Last Forty Years', *British Rainfall 1900* (1901), 23–45.
D.P. Miller, 'The Revival of the Physical Sciences in Britain, 1815–1840', *Osiris* 2 (1986), 107–134.
D.E. Pedgley, 'A Short History of the British Rainfall Organization', Occasional Papers on Meteorological History No. 5 (2002).
P.P. Pennant, 'To the Editor of the Meteorological Magazine', *Symons's Monthly Meteorological Magazine* (1871), 72–73.
J. Phillips, 'Report of Experiments on the Quantities of Rain falling at Different Elevations above the Surface of the Ground at York', *British Association for the Advancement of Science, Transactions of Section* 3 (1833), 401–412.
J. Phillips, 'Third Report of Experiments on the Quantities of Rain Falling at Different Elevations Above the Surface of the Ground at York', *British Association for the Advancement of Science, Transactions of Section* 5 (1835), 171–179.
F.W. Stow, 'To the Editor of the Meteorological Magazine', *Symons's Monthly Meteorological Magazine* (1871), 183.
G.J. Symons, 'On the Rainfall of the British Isles', *British Association for the Advancement of Science, Report, London* (1865), 192–242.
G.J. Symons, 'Notes on some Results of the Various Sets of Experimental Gauges', *British Rainfall 1866* (1867), 21–35.
G.J. Symons, 'Interpretation of the Rotherham Experiments', *British Rainfall 1869* (1870), 17–24.
G.J. Symons, 'Editor's Comment', Symons's Monthly Meteorological Magazine (1871), 65.
G.J. Symons, 'Difference of Rainfall with Elevation', *British Rainfall 1878* (1879), 21–30.
G.J. Symons, 'On the Rainfall Observations made upon Yorkminster by Professor John Phillips F. R. S.', *British Rainfall 1881* (1882), 41–45.
G.J. Symons, 'Dr. Wild on the Influence of Dimensions and Position upon the Indications of Rain Gauges,' *British Rainfall 1885* (1886), 14–23.

Going Right and Making It Wrong: The Reception of Fizeau's Ether-Drift Experiment of 1859

Jan Frercks

Introduction

For Hippolyte Fizeau, everything in his ether-drift experiment of 1859 went right.[1] He had expected a change in an optical effect caused by the motion of the Earth, and, indeed, he measured this change. The data corresponded neatly with his theory-based prediction. This was the first demonstration of the motion of the Earth with respect to the luminiferous ether. In fact, it remained the only experiment which proved such an effect.

According to present-day knowledge, however, the effect deduced and measured by Fizeau is known not to exist, signaling that something must have gone fatally wrong in Fizeau's experiment. The problem is not that the experiment is based on the assumption of an ether, which is today is regarded as non-existent. The problem rather is that the data themselves must be judged as "wrong" in the sense of referring not to nature, but to some (unknown) property of the apparatus or the measuring procedure.

This makes the experiment interesting for both the history of science as well as the philosophy of science. How can we deal adequately with a situation in which theory and experimental data agree perfectly, but nevertheless turn out both to be wrong?

After presenting Fizeau's experiment in some detail, I will discuss this methodological question head-on. I suggest that by focusing on the change of how Fizeau's experiment was perceived over time, we can employ a method of analysis that integrates both Fizeau's "right" as well as our "wrong". This is then applied to Fizeau's ether-drift experiment. It will be shown that reactions to Fizeau's experiment are remarkably general, even if firm judgements are made. This will provide some insight as to how to scrutinize others' scientific contributions.

J. Frercks (✉)
Leuphana College, University of Lüneburg, Scharnhorststr. 1, 21335 Lüneburg, Germany
e-mail:jan.frercks@t-online.de

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_11,

Fizeau's Ether-Drift Experiment

Since percipients could in general only rely on Fizeau's published accounts, I use only these here to render the experiment. Fizeau published his experiment in an abridged version in the *Comptes rendus* of the *Académie des Sciences* and in full length in the *Annales de chimie et de physique*.[2] A German translation of the latter was printed in the *Annalen der Physik und Chemie*.[3]

The Aim of the Experiment

In the middle of the nineteenth century, the wave theory of light was firmly established. According to this theory, light consists of transverse waves in an all-pervasive, subtle medium, the ether, assumed to be a lattice of ether atoms held together by attractive and repulsive forces (similar to our present picture of crystals). When disturbed, the ether particles oscillate around their rest positions, propagating the disturbance as waves.

One of the salient problems of this theory was the relation between the ether and ordinary matter, in particular matter in motion. In 1851, by means of a sophisticated interference apparatus, Fizeau showed that while the ether is barely dragged in moving air, it is partially dragged in moving water, which was in agreement with an equation given long before by Fresnel.[4] If one can generalize from the motion of the water in the laboratory and apply this principle to the orbital motion of the Earth (of about 30 km/s), an ether wind should be produced. This passing along of the ether at the surface of the Earth is called ether drift.[5] Fizeau was sure that he had found a workable method to prove this.

The Brewster-Fizeau Effect

Fizeau's experiment is based on an effect which I call the Brewster effect, having been first investigated empirically by Brewster. Fizeau presents the final equation, but neither derives nor explains it, as it was obviously widely known.[6] The best contemporary account of its theoretical explanation was given by Jamin.[7]

Imagine a beam of plane-polarized light travelling in the direction of the x-axis (Fig. 1, on the left) and landing obliquely on a glass plate, containing y (not shown). The plane

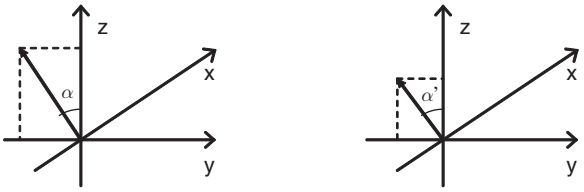


Fig. 1 Orientation of a beam of plane-polarized light before (left) and after (right) the oblique passage through a glass plate

of incidence is defined as the common plane of the incident, reflected and refracted rays (the x-z-plane in Fig. 1). The arrow indicates the orientation of the plane of polarization and the amplitude of the wave. In general, the plane of polarization subtends an angle α with the plane of incidence, but the wave can be regarded as a superposition of a wave *in* the plane of incidence and another one *perpendicular* to it, as indicated by the dotted lines. When hitting the interface, the light is partially reflected and refracted, but, according to Fresnel's equations, the proportion of the intensities of these two resulting partial beams is different for the parallel and for the perpendicular component. Accordingly, the plane of polarization of the leaving beam subtends a different angle α' with the plane of incidence. The rotation is given by

$$\tan \alpha' = \frac{\tan \alpha}{\cos^2(i - r)'}$$

where i is the angle of incidence, and r the angle of refraction.

How does the motion of the Earth change this effect according to Fizeau? I quote it in full length:

> The angle of incidence and the azimuth remaining unchanged, the larger the refractive index of the material [la matière] from which the plate is built, the larger the rotation; since the refractive index of a body is inversely proportional to the speed of light in this material [ce milieu], it follows that the magnitude of rotation depends on the speed with which the light propagates in this substance, the lower the speed, the larger the rotation.
>
> If now the speed of light inside the substance happens to change for any reason [une cause quelconque], one can foresee that the rotation will undergo a corresponding change; the study of the speed of light can therefore be traced back to an easily detectable phenomenon like the rotation of the plane of polarization.[8]

Thus, because the glass plate moves, the relative speed of light changes, which in turn changes the angle of refraction r and thereby alters the Brewster effect. I call this change the Brewster-Fizeau effect.

Since the quantitative calculation of the Brewster-Fizeau effect would have been arduous, in particular for more than one glass plate, Fizeau empirically determined the expected effect by comparing the Brewster effect for two media with different, but well-known refractive indices, such as crown glass and flint glass. From this he deduced the Brewster-Fizeau effect to be only 1/1500 of the rotation at rest.

To repeat, the Brewster effect is the rotation of the plane of polarization when a polarized light beam passes obliquely through a glass plate. The Brewster-Fizeau effect is the *variation* of the Brewster effect due to the motion of the glass plate with respect to the ether.

Implementation of the Apparatus

Building the apparatus was no doubt difficult, and Fizeau did not conceal a number of problems. However, although the process was not straightforward – Fizeau mentioned "long tinkering" no less than four times – he found workable solutions for most problems in the end.

Fizeau used packages of four conical glass plates to avoid multiple reflections in and between the glass plates. The different rotation for different wavelengths was corrected by placing a quartz or a vessel filled with appropriately diluted essence of lemon or turpentine behind the piles of glass plates. Elliptical instead of linear polarization, most probably caused by imperfect homogeneity of the glass, was eliminated by the use of a sequence of glass plates with different angles of incidence. The effect was increased to a measurable size by the use of further piles, adjusted to a particular angle.

The final apparatus consisted of a chain of optical pieces which were fixed into short segments cut from telescope tube, see Fig. 2. These were placed in a V-shaped channel which was especially suitable for rotating many pieces around the optical axis. The polarizing prisms at both ends were equipped with a pointer and a scale, and probably a mechanical gadget to turn them carefully. The whole optical bench could be rotated horizontally.

Nevertheless, even during the measurements, Fizeau made some improvements to the apparatus, which again indicate the difficulties associated with running the apparatus, see Fig. 3. On June 20th, a telescope was added for adjusting the apparatus, on June 24th, two glass tubes were included to make the apparatus stiffer, and as of July 1st a plumb line assured the correct direction of the vertical axis.

Measurements

For measurements, sunlight coming from a heliostat was introduced from either side by means of two fixed mirrors, located left and right of the apparatus (not shown in Fig. 2). An angle of the polarizing prism was adjusted at the polarizing prism on one side, and the observer looked through the polarizing prism on the other

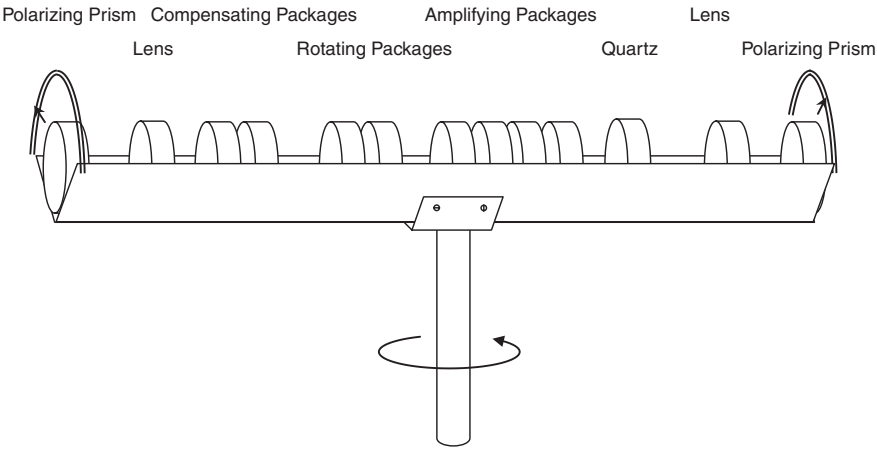


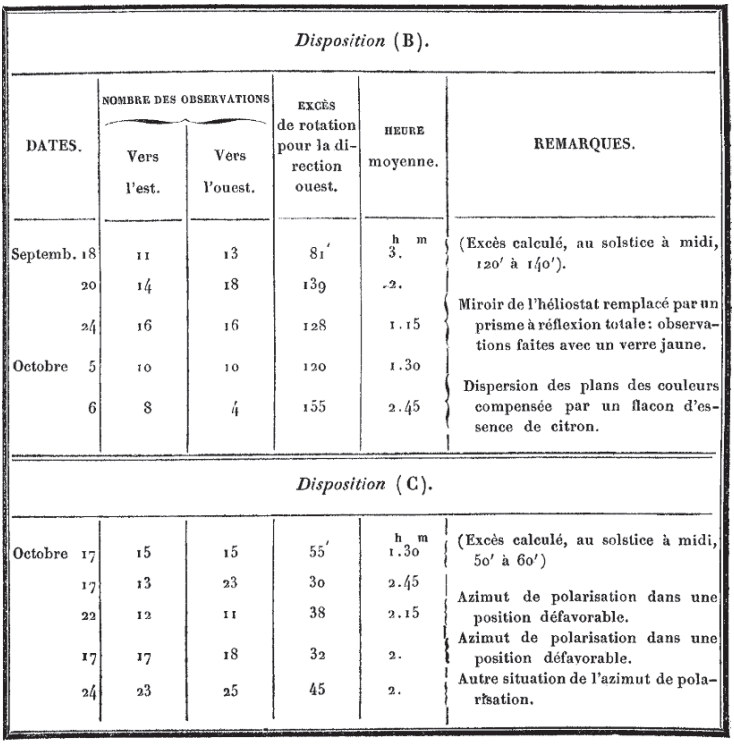
Fig. 2 Sketch of Fizeau's apparatus according to the description in Fizeau (1860a)

(159)

Disposition (A).

DATES.	NOMBRE des observat. Vers l'est.	Vers l'ouest.	EXCÈS de rotation pour la direction ouest.	HEURE MOYENNE.	RÉMARQUES.
			° ′ ″	h m	
Juin 2	11	18	33	4	(Excès calculé, au solstice à midi, 45′ à 65′).
3	34	32	45	2.30	
4	54	57	60	12	
5	46	55	66	12	
6	15	15	90	11.30	
	15	15	20	1 45	
	20	20	25	4	
7	15	15	53	11.30	
8	25	25	38	2.30	
9	30	27	25	3.30	
13	30	31	54	12	
15	17	19	73	1	
	20	22	8	4	
16	12	13	1.29	11.45	Dans ces trois séries on a introduit à dessein une erreur constante dans l'appareil en inclinant l'axe de rotation, afin d'observer l'influence de l'heure dans des conditions différentes des précédentes.
	12	15	1.15	2.15	
	21	18	1. 1	4	
20	17	21	42	3	
21	27	29	57	12.15	A partir de cette série, on a ajouté à l'appareil une lunette accessoire destinée à assurer l'identité de direction du rayon dans les deux situations de l'appareil.
	21	15	31	4	
24	40	41	46	12.15	
	20	22	— 7	4	Excès inverse, c'est-à-dire pour la direction est.
27	10	10	53.30	1.30	A partir de cette série l'appareil est consolidé avec deux longs tubes de verre mastiqués pour éviter les flexions.
	10	10	37	3	
	10	10	23.30	4	
28	11	12	60	12	
30	20	20	32	2.30	
Juillet 1	26	23	53.30	12.45	
2	24	20	49	11.30	Un fil à plomb est ajouté à l'appareil pour maintenir l'axe vertical et éviter les flexions.
	15	15	23.30	4	
3	25	15	39	11.15	
	15	15	19	4	
4	10	10	39	1	
	16	16	9.30	4	Un des miroirs (celui de l'est) ayant paru défectueux, l'autre est divisé en deux parties, la première pour l'est, la seconde pour l'ouest.
5	10	20	56.30	1	
	10	10	26	3	
6	20	20	55.30	12.15	
	10	10	25	2.30	
	10	10	23.30	3.45	
7	10	15	47	2.30	
	10	14	30	4	
8	10	20	62	11.15	Amélioration des images par un petit changement de direction du rayon et par l'addition d'un écran.
	10	20	50	12.45	
	11	12	43	2.45	
	10	10	19	4	
9	8	8	55.30	10.45	Observations alternées de deux en deux pour diminuer l'influence des changements de température.
	10	10	59	12.30	
	10	10	43	2.45	
	10	10	26	4	
11	10	10	44	10.30	Série de 4 h. faite avec des précautions particulières.
	10	10	59	12.30	
	14	14	28	4	
12	10	10	59	1	
	10	10	27	4	
13	16	16	50	12.30	
	14	14	31	4	
14	10	10	43	1	Le 14 on a interverti les positions des miroirs ; une pile est devenue oscillante sur son support par l'effet de la chaleur sur les liéges.
	10	10	42	2	
	10	10	3	3.45	
15	10	10	59	12.15	

Fig. 3 Fizeau's data, reproduced from Fizeau (1860a, 159 and 161)



Disposition (B).

DATES.	NOMBRE DES OBSERVATIONS Vers l'est.	Vers l'ouest.	EXCÈS de rotation pour la direction ouest.	HEURE moyenne.	REMARQUES.
Septemb. 18	11	13	81′	h m 3.	(Excès calculé, au solstice à midi, 120′ à 140′).
20	14	18	139	2.	
24	16	16	128	1.15	Miroir de l'héliostat remplacé par un prisme à réflexion totale: observations faites avec un verre jaune.
Octobre 5	10	10	120	1.30	
6	8	4	155	2.45	Dispersion des plans des couleurs compensée par un flacon d'essence de citron.

Disposition (C).

Octobre 17	15	15	55′	h m 1.30	(Excès calculé, au solstice à midi, 50′ à 60′)
17	13	23	30	2.45	
22	12	11	38	2.15	Azimut de polarisation dans une position défavorable.
17	17	18	32	2.	Azimut de polarisation dans une position défavorable.
24	23	25	45	2.	Autre situation de l'azimut de polarisation.

Ann. de Chim. et de Phys., 3e série, t. LVIII. (Février 1860.) 11

Fig. 3 (continued)

side and turned this until the light vanished, and the same was done in the opposite direction. This double procedure was repeated with the apparatus rotated by 180°, and the measured values were compared.

Fizeau published those data that he considered reliable in two tables, reproduced in Fig. 3. Every number in the table represents the mean of a series of measurements consisting of 12 to 111 single measurements, and totaling over 2000 measurements.

Fizeau concludes that

> … one is therefore led by reasoning and by experiment to admit as very probable that the azimuth of polarization of the refracted ray is really influenced by the motion of the refracting medium [milieu], and that the motion that drags the Earth along through space exerts an influence of this kind on the rotations produced in polarized light by piles of inclined glass plates.[9]

Data: Right or Wrong?

In order to demonstrate how convincingly the data indicate the existence of the sought-after effect, I have plotted the excess rotation of the plane of polarization (in minutes of arc) against time.

Figure 4 shows that there are many more points after 12:00 than before. The easiest explanation for this is that Fizeau was a late riser, although it is possible that the experiment required sufficient preparation time every day. Anyhow, since Fizeau strove for the effect as such, not for its change with the time of the day, there was obviously no need to start earlier.

More important, however, is the top-bottom asymmetry. All but one of the points are above zero. Even more impressive is that the measured values decrease on either side of 12:00. With the apparatus fixed in east-west direction, the rotation of the Earth would reduce the effect to zero at 6:00 and 18:00. Indeed the measured values reveal such a trend, at least for the afternoon.

Fizeau had even made a quantitative prediction, saying that the effect would be between 45 and 65 min of arc for 12:00 (Fig. 3, "Juin 2"). I have added the respective decrease curves based on the fact that at the solstices the shape of the decrease curve is sinusoidal (for other days of the year it is nearly so). Given that the range of 20 min of arc neither encompasses the maximum margin of error, nor has a particular statistical meaning (at a time when calculating the standard deviation was not yet common), the fact that 65% of the points lie between the curves adds to the quality of the measurement.

The problem with these data is that present-day knowledge disproves their validity, because relativity theory does not allow for such an effect. The data should be scattered symmetrically above and below zero, independent of time.

Can we save Fizeau's data *and* relativity by questioning the derivation of the null effect from relativity? We can not, because the non-detectibility of a uniform motion inside an inertial system (which the Earth almost is) is a *postulate* of relativity rather then a *consequence*. So if we accept relativity, we must bluntly say that something in Fizeau's experiment went severely wrong.

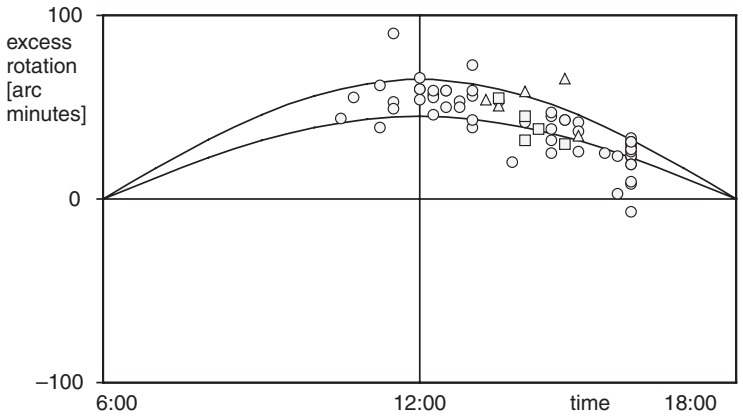


Fig. 4 Excess rotation of the plane of polarization, data from Fig. 3. Circles indicate readings from "disposition A", triangles from "B", and squares from "C", the data of "B" and "C" have been recalculated for the expected values in "A"

We can even tell a little more precisely *what* went wrong. For this, it is important to note that Fizeau's experiment does not hinge on the theory of the luminiferous ether, that is of an entity which according to present knowledge does not exist. We do not have to enter the discussion about whether a statement about a non-existing thing is necessarily wrong, since the effect to be proven or disproved is independent from ether theory.

There is a positive example of the separation between an experiment and its theoretical interpretation. Fizeau's previously mentioned ether-drag experiment of 1851 was based on the ether theory of light and it had proven the partial ether drag in moving water. Yet today, physicists do not view this experiment to be wrong. Instead they keep the data, strip off the ether background and take the experiment as a proof of the relativistic formula for the addition of velocities. This experiment receives extensive coverage in modern physics textbooks and in history of science surveys, not the least encouraged by the ample references made to it as well as to trains by Einstein.

I would not generally argue for a division of empirical data from theoretical interpretation. There is no non-interpretative production of data. In contrast, the very neutrality of data perceived by practitioners is an indication of a common ground of understanding about the conditions under which the data are produced. The existence of such a common ground is not self-evident, but it can be established. Only then can data be used for extending or changing knowledge.

In our case, there *is* such a common ground. There is no reason to assume that our understanding of plane of polarization, glass plates, east and west, morning and afternoon differs fundamentally from Fizeau's. The experimental procedure and its results are independent of ether theory.[10]

This allows us to specify that something went amiss in the complex of apparatus, measuring procedure, actual measurements, and choice of valid readings. After all, the published data are not merely ambiguous, they are wrong in a remarkably stable way and in total agreement with Fizeau's theoretical prediction. For the moment, however, we cannot locate the mistake inside this complex. At least, it cannot be explained by a general lack of skill on the part of the experimenter. Fizeau had built and run several difficult experiments which were primarily admired (both in his time and today) for their technical performance.[11]

Furthermore, there must be a mistake in the theoretical derivation of the effect. But here again, we cannot tell whether the basic assumptions of the ether theory or Fizeau's derivation (or both) are wrong.

Present-day physicists have made their decision. They accept relativity as true, and they have forgotten Fizeau's experiment.[12] If they remember the experiment at all, it is "only of historical" interest. What, then, do we as historians or philosophers of science do with this experiment? How can we deal with this "only"?

This requires some methodological considerations. Although using the case at hand for illustration, they are intended to be of more general relevance. The conclusions to be drawn will then be applied to the case at hand.

Methodological Considerations

The following two ways to deal with this situation seem to be irreconcilable.[13] Position A adopts the actors' perspective and tries to avoid any influence of current knowledge on the reconstruction of scientific endeavors in the past science. In contrast, position B uses current knowledge to judge past science.

Applied to the case at hand, these positions lead to opposite conclusions. According to A, Fizeau's experiment was right because Fizeau took it to be right. According to B, it was wrong, because we know that it contradicts relativity.

The main criticism that can be leveled at A is that, since the historian does not dare to state that the experiment was *really* right, s/he hides the judgement behind an agnosticism which is said to be methodologically justified. The avoidance of talking about right and wrong altogether thus circumvents what the whole enterprise of science is basically about.

The main criticism that can be leveled at B is that using later established knowledge contributes nothing to understanding past scientists who could not have had this knowledge. Furthermore, since every science is historically situated, it precludes the understanding of scientific practice in general, including the generation of the very knowledge that is used in judging the past.[14] In not explicitly stating whose knowledge is the "current knowledge", it runs the risk of suggesting that the knowledge used is *necessarily* the best available knowledge or even a-historical knowledge.

Thus A belongs to historiography of science whereas B belongs to philosophy of science. A is descriptive, whereas B is judgmental. Presented in this – admittedly schematical – way, the positions seem irreconcilable.

I claim that they are not. The impression of irreconcilability results from choosing a too restricted time frame in each of the positions, in A around the past event and in B around the present. The solution is to take the time between Fizeau and us into account. Position C strives to fill in the gap between 1859 and 2008 (although in this paper, for practical reasons, I stop at around 1910). Applying position C means retracing reactions of any kind to an event of past science, including judgements about right or wrong. The strong dichotomy of Fizeau's clear-cut result on the one hand and its total rejection nowadays on the other is not watered down, but bridged by examining intermediate persons and papers. Both 1859 and 2008 are not outside history. In this respect, there is no principle difference between then and now.

There must have been at least one change in opinion regarding Fizeau's experiment between 1859 and 2008, but this change was not necessarily abrupt. There may have been doubts and uncertainties. The assessment of Fizeau's experiment may have been stable for long periods, the experiment may have been temporarily forgotten, the discussion may have faded away, but this is all part of the (contingent) story, not of the methodological position. Rather than trying to find out myself what really went wrong in Fizeau's laboratory (and at his desk when calculating the effect to be expected), I let contemporaries and followers do this work. Special emphasis will be put on which part of Fizeau's experiment (main hypothesis, derivation of the effect, apparatus, experimental procedure, evaluation of the

data) is challenged and on which notions the authors use in order to label what actually "went amiss".

Position C embraces both A and B. These are extreme ends of a continuum, but nevertheless they belong to one and the same continuum.

Position C is compatible with A, because it adopts the actors' perspectives. But the actors are not merely Fizeau and his contemporaries, but also all later persons in relation to the experiment, present-day scientists included. As historians of science, we do not have to say that Fizeau was right, because that would dismiss the opinion of other actors. It is even justified to say that Fizeau was wrong, because this is what *has become* the dominant opinion *among* the actors.

Likewise, C is compatible with B, if proponents of B admit that the "current knowledge" is in fact the knowledge of today's physicists. What the historian or philosopher of science is doing in B is simply adopting the opinion of his or her physicist colleagues.[15] Using hindsight is nothing more than frankly siding with present-day scientists. This shows that today's knowledge is not a-historical, but simply the (preliminary) end of a chain of judgements.[16] It is a recently taken snapshot of the discussion. There is nothing special about the present.

As extreme ends of C, A and B remain viable positions for particular cases. For example, one may reconstruct Fizeau's experiment from Fizeau's perspective. This is especially interesting for questions about the inner structure of a research program and about publication strategies.[17] It necessitates, however, the awareness that the absence of criticism is not methodologically justified agnosticism, but simply a consequence of excluding later critics.

If on the other hand, no present-day physicist discusses Fizeau's experiment, one may legitimately fill in the gap and judge Fizeau's experiment based on today's knowledge. Remaining laboratory notes in the possession of the *Archives de l'Académie des Sciences* would make this an interesting project. This, however, would be practicing science itself rather than the history or the philosophy of science.

Thus the choice of position is basically a choice of the period of time from which criteria for evaluation are taken. This choice has nothing to do with the problem of Whig history. Whether or not a case study is whiggish depends on the philosophy of history underlying the narrative. There can be whiggish and non-whiggish case studies for each position – A, B, and C.

I will now adopt position C and try to retrace the reception and assessment of Fizeau's experiment.

Reception and Assessment of Fizeau's Experiment

The three journals in which Fizeau's experiment was published were the most prestigious and the most widely read physics journals in the second half of the nineteenth century in France and Germany, so any disregard of Fizeau's experiment cannot merely be attributed to it having been published in a remote journal.

Even the French journals were easily available in Germany. In these journals, I have searched through all papers belonging to those four research fields in which I expected to find Fizeau's experiment mentioned. Likewise, I have searched through most French and German textbooks, both on physics in general and optics in particular, that were published or re-edited between 1859 and 1910. Further texts have been included if they have been referred to.

Far-Reaching Neglect

All in all, I have found references to Fizeau's experiment in only thirteen of these texts. These texts will be dealt with in turn below. But first the far-reaching neglect has to be discussed. Among the textbooks, we find Fizeau's experiment mentioned only in multi-volume textbooks on optics, written by leading theoreticians. Among the other textbooks, according to scope and envisaged audience, some explain the Brewster effect, some only polarization in general, and some do not even mention polarization at all.

In the journals, there are a significant number of papers related to the topics of the rotation of the plane of polarization, of the theory of light waves in moving media, and of the motion of the solar system. Obviously, all of these topics are close to Fizeau's experiment, but only Faye mentions it. This can be explained by a kind of patchwork epistemology. Scientists may have been aware of Fizeau's experiment, but purposely chosen different topics. Researchers attached their work to others, but tried to avoid overlap. One had little to gain in repeating or commenting someone else's work after all. There is one field, however, for which Fizeau's experiment is doubtlessly relevant.

Interdisciplinary: The Theory of Stellar Aberration

Stellar aberration is the long-standing, empirically supported effect of an apparent shift in the position of celestial bodies due to the motion of the Earth-bound observer.[18] When light is considered to be made up of particles, this effect can be explained easily, but according to the wave theory of light it is much more complicated, and intimately bound to the problem of ether drag. If ether drag exists in air and thus in the terrestrial atmosphere, stellar aberration must occur at the entrance of the atmosphere during the light's passage through it. If there is no ether drag in the atmosphere, stellar aberration must occur in the telescope. This is precisely the question Fizeau's experiment examines. Therefore, it is the broad neglect among the vast majority of papers which requires explanation rather than the few cases in which Fizeau's experiment is discussed or at least mentioned.

My explanation is that the problem was too complex for Fizeau's contemporaries to comment on his experiment. At the time, the most prominent experiment discussed was still Arago's much simpler prism experiment of 1810.[19] Arago

had detected that the angle of refraction does not depend on whether the prism approaches or recedes from the light ray, a result that prompted Fresnel to propose his theory of partial ether drag.

Problems dealt with in the discussion of stellar aberration include: What happens if a diffraction grating is used instead of a prism in experiments like Arago's? Is precision spectroscopy affected by the motion of the Earth? Is stellar aberration the same in a telescope filled with an optically dense medium like water or turpentine? A parallel shift of a plane of ether molecules is still the same plane, so how can ether drag in this plane have an effect? On the assumption that the ether is dragged in the atmosphere, is the light ray still perpendicular to the wave surface? Is Arago's experiment useless, because Arago used an achromatic prism which compensates all changes in wavelengths?

The following questions are directly related to Fizeau's experiment. Is it the wavelength, the frequency or the velocity of light which determines the angle of refraction? What happens with light (in particular its direction and its wavelength) if it is reflected off a moving mirror? According to Fresnel's theory, is the ether itself dragged by optically dense matter or only the ether waves? Is it appropriate to generalize from a laboratory scale air-stream to the motion of the terrestrial atmosphere?

I have mentioned these points in some detail because they demonstrate the complexity of this apparently confined problem. Not all of these problems, to be sure, have been solved in a satisfactory manner. This explains sufficiently, I suggest, the reason why most texts neglect to mention Fizeau's experiment. It would be expecting too much to review the most current state of the discussion, as well as to apply it to another experiment.

However, there is another possible interpretation for instances of neglect of Fizeau's experiment. The fact that four papers touch directly on the content of Fizeau's experiment without actually mentioning it, I interpret as tacit criticism. Criticism, if tacit, is obviously a kind of assessment, therefore I have included these into the sequence of instances of reception.

In the remaining part of this section, I will present – in chronological order of publication – the sequence of instances of reception of Fizeau's experiment. The limited number of these instances allows doing this completely. This provides the material for discussion of the particular historical case and of the method in the final section.

Faye (1859)

The first comment on Fizeau's experiment was published only one week after Fizeau's paper by Faye.[20] Faye relates the experiment to the question of the motion of the solar system. The direction of this motion toward the constellation of Hercules was well established, but its speed – Faye uses Struve's and Peter's value of 7894 m/s – was very uncertain (giving all digits did not mean an accuracy of 1 m/s at that time).

As an astronomer, Faye considers himself incapable of scrutinizing the apparatus and the method, and he recommends the Academy's *Section de Physique* to do so. In spite of – or because of – this, Faye takes the data as given. These data, however, applied to the new evidential context of the motion of the sun rather than of the ether drag in the terrestrial atmosphere do not remain equal.[21] Faye assigns them several different functions and degrees of reliability.

Faye distinguishes data produced at noon, when the effect due to the sun's motion is negligible, from those produced at 16:00. The first set serves as a reference to fix

Date.		Vitesse.	Déviation calculée.	Déviation observée.	Différence.
	h m	m			
4 juin	0.00	30200	65′	60′	+ 5′
11 juillet	0.30	25500	55	59	— 4
12 »	1.00	23900	51	59	— 8
13 »	0.30	25300	54	50	+ 4
23 octobre	2.00	22600	48	45	+ 3

Fig. 5 Recalculation of the expected values, including the motion of the solar system, Faye 1859, 872

the proportionality constant between speed and rotation of the plane of polarization. In order to reduce arithmetical effort, Faye only uses five values. Two are taken from the first and the last measurements, because he wants to make sure that the apparatus did not change over time. Three values are taken from those mea-

Date.		Vitesse.	Déviation calculée.	Déviation observée.	Différence.
	h	m			
11 juillet	4	7600	16′	28′	— 12′
12 »	4	7500	16	27	— 11
13 »	4	7500	16	31	— 15

Fig. 6 Recalculation of the expected values, including the motion of the solar system, Faye 1859, 872

surements which, according to Fizeau, have been taken with "particular care"[22]. Figure 5 lists the resulting "calculated deviation". Consequently, the differences are zero in the mean. Note that "difference" is defined as the deviation of the calculated values from the observed values, not vice versa.

The next calculation has been done including the motion of the solar system, however, the calculated values for 16:00, when the position of the apparatus is such

		Vitesse.	Calcul.	Observation.	Différence.
Vers midi	4 juin	29400m	57′	60′	— 3′
	11 juillet	29400	57	59	— 2
	12 »	28500	55	59	— 4
	13 »	29400	57	50	+ 7
	24 octobre	24500	48	45	+ 3
à 4 heures	11 juillet	14100	27	28	— 1
	12 »	14100	27	27	0
	13 »	14000	27	31	— 4

Fig. 7 Recalculation of the expected values, without the motion of the solar system, Faye 1859, 872

that an effect due to the motion of the solar system is to be expected, are much too small (Fig. 6).

Faye therefore repeats the procedure, but without the motion of the solar system. Now "agreement [accord] is complete" (Fig. 7).

Does this prove that the solar system does not move? Faye's statements are ambiguous:

> If the experiments of M. Fizeau really have the exactitude that they appear to have, the proper motion towards the constellation of Hercules which astronomers attribute to the solar system does not exist. If, in contrast, the astronomical determinations of this motion are sound, one has to admit that the experiments of the learned physicist are affected by a systematic error or that his theory contains an important gap.[23]

This either-or is not resolved in the end. Faye does not jump to the conclusion that the solar system does not move, but

> despite these motives, by no means do I want to pronounce myself against Fizeau's experiments. … I restrict myself to saying that the contradiction I just mentioned seems to require that this eminent physicist subject his theory and his apparatus to a special examination.[24]

It is important to note that Faye's discussion concerns the aptitude of Fizeau's experiment in its present state to detect a speed of only 8 km/s. The discrepancy between data and astronomical knowledge only arises from granting the data produced with a speed of 30 km/s a great deal of accuracy. Thus, a "systematic error" can only refer to the degree of accuracy, not to the experiment as such.

Maxwell (1868)

Maxwell mentions Fizeau's experiment in a letter to Huggins in the course of a discussion about moving spectroscopes.[25] According to Maxwell, if Fresnel's theory is correct, no effect is to be expected in the spectroscope, but

> on the other hand, M. Fizeau has observed a difference in the rotation of the plane of polarization according as the ray travels in the direction of the Earth's motion or in the contrary direction, and M. Ångström has observed a similar difference in phenomena of diffraction. I am not aware that either of these very difficult observations has been confirmed by repetition.[26]

Does this suggest agreement or rejection? The crucial word seems to be "observed". On the one hand, this contrasts with the expression "I have found that … ",[27] which Maxwell uses for his own result. "Finding" presupposes something having been there all along, while "observing" may describe subjective, even deceptive perception. On the other hand, in the expression "these very difficult observations", "observation" is used for the whole experiment rather than for the act of watching. Therefore "observed" can be read as neutral – and neutrality is positive in science.

Maxwell's reference to Fizeau's ether-drag experiment starts:

> In another experiment of M. Fizeau, which seems entitled to greater confidence, he has observed that …[28]

This however, does not further elucidate the case. While on the one hand, "greater confidence" emphasizes the contrast between both experiments (although watered down by "seems entitled"), on the other hand, "observed" is used for something trustworthy here.

Instead of trying to decipher Maxwell's opinion, I view this ambiguity as a result. There is a continuum of judgement between right and wrong, and by carefully chosen words, Maxwell managed to maintain ambiguity in a situation when

> the whole question of the state of luminiferous medium near the Earth, and of its connection with gross matter, is very far from being settled by experiment.[29]

Verdet (1870)

The predominant explanation for the null result in refraction experiments like Arago's is that there is an effect due to the motion of the prism which is cancelled out by aberration so that there is no effect from the perspective of the observer in motion. This makes Fizeau's experiment plausible for Verdet, since

> there is no similar compensation for the deviation of the plane of polarization, so that, according to the direction in which the ray travels parallel to the motion of the Earth, there must be a difference between the deviations of the plane of polarization. […] M. Fizeau has verified this conclusion from theory by experiment.[30]

Verdet is the most decided proponent of Fizeau's experiment, although he mentions a number of technical difficulties. How is this possible? Verdet refuses to discuss the quantitative accuracy of Fizeau's data, because the motion of the solar system is said to be unknown. Therefore, one cannot compare Fizeau's data with any reliable quantitative prediction. Instead, as a qualitative experiment, it is beyond doubt:

> M. Fizeau has given a new physical proof [preuve physique] of the translational motion of the Earth […].[31]

Boussinesq (1873)

Boussinesq's paper was reported to the *Académie des Sciences* by Fizeau.[32] Boussinesq applies his sophisticated theory of the interplay between ether and matter to the motion of the Earth. He concludes that no effect should be expected if the apparatus as

a whole (including the observer) moves with respect to the ether. He mentions Fizeau's experiment on ether drag in moving air and water, but he interprets this as a partial drag of the ether *waves* rather than of the ether itself. He clearly states that the generalization from a laboratory scale air-stream to the motion of the terrestrial atmosphere is by no means self-evident, so an experiment which deals with this problem head-on is badly needed. If Boussinesq does not mention Fizeau's polarization experiment in this respect, this must be interpreted as a belief that its conclusions are wrong.

Ketteler (1873)

Ketteler has written a series of papers on the theory of stellar aberration, which were put together in revised form in a book.[33] Regarding optical experiments to prove the motion of the Earth, Ketteler states:

> The first step in this direction, as completed by Fizeau, has rightly [mit Recht] been received with lively and universal interest.[34]

This work "has rightly become famous".[35] Considering the rotation of the plane of polarization, Ketteler states,

> … the applicability of this method seems to be already practically proven by Fizeau.[36]

Ketteler's own calculations reveal no doubt about the theory of the Brewster-Fizeau effect, but Fizeau's equation relating rotation and refractive index is described as a "rough empirical equation",[37] which explains the difference between data and theoretical prediction. The experiment itself is reported, but not discussed or commented upon.

Calling Fizeau's experiment "broken off",[38] together with the emphasis on the immature theory leaves Fizeau's experiment as nothing more than the first step. Like all others before, Ketteler is skeptical with regard to the quantitative results, but unlike the others, he is vague even about the qualitative result. At least he judges the method as positive.

Fizeau (1874)

In 1874, Fizeau himself enters – or better: avoids – the discussion.[39] He reports a commission's decision to give Mascart the *Grand Prix des Sciences Mathématiques* for 1872 for a work on the influence of the motion of the light source and the observer. Fizeau mentions a number of attempts to detect the motion of the Earth by strictly optical means, based on different effects such as "phenomena of interference, of gratings, of aberration, of refraction, and of polarization".[40] For those familiar with the topic, it is easy to trace these back to the experiments of Babinet, Ångström, Airy, Arago, and Fizeau. Fizeau concludes:

> And finally (except for one or two exceptions, the results of which remained doubtful [douteux]), all experiments of this kind have led to completely negative results, as if a general law of nature was always opposed to their success.[41]

Since Fizeau's experiment on the Brewster-Fizeau effect *did* have a positive result, it must be among the "exceptions" and is thus classified by Fizeau himself as "doubtful".[42] It is not clear when and why Fizeau changed his mind. Lab-notes in the *Archives de l'Académie des Sciences* in Paris reveal that Fizeau was still experimenting on an ether-drift experiment based on the measurement of radiant heat. Most probably, even Fizeau himself was not able to detect the mistake. In addition he might have discussed the topic with other French physicists, although this cannot be traced through published papers.

Lorentz (1887–1904)

Lorentz wants experiments rather than the "degree of probability or simplicity"[43] to decide whether or not the ether wind exists:

> I know of only two experimental inquiries that are relevant for this question.[44]

One is Michelson's ether-drift experiment of 1881, of which Lorentz remains skeptical, not least because of his own 10-page calculation that demonstrates that the effect due to the ether wind would only be half of what Michelson had expected. The other is Fizeau's experiment, in which Fizeau "has found"[45] [a trouvé] an effect.

> No objection could be made [saurait être faite], it seems to me, to the conclusion of this scientist, namely that close to the surface of the Earth the ether is not at rest with respect to it, but, in my opinion, it has not been demonstrated by these experiments that the relative speed is exactly equal to the speed of the Earth.[46]

This demonstrates – as was the case with Verdet – belief in the qualitative effect, but not in the quantitative data. Lorentz continues:

> I will not enter the discussion about these observations here, because this has to be based on a study of the modifications which, due to the motion of ponderable matter, undergo the conditions that determine reflection and refraction at the surface of bodies.[47]

In 1895, Lorentz accepts the results of Michelson and Morley's further experiments,[48] and he saves his theory by the introduction of the "auxiliary hypothesis"[49] [Hülfshypothese] of the contraction of moving bodies. In turn, this makes Fizeau's experiment problematic, and Lorentz discusses it at length.[50] He demonstrates that his theory precludes the Brewster-Fizeau effect. Since he did not succeed in "developing"[51] [entwickeln] his theory in accordance with Fizeau's result without losing explanatory power for the other experiments, he concludes that Fizeau's result stems from an "observational error".[52] Although he suspects that the mirrors were at fault, a

rough calculation shows that the error would be much smaller than Fizeau's measuring values. Hence the whole experiment provides "a contradiction which I am not able to resolve".[53]

In 1905, then, Fizeau's "completely inexplicable"[54] [völlig unerklärliches] result has gained the status of the only exception. Although Lorentz is still not sure whether to trust it or not, the experiment no longer threatens the whole enterprise. It has – literally – become a footnote to the increasingly homogeneous complex of theory and experimental results.

Des Coudres (1889)

Des Coudres describes an experiment which he characterizes as "the translation of an experiment, which Fizeau had proposed in 1852 for radiant heat, into the field of electrodynamics".[55] This refers to an experiment, which had not produced a result because of severe technical problems.[56] Although Des Coudres's experiment is much more similar to this experiment than to the one based on the Brewster-Fizeau effect, it is nevertheless striking that the one that *had* produced a result is not mentioned.

Mascart (1893)

Mascart acknowledges that Fizeau's data – although scattered – clearly show an effect of the expected order of magnitude, but he is skeptical of Fizeau's conclusion.[57] Three times he points to some weak feature of the experiment and uses judgmental expressions:

> An experiment by M. Fizeau, the explanation of which appears to be unsatisfactory and which would seem to indicate [semblerait indiquer] that …[58]

> Despite certain causes of error [causes d'erreur] which he was not able to be specify nor to eliminate completely, M. Fizeau believes [croit] that …[59]

> This experiment is the only one up to now that would allow thinking that …[60]

To be inexplicable, to be proven faulty, and to be the only experiment gives reason to doubt that the experiment proved the existence of the ether wind. It only "seems to be", it only "would allow thinking of", and Fizeau only "believes".

The technical aspects receive praise, but even this can be read critically:

> Apart from the problem to be solved, this experiment is particularly remarkable for the practical solution of technical problems, which it presented.[61]

Suggesting a new evidential context (polarization measuring techniques) can be seen as another way of saying that the apparatus is unsuitable for the purpose for which it was built.

The most concrete criticism concerns the theoretical aspects. The crucial point for Mascart is Fizeau's assumption "that rotation is governed [réglée] by the absolute

speed of light in the moving medium [milieu mobile], not by the apparent refractive index".[62] Like Lorentz, however, Mascart is not able to give a quick answer and resumes "it would be necessary to see if the theory of reflection underpins [comporte] such an interpretation".[63]

Zehnder (1895)

Zehnder describes a near repetition of Fizeau's ether-drag experiment, although the ether itself is accelerated rather than air or water. The null result fits well to Fresnel's equation and to Fizeau's ether-drag experiment, which is dealt with at length. Likewise, Zehnder has done an ether-drift experiment with mercury employed to slow down the ether stream in a glass tube. He admits that its null result is inconclusive. Strikingly, Zehnder does not mention Fizeau's ether-drift experiment. Moreover, he speaks of

> the many experiments which have been performed on this and which all have lead to a negative result,[64]

as if Fizeau's did not exist. This is the easiest and the most blatant way of saying that Fizeau's experiment and its result are irrelevant.

Wien (1898)

Wien summarizes for the *Gesellschaft Deutscher Naturforscher und Ärzte* all experimental attempts to detect the motion of an apparatus with respect to the ether, and compared them to Lorentz's theory of an immobile ether. Wien presents the experiments subdivided into those with positive and those with negative results. Fizeau's experiment is presented as follows:

> Fizeau's experiment about the influence of the motion of the Earth on the rotation of the plane of polarization by piles of glass plates. The positive result of this experiment has been doubted recently. It would not be consistent with the assumption of an immobile ether according to the investigations of H. A. Lorentz.[65]

Strikingly, Wien ranges Fizeau's experiment under "experiments with negative result",[66] i.e. according to what *should* have occurred. Or did he use "negative" for both null results and poor results?

Drude (1900)

Drude's presentation of Fizeau's experiment makes it clear that it is theory that sets the frame of judgment:

> While according to the developed theory, apart from aberration and the change of the period of oscillation according to Doppler's principle, no influence of the motion of the Earth on

> optical phenomena observed on Earth is to be expected, and in fact is generally not observed, Fizeau believed that he verified the effect of the motion of the Earth in one case.[67]

The positive results of stellar aberration and of the Doppler effect have been integrated into theory, but Fizeau's positive result has not. Drude explains why Fizeau's experiment must be wrong:

> According to the theory given here, such a difference cannot exist. If in any position of the apparatus the analyzer is adjusted to darkness, this means that the motion of the light is limited to a space that does not extend over the analyzer. As we have discussed above, …this space does not change its boundaries due to the motion of the Earth if the rays are kept unchanged relative to the apparatus, even if crystalline media are used to produce the boundary surface S of the light space. Therefore the dark position of the analyzer should be independent of the orientation of the apparatus with respect to the motion of the Earth.
>
> Anyway, it is desirable that Fizeau's experiment is repeated once more; for the time being we can consider it doubtful, whether there is really a contradiction on this point between the theory given here and experience.[68]

This means the relation between Fizeau's result and Drude's theory is not even a contradiction. For a "real" contradiction, the experiment must be repeated. This is a remarkable relation between theory and experiment. It is theory which defines that the data count for nothing, but Drude allows for having a contradiction between a better experiment and theory in principle. How, then, do we know that another experiment is better? And what would Drude do with a sound experiment in contradiction with his theory?

Poincaré (1901)

For Poincaré, the expectation of a null result in experiments like Fizeau's reaches the status of a theorem:

> Theorem. The motion of the Earth does not influence optical phenomena if the squares of ξ, η, ζ are neglected.[69]

It follows a mathematical proof based on Lorentz's theory. Meanwhile, the interesting question is whether a second order effect is to be expected.

Lorentz's theory is firmly founded in many experiments, leaving Fizeau's experiment as an exception:

> All these experiments have given negative results. There is, however, an exception: M. Fizeau believes that he observed [a cru observer] an influence of the motion of the Earth on the rotation of the plane of polarization in vitreous reflection [*sic*] of polarized light. But these experiments are extremely delicate, and M. Fizeau has told me himself the doubts which he maintained with regard to the afore-mentioned result.[70]

Since theory forbids a result like Fizeau's, it is no longer necessary to find out in detail what went wrong in Fizeau's theoretical derivation and in the experimental procedure.

Brace (1905)

It took about 50 years before two experimental attempts to scrutinize Fizeau's results experimentally had been made. Brace repeats the general view that Fizeau's experiment is the only ether-drift experiment with a positive result and that everyone – including Fizeau – doubted its result. Two reasons are invoked to legitimize his own repetition of the experiment:

> Notwithstanding these facts, reference is still made to the positive results of this experiment.[71]

The few references which are, in fact, made to Fizeau's experiment are still too many for Brace. Obviously, physicists preferred to be allowed to say "*all* experiments have null results".

The second reason concerns the experimental procedure:

> The undoubted care and skill devoted to this experiment has left an impression that, perhaps, after all the test involved surface conditions which did not enter in the other experiments and which might give the positive results obtained.[72]

If the "surface conditions" are crucial, Fizeau's experiment, at least its technical part, might have been right! In saying that this part is "undoubted", the experiment would be saved, if these surface conditions could be specified.

I will not go into the details of Brace's implementation of his experiment, which eventually gave a clear-cut null result. His conclusion concerning Fizeau's experiment is:

> The results of my observations, made under similar conditions, show that the effects which Fizeau obtained must have been due to other causes than those whose effects he supposed he was examining.

Brace's own result shows that Fizeau's result is not caused by the Brewster effect. Nevertheless, Brace does not call the technical part of Fizeau's experiment "wrong". He had initially tried to repeat Fizeau's experiment by similar means, but soon made significant changes in the apparatus. This makes it impossible to specify what went wrong in Fizeau's experiment. All that can be said is that there are "other causes", concealed in the replaced parts of the apparatus.

Strasser (1907)

The second experiment, initiated independently from Brace, was published somewhat later by Strasser. He judges Brace's work as not yet conclusive (obviously to legitimize his own attempt), but – like Brace – bluntly rejects Fizeau's results. He expresses that "only Fizeau claimed [gab an] to have observed such an influence in one case".[73] The main source of skepticism is its contradiction with the "theory of the electrodynamics of moving bodies",[74] but he ventures a possible technical explanation for Fizeau's results as well:

> It is possible after all that the differences in rotation which Fizeau got for the east and west direction are caused by the different positions of the mirrors with respect to his apparatus.[75]

His conclusion could not be more self-assured:

> From the available observations it can be concluded with certainty that the effect observed by Fizeau does not exist [nicht besteht], and hence that so far no experiment has proven an influence of the Earth's motion on any optical phenomenon on Earth.[76]

Note that Strasser does not mention Einstein's seminal paper about relativity, which was published in the same journal two years before. Relativity is still an *explandum* rather than an *explanans* at that time. The rejection of Fizeau's experiment and its results has been achieved without it.

These were the instances of reception of Fizeau's experiment up to 1910. This material will now be discussed. First, I will gather the results from the case study, both for the history and the philosophy of science. Then I will discuss the method employed in the case study in its relation to the history of science and to the philosophy of science.

Discussion

Growing Doubts Without Reasons

At first glance, the result of the inquiry is disappointing. Most writers, not only those of textbooks, but also researchers on topics close to Fizeau's experiment like polarization, ether theory, or the motion of celestial bodies, do not say one word on it. The few who do are either brief or they report rather than discuss the experiment.

At least it was possible to retrace the broad line of the changing assessment of the experiment. Faye (1859) accepted Fizeau's result, Maxwell (1868) was undecided, Verdet (1870) agreed with the qualitative result, Ketteler (1873) and Lorentz (1887) had doubts, but still agreed rather than disagreed. The neglect of Fizeau's experiment by Boussinesq (1873), Fizeau (1875), Des Coudres (1889), and Zehnder (1895) indicates an at least skeptical position. With Mascart (1893) and Lorentz (1895), doubts prevailed, and Wien (1898) bluntly twisted the result into what should have come about. Drude (1900) disagreed on account of theoretical reasons, and finally Poincaré (1901), Brace (1905), and Strasser (1907) were sure that the experiment had gone amiss and the results were wrong.

All above mentioned parts of the experiment (main hypothesis, derivation of the effect, apparatus, experimental procedure, evaluation of the data) have been challenged. The judgements, however, were mostly uttered without specific arguments (except for Faye's criticism of Fizeau's evaluation of the data, which, although governed by his particular interest, is comprehensible in detail). The critics refused to do what I wanted them to do, namely to scrutinize Fizeau's experiment. Faye's "affected

by a systematic error" for the apparatus and "contains an important gap" for theory is the most concrete diagnosis. Except for Lorentz' and Strasser's hinting at the mirrors, there is no suggestion, even by the staunch critics, of what had gone wrong in Fizeau's experiment. Nor was there a discussion of Fizeau's derivation of the Brewster-Fizeau effect and its magnitude. Even Lorentz, who argued from his theory that no effect is to be expected was not able to reveal the weak point in Fizeau's derivation. Judgements were vague, and comments were remarkably general.

Therefore, while we know *that* something in the experimental procedure went wrong, we still do not know *what*. Likewise, even today it is still unknown whether Fizeau's derivation of the effect – given the general framework of the ether hypothesis – is right or wrong. Relativity has *cut* the Gordian knot of the theoretical part of Fizeau's experiment, no one has yet *solved* it. One may be tempted to try to do this. However, when reading through the discussion on the theory of stellar aberration and making some calculations myself, I was too often surprised by problems and difficulties which I had not foreseen to be confident about not having overlooked something important.[77] This is compounded by the fact that, today, there are no interlocutors and no new contributions.

Classification of Assessments

Upon closer inspection, however, the sequence of references tells us a lot about the pursuit of science. First, if my interpretation is correct, there are at least four different reasons for not mentioning some piece of scientific work. For textbooks, it may be too specific and too complicated. For experimental science, it may be seen as an already solved problem which should be avoided if one wants to accomplish something new. For the theoretical debate, it may just be too difficult for an easy assessment. And finally, in an appropriate context, neglecting to mention someone's work can be perceived as blatant criticism.

Second, it is remarkable that the appraisal of Fizeau's experiment changes at all given that there is no detailed discussion, either of the apparatus or of the theoretical derivation. To say that assessments are vague is itself too vague. It is worth taking a closer look at the way general terms are used without having something substantial to say. Although transitions are fluent, six different ways of expressing opinions in general, non-argumentative ways can be identified. A common feature is that the critics do not use strong words like "wrong" or "failure". Mascart speaks of "error", but shifts the accusation from the person to the apparatus.

(1) It is possible merely to repeat what others have said. This can be positive (Ketteler: "rightly been received with lively and universal interest"; "has rightly become famous") or negative (Wien: "has been doubted recently"). Stating that the experiment has been doubted is itself an act of doubting. This is purely performative, without residual argumentation.
(2) The intervention of Fizeau as a person can be emphasized, implicitly suggesting deception. I have discussed the ambiguity of Maxwell's "observed".

More explicit skepticism is expressed in saying that Fizeau "believed" (Mascart, Drude, and Poincaré) or "claimed" (Strasser) to have observed an effect. Positive attributes to Fizeau (Faye: "learned physicist"; "eminent physicist") merely serve to mitigate the tone of the criticism.

(3) It can be pointed out that the experiment is particularly difficult (Maxwell: "very difficult observations"; Poincaré: "extremely delicate"). This at once questions the experiment and excuses its failure. If the result is accepted, the stated difficulty of the experiment adds to the fame of the experimenter, but if the status is still vague, it is an open flank for an attack. Here again, praise (Brace: "undoubted care and skill devoted to this experiment") serves as an excuse for the attack.

(4) A kind of criticism appearing in the guise of well-meaning advice is to propose – explicitly or implicitly – to find a new evidential context to save at least some part of an otherwise failed experiment. We may hear this in Mascart's "particularly remarkable for the practical solution of technical problems", and indeed Mascart mentions Fizeau's solution for the isolation of the main ray by prismatic glass plates in a different context.[78] Even Brace's cool statement that Fizeau's effect "must have been due to other causes" implicitly proposes investigating *these* causes rather than saying that the experiment is irredeemably and mystically flawed.

(5) It is possible to emphasize that Fizeau's experiment is the only one with this result (Ketteler: "first step", "broken off"; Lorentz: "only two experimental inquiries"; Mascart: "the only one up to now"; Poincaré: "exception"; Strasser: "only Fizeau").

(6) Finally, a way of judging an experiment is to invoke its relationship to theory. Lorentz and Mascart call for an examination of whether Fizeau's derivation is right according to present theory. Later-on, it will suffice for Wien, Drude, Lorentz, Poincaré, and Strasser to point out that Fizeau's effect cannot exist because it contradicts the most recent theories. A detailed scrutiny of Fizeau's derivation of the Brewster-Fizeau effect is no longer necessary then. Fizeau's experiment is past saving because theories have been designed in ways that preclude positive results like Fizeau's right from the start.

Explanation of the Change of Assessment

The empirical inquiry has revealed a number of significant features of the changing assessment of Fizeau's experiment.

(1) Immediately after the event there was total agreement, while at the end of the time frame of my inquiry, there was total rejection.

(2) The transition has been a gradual and almost monotone increase in skepticism.

(3) Apart from Maxwell and Verdet, there has been no opposite assessment at the same time. This is all the more significant, since all percipients (except Strasser citing Wien and Drude) refer to Fizeau's papers directly, not to preceding percipients.

(4) The state of irreversible rejection had been reached before the experiment had been replicated by Brace and Strasser. Are we then justified to say that they were as theoretically biased as Fizeau?
(5) The state of irreversible rejection had been reached before relativity theory.
(6) There was no significant difference in reception in France and Germany.

These results call for an explanation. How can assessment change smoothly, but profoundly with neither a repetition nor a detailed scrutiny of the experiment?

My suggestion is that the changing role of theories in physics during the second half of the nineteenth century is the most important factor.[79] The early ether theory was confined to reproducing known effects without contradiction. Later, ether theory did not only merge optics with electrodynamics, it changed its role. This included prediction, and – perhaps more importantly – securing theory against single empirical results. It is no coincidence that during this time Pierre Duhem, himself theoretical physicist, stated that scientific theories are irrefutable by single experimental findings. Huge amounts of work have been invested in the construction and elaboration of theories, too much to be threatened by single experiments. It would have been possible to integrate the recalcitrant result into theory by some auxiliary hypothesis. Our case, however, shows that in practice, it might work much easier. Although a convincing diagnosis was never made, the experiment was simply assumed to be wrong.

The Theory-Experiment Relation

The case study provides interesting material for the philosophy of science as well, in particular for the problem of the relation between theory and experiment. Fizeau's experiment is not at all typical for experimental science, rather it is the rare, but almost ideal case of an experimental test of theory. An observable effect is mathematically derived from a fundamental theory, an apparatus capable of detecting this effect is conceived and built, measurements are taken, the data happen do be in full correspondence with the expectation, which corroborates the underlying theory.

For Fizeau's experiment alone, there would have been little more to say. Taking later assessment into consideration, however, enlarges significantly the spectrum of relations between theory and experiment. Including the present-day assessment of Fizeau's experiment as wrong makes it an intriguing instance of the malleability of experimental procedures in terms of stabilizing particular results. Including the intermediate assessments, however, adds much more. Although one might expect correspondence between theory and experimental results as the goal of experimental science, this state, after having been achieved, was nevertheless broken open again. Theories have been built in contradiction to the results of Fizeau's experiment. The requirement that a new theory should explain all known facts was bluntly ignored in this case. As a consequence, the experiment was rejected, even though no one had any idea of what had gone wrong in it.[80] Neither was rejection based on replication.[81] If according to theory there should have been a null result in the experiment, it could

have been elucidative to scrutinize Fizeau's derivation in order to understand better the fundamental assumptions of ether theory (after all, the new theories *were* still ether theories), but this was not done either. Furthermore, it turns out that assessments do not have to be clear-cut. In contrast, there was a continuum of doubts, again without explicitly stating technical or theoretical reasons for the degree of rejection.

(1) The malleability, (2) breaking open a firm correspondence between theory and experimental results, (3) change of assessment without new evidence and without technical scrutiny, and (4) a continuum of doubts: At least these points have to be taken into account in any system – whether descriptive or normative – of the theory-experiment relation.

Recounting Reception and the History of Science

It is now widely acknowledged that the reconstruction of past events is an integral part of science itself. If accounts in published papers, in textbooks, or in conference speeches do not neatly correspond to the event, this is not bad historiography, but a part of science proper, and even an important part.[82]

Apart from Faye, who used Fizeau's experiment as a measurement of the motion of the sun, the meaning of the experiment remained stable across the changing assessment over the first 50 years. This is not necessarily the case. Events may change their meaning in later reconstruction, sometimes beyond recognition.[83] This makes studying the reception more than a mere supplement to studying the event itself. Of course, an event can still be studied in its own right but it precludes understanding its non-local significance, because this is only established in reception.

The focus on continuous reception is especially well-suited for the study of error because reception shares three essential features with error. First, error is cognitive. An apparatus, a procedure, a notion, even a calculation or a chain of reasoning cannot be wrong as such, it has to be judged as wrong with respect to some frame of meaning. Errors are errors only with respect to some particular interest.

Second, this necessitates transforming a situation into a text – if only a text spoken tacitly by the experimenter to himself or herself. Hence error is discursive. Remedying an error necessitates a textual or verbal account of the error.[84] Explicitly stating what went wrong defines what it is to go right. Granted that the category of "right" is as complex as the category of "wrong" (and not just the absence of the different kinds of wrongness), it is the spectrum of identified errors that reveals the frame of meaning.[85]

Third, recognition of error is always retrospective (thus re-cognition). After being detected, an error is no more an error but at most a problem to be solved.

Reception is itself cognitive, discursive, and retrospective. It deals with an account of an event (rather than with the event itself), it relies on having this account communicated by whatever means, and it comes after the event by definition.

The act of reception, however, is itself an event to be accounted for, to be communicated and to be set into a particular evidential context. (Faye and Verdet judged Fizeau's experiment differently, because of different esteem for astronomy; Mascart's judgement depended on whether it was related to ether theory or to

polarization physics.) Retrospective judgement is no idle activity, but intimately bound to the interests and frames of meaning of the percipients.[86] This allows for a variety of reactions to one and the same event and to a variety of definitions as to what it means to have gone right. Hence, the reactions perhaps tell more about the percipient than about the event in question.

This points to the main shortcoming of the method. Although one has to examine an enormous amount of text, the result is still a set of dispersed reactions. The problem is not that there might be few or no reactions, which would itself be a result. The problem is that they have been assembled by the single criterion of being related to a particular event in the past. Thus, it is hardly possible to reconstruct the context in which every single reaction occurred, let alone to identify causes for the judgements which are made in the reactions. All what can be done with reasonable effort is to identify the fields and the literary genre in which the past event is received and to classify the types of reactions. Explanation, however, will mostly be possible for the broad line of reception, not for every single case.

Recounting Reception and the Philosophy of Science

Rather than assessing past events based on present-day knowledge, I have proposed to retrace the chain of actual instances of assessment, including the present-day view as but the last of its instances. Where does this leave philosophy of science? Does it come down to substituting philosophy of science by history of science?

This depends on what is meant by philosophy of science. If the rational reconstruction of the present-day opinion of a past event is meant, then indeed I argue for substituting the philosophy of science by the history of science. Both endeavors, recounting the history of reception and rational reconstruction, have the same starting point and the same endpoint. They even share the aim of *explaining* this endpoint, although the categories of explanation may be fundamentally different ones. Therefore I do not see why for the rational reconstruction it should be of any virtue, *not* to look for earlier instances of assessment. Of course, it may be the case that, after having examined them, the philosopher of science is not content with any of them or that due to contingent reasons, there have been no assessments. In these cases, it is perfectly legitimate to venture one's own assessment. Philosophy of science, as Hans Radder has emphasized, is at once in and about the world.[87]

However, the question is to which world such an endeavor belongs. Certainly, it would not be history of science. Furthermore, at least if present-day scientists are no longer interested in this event, it will not count as science proper either. Is it philosophy of science? This, as always, may depend on how and by whom this assessment is itself perceived. Anyway, I do not see how such a new assessment adds *in principle* to the history of assessments, apart from being its $(n+1)$th instance. At least, it does not contribute to understanding scientific change.

Things are different for systematic questions, such as, for example, the question about the relation between theory and experiment. Here, the proposed method in no way undermines or substitutes basically philosophical questions. Broader con-

cepts for understanding science do not evolve from the empirical material after all. Extending the empirical material by including the history of reception does not change this (except if questions of temporal change and reception are at stake). Nevertheless, including the history of perception will do no harm to these questions either. In contrast, as the case study has shown, it may significantly enrich the questions to be asked by philosophy of science.

Conclusion

At the beginning, I asked how one, as historian or philosopher of science, can adequately deal with an experiment which for present-day scientists is of "only" historical interest – generally a polite expression for saying that it is no longer of interest at all. Indeed, if we focus on the experiment itself, there is nothing very interesting in it. If we focus on its present-day assessment, it likewise can be put aside easily. In contrast, the most interesting question is studying the very process that gradually made the experiment "of only historical interest" today. Both for the history of science and for the philosophy of science, Fizeau's "right" alone is as uninteresting as our "wrong". What is interesting is the gradual transformation of the former into the latter.

Acknowledgements For important discussions and valuable comments on earlier drafts of this paper I thank Christine Blondel, Olaf Breidbach, Giora Hon, Falk Rieß, Jutta Schickore, Friedrich Steinle, and Gerhard Wiesenfeldt. Furthermore, I thank Rachel Dinisman and David Petroff for correcting my English.

Notes

1. The expression "right" is to be understood here to be as neutral as possible. Elsewhere (Frercks 2005), I have shown that Fizeau regarded an experimental result as right if it was both expected and positive. Since he only published his findings in these cases, we may safely conclude that, for Fizeau, this experiment went right.
2. Fizeau 1859a and Fizeau 1860a, respectively.
3. Fizeau 1860b.
4. Fizeau 1851 and 1859b.
5. This inference is not straightforward, because it is not clear how the ether drag occurs. Fresnel saw the ether as dragged along by single molecules of matter. In contrast, Stokes conceived ether drag as a macroscopic phenomenon caused by the motion of the Earth as a whole. Thus, according to Stokes, there is no ether wind regardless of whether or not there is ether drag in moving air.
6. Fizeau 1860a, 132.
7. Jamin 1858–1866, III, 667–677.
8. Fizeau 1860a, 131.
9. Ibid, 163.
10. This is true except for a possible bias in the form of a particular theory-related result. Other statements cannot be separated from the underlying theory, which causes different historiographical problems. Jardine 1991, for example, discusses how to deal with pronouncements like Oken's "The nose is the thorax repeated in the head". He suggests trying to understand the past scientist's questions rather than the answers.

11. Among the most admired experiments are the measurement of the speed of light (Fizeau 1849, see Frercks 2000) and the previously mentioned experiment on ether drag in running water (Fizeau 1851).
12. And so have philosophers of science. In their comprehensive rational reconstruction of the theory of aberration, Janssen and Stachel 2004 do not even mention Fizeau's experiment.
13. There are of course further possibilities. One of them is not to accept the choice made by our contemporaries. We could proclaim that relativity is wrong and use Fizeau's experiment as a falsification. Another possibility is to suspect that nature has changed since Fizeau's experiment, which is not as ridiculous as it sounds. Goldstein et al. 1973 use Rømer's measurements of 1676 to find out whether the speed of light has changed significantly over the last three centuries.
14. Collins 1981 regards this as a vicious circle and therefore recommends restricting empirical work to present-day science.
15. Canguilhem 1979 develops a viable niche for B as part of philosophy of science, alongside practicing scientists.
16. Even among those who still question relativity, Fizeau's experiment seems to be forgotten, see for example http://members.lol.li/twostone/E/physics1.html.
17. In Frercks 2005, I attempt this for Fizeau's ether-related research between 1847 and 1852.
18. Acloque 1991 offers the most comprehensive account of this discussion.
19. Published as Arago 1853.
20. Faye 1859.
21. The notion of "evidential context" is from Pinch 1986.
22. Faye 1859, 871.
23. Ibid., 870–871.
24. Ibid., 875.
25. The letter has been published in Huggins 1868, 532–535.
26. Huggins 1868, 535.
27. Ibid.
28. Ibid.
29. Ibid.
30. Verdet 1870, 514.
31. Ibid., 515.
32. Boussinesq 1873.
33. Ketteler 1873.
34. Ibid., 406.
35. Ibid., 128.
36. Ibid., 125.
37. Ibid., 127.
38. Ibid., 134.
39. Fizeau 1874.
40. Ibid., 1533.
41. Ibid.
42. According to Lorentz 1895, 2, Fizeau has told van de Sande Bakhuijzen that he no longer considers his experiment to be decisive.
43. Lorentz 1887, 164.
44. Ibid.
45. Ibid.
46. Ibid.
47. Ibid.
48. Michelson and Morley 1887.
49. Lorentz 1895, 7.
50. Ibid., 125–138.
51. Ibid., 126.
52. Ibid., 7.
53. Ibid., 127.

54. Lorentz 1904, 267, note 95.
55. Des Coudres 1889, 78–79.
56. Fizeau 1852, translated as Fizeau 1854. On this experiment, see Frercks 2001.
57. Mascart 1889–1893, III, 121–126.
58. Ibid., 121.
59. Ibid., 124.
60. Ibid., 125–126.
61. Ibid., 121.
62. Ibid., 124.
63. Ibid.
64. Zehnder 1895, 79.
65. Wien 1898, 54.
66. Ibid., 53.
67. Drude 1900, 437.
68. Ibid., 437–438.
69. Poincaré 1901, 528.
70. Ibid., 517–518.
71. Brace 1905, 591.
72. Ibid., 591.
73. Strasser 1907, 137.
74. Ibid., referring to Wien and Drude, not to Einstein.
75. Ibid., 142.
76. Ibid., 144.
77. This was in the course of the replication of Fizeau's ether-drift experiment based on the measurement of radiant heat, see Frercks 2001.
78. Mascart 1889–1893, II, 413.
79. A good introduction to this topic is Buchwald 1988.
80. This is unlike the case of Miller's ether-drift experiment of the early twentieth century, which seemed to indicate a positive effect until a commission was set up which came to the conclusion that the data had been caused by temperature effects on the interferometer, see Swenson 1972.
81. Of course, there was increasing evidence from other experiments. Nevertheless it should be noted that Fizeau's experiment gave a positive result (rather than a null result) to first order (in terms of v/c), unlike for example Michelson and Morleys's experiments.
82. See, for example, Brannigan 1981, Nickles 1988, and Caneva 2001.
83. This has been further developed and examined in a case study in Frercks et al., submitted.
84. There are of course tacit, pre-cognitive ways of making things right, both in the handling of an apparatus and in reasoning, which have nothing in common with reception.
85. Hon 1998; Law and Mol 2001 have shown how one and the same case of going wrong was accounted for in very different ways, which reveal different opinions of what it means to go right.
86. On the different functions of references to past events in science, see Staley 1998 and Caneva 2001.
87. Radder 1996.

References

Acloque, P., 1991. "L'aberration stellaire: Un mirage qui a destitué l'éther," *Cahiers d'Histoire et de Philosophie des Sciences*, *Nouvelle série* 36: 1–258.

Arago, F., 1853. "Mémoire sur la vitesse de la lumière, lu à la première Classe de l'Institut, le 10 décembre 1810," *Comptes rendus hebdomadaires des séances de l'Académie des Sciences de l'Institut de France* 36: 38–49.

Boussinesq, J. V., 1873. "Sur le calcul des phénomènes lumineux produits à l'intérieur des milieux transparents animés d'une translation rapide, dans le cas où l'observateur participe lui-même à cette translation," *Comptes rendus hebdomadaires des séances de l'Académie des Sciences de l'Institut de France* 76: 1293–1296.

Brace, de W. B., 1905. "A Repetition of Fizeau's Experiment on the Change Produced by the Earth's Motion on the Rotation of a Refracted Ray," *The London, Edinburgh, and Dublin Philosophical Magazine* (6) 10: 591–599.

Brannigan, A., 1981. *The Social Basis of Scientific Discoveries*. Cambridge etc.: Cambridge University Press.

Buchwald, J. Z., 1988. "The Michelson Experiment in the Light of Electromagnetic Theory Before 1900." In: S. Goldberg and R. H. Stuewer, eds., *The Michelson Era in American Science, 1870–1930*. New York: American Institute of Physics, 55–70.

Caneva, K. L., 2001. *The Form and Function of Scientific Discoveries*. Washington D.C.: Smithsonian Institution Libraries. Also published electronically at www.sil.si.edu/silpublications/dibner–library–lectures/scientific–discoveries/.

Canguilhem, G., 1979. *Wissenschaftsgeschichte und Epistemologie*. Frankfurt/Main: Suhrkamp.

Collins, H. M., 1981. "Understanding Science," *Fundamenta Scientiae* 2: 367–380.

Des Coudres, T., 1889. "Ueber das Verhalten des Lichtäthers bei den Bewegungen der Erde," *Annalen der Physik und Chemie* (3) 38: 71–79.

Drude, P., 1900. *Lehrbuch der Optik*. Leipzig: Hirzel.

Faye, H. A., 1859. "Sur les expériences de M. Fizeau considérées au point de vue du mouvement de translation du système solaire," *Comptes rendus hebdomadaires des séances de l'Académie des Sciences de l'Institut de France* 49: 870–875.

Fizeau, H., 1849. "Sur une expérience relative à la vitesse de propagation de la lumière," *Comptes rendus hebdomadaires des séances de l'Académie des Sciences de l'Institut de France* 29: 90–92.

Fizeau, H., 1851. "Sur les hypothèses relatives à l'éther lumineux, et sur une expérience qui paraît démontrer que le mouvement des corps change la vitesse avec laquelle la lumière se propage dans leur intérieur," *Comptes rendus hebdomadaires des séances de l'Académie des Sciences de l'Institut de France* 33: 349–355.

Fizeau, H., 1854. "Idee zur Messung der Umlaufsbewegung der Erde," *Annalen der Physik und Chemie,* (2) 92: 652–655.

Fizeau, H., 1859a. "Sur une méthode propre à rechercher, si l'azimut de polarisation du rayon réfracté, est influencé par le mouvement du corps réfringent. Essai de cette méthode," *Comptes rendus hebdomadaires des séances de l'Académie des Sciences de l'Institut de France* 49: 717–723.

Fizeau, H., 1859b. "Sur les hypothèses relatives à l'éther lumineux. Et sur une expérience qui paraît démontrer que le mouvement des corps change la vitesse avec laquelle la lumière se propage dans leur intérieur," *Annales de chimie et de physique* (3) 57: 385–404.

Fizeau, H., 1860a. "Sur une méthode propre à rechercher si l'azimut de polarisation du rayon réfracté est influencé par le mouvement du corps réfringent. Essai de cette méthode," *Annales de chimie et de physique* (3) 58: 129–163.

Fizeau, H., 1860b. "Ueber eine Methode, zu untersuchen, ob die Bewegung des brechenden Körpers einen Einfluß habe auf das Polarisationsazimut des gebrochenen Strahls," *Annalen der Physik und Chemie* (2) 109: 160–169.

Fizeau, H., 1874. "Grand Prix des Sciences Mathématiques. Rapport lu et adopté dans la séance du 14 juillet 1873," *Comptes rendus hebdomadaires des séances de l'Académie des Sciences de l'Institut de France* 79: 1531–1534.

Frercks, J., 2000. "Creativity and Technology in Experimentation. Fizeau's Terrestrial Determination of the Speed of Light," *Centaurus* 42: 249–287.

Frercks, J., 2001. *Die Forschungspraxis Hippolyte Fizeaus. Eine Charakterisierung ausgehend von der Replikation seines Ätherwindexperiments von 1852*. Berlin: Wissenschaft und Technik Verlag.

Frercks, J., 2005. "Fizeau's Research Program on Ether Drag. A Long Quest for a Publishable Experiment," *Physics in Perspective* 7: 35–65.

Frercks, J., H. Weber, G. Wiesenfeldt. “Reception and Discovery. The Nature of Johann Wilhelm Ritter’s Invisible Rays,” *Studies in History and Philosophy of Science Part A,* in press.

Goldstein, S. J., J. D. Trasco, T. J. Ogburn. 1973. “On the Velocity of Light Three Centuries Ago,” *The Astronomical Journal* 78: 122–125.

Hon, G., 1998. “‘If This Be Error’. Probing Experiment with Error.” In: M. Heidelberger and F. Steinle, eds., *Experimental Essays: Versuche zum Experiment*. Baden-Baden: Nomos, 227–248.

Huggins, W., 1868. “Further Observations on the Spectra of some of the Stars and Nebulæ, with an Attempt to Determine therefrom whether these Bodies are Moving Towards or from the Earth, also Observations of the Spectra of the Sun and of Comet II,” *Philosophical Transactions of the Royal Society of London* 158: 529–564.

Jamin, J., 1858–1866. *Cours de physique de l’École polytechnique*. 3 vols., Paris: Mallet-Bachelier and Gauthier-Villars.

Janssen, M., J. Stachel, 2004. “The Optics and Electrodynamics of Moving Bodies,” *Preprints of the Max Planck Institute for the History of Science* 265.

Jardine, N., 1991. *The Scenes of Inquiry. On the Reality of Questions in the Sciences*. Oxford: Clarendon Press.

Ketteler, E., 1873. *Astronomische Undulationstheorie oder die Lehre von der Aberration des Lichtes*. Bonn: Neusser.

Law, J., A. Mol, 2001. “Local Entanglements or Utopian Moves: An Inquiry into Train Accidents,” published by the Centre for Science Studies and the Department of Sociology, Lancaster University at http://www.lancs.ac.uk/fass/sociology/papers/law-mol-local-entanglements-utopias-and-train-accidents.pdf

Lorentz, H. A., 1887. “De l’influence du mouvement de la terre sur les phénomènes lumineux,” *Archives néerlandaises des sciences exactes et naturelles* 21: 103–176.

Lorentz, H. A., 1895. *Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern*. Leiden: Brill.

Lorentz, H. A., 1904. “Weiterbildung der Maxwellschen Theorie. Elektronentheorie,” In: Sommerfeld, A., ed., *Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen*. Vol. 5, Part 2. *Elektrizität und Optik*. Leipzig: Teubner, 1904–1922, 145–280.

Mascart, É. É. N., 1889–1893. *Traité d’optique*. 3 vols., Paris: Gauthier-Villars.

Michelson, A. A., E. Morley, 1887. “On the Relative Motion of the Earth and the Luminiferous Ether,” *American Journal of Science, Silliman’s Journal* (3) 34: 333–345.

Nickles, T., 1988. “Reconstructiong Science: Discovery and Experiment.” In: D. Batens and J. P. Bendegem, eds., *Theory and Experiment: Recent Insights and New Perspectives on Their Relation*. Dordrecht: Reidel, 33–53.

Pinch, T., 1986. *Confronting Nature. The Sociology of Solar-Neutrino Detection*. Dordrecht: Reidel.

Poincaré, H., 1901. *Électricité et optique. La lumière et les théories électrodynamiques. Leçons professées à la Sorbonne en 1888, 1890 et 1899*. 2nd ed., Paris: Gauthier-Villars.

Radder, H., 1996. *In and About the World. Philosophical Studies of Science and Technology*. Albany/New York: State University of New York Press.

Staley, R., 1998. “On the Histories of Relativity. The Propagation and Elaboration of Relativity Theory in Participant Histories in Germany, 1905–1911,” *Isis* 89: 263–299.

Strasser, B., 1907. “Der Fizeausche Versuch über die Änderung des Polarisationsazimuts eines gebrochenen Strahles durch die Bewegung der Erde,” *Annalen der Physik* (4) 24: 137–144.

Swenson, L. S., 1972. *The Ethereal Aether. A History of the Michelson-Morley-Miller Aether-Drift Experiments, 1880–1930*. Austin and London: University of Texas Press.

Verdet, M. É., 1870. *Leçons d’optique physique*. Vol. 2 (=Œuvres, Vol. 6), Paris: Imprimerie impériale.

Wien, W., 1898. “Ueber die Fragen, welche die translatorische Bewegung des Lichtäthers betreffen,” *Verhandlungen der Gesellschaft Deutscher Naturforscher und Ärzte* 70: 49–56.

Zehnder, L. A., 1895. “Ueber die Durchlässigkeit fester Körper für den Lichtäther,” *Annalen der Physik und Chemie* (3) 55: 65–81

The Spectrum of β Decay: Continuous or Discrete? A Variety of Errors in Experimental Investigation

Allan Franklin

In this paper I will examine how the physics community came to accept the idea of a continuous energy spectrum of electrons emitted in β decay, the spontaneous transformation of an atomic nucleus into another atomic nucleus with the emission of an electron and a neutrino. In this history we shall see several different ways in which scientists went wrong. These include two instances of apparent disagreement between pairs of experimental results in which all of the results were, in fact, correct. One of each discordant pair of results had been misinterpreted. There will also be a case of an incorrect experimental result, which was an artifact produced by the experimental apparatus. It was an unsuccessful attempt to replicate this incorrect result, a discrete energy spectrum in β decay, that led to an experiment which, in retrospect, demonstrated the continuous energy spectrum. This experiment itself initially provided inconsistent results using different detectors. At the time of its publication, however, this experiment did not persuade physicists that the energy spectrum was continuous. I will examine some of the reasons for the lack of acceptance of this result and end my story with the experiment that ultimately resolved the issue. It persuaded the physics community that the energy spectrum in β decay was continuous. The process lasted approximately 30 years.

William Wilson and the Absorption of β Rays

In the first decade of the twentieth century, physicists believed that the β particles emitted in radioactive decay were monoenergetic and that such monoenergetic electrons would be absorbed exponentially in passing through matter.[1] Conversely, they also believed that if electrons followed an exponential absorption law then they were monoenergetic. There was evidence supporting this view. William Wilson, however, with some supporting evidence from other experimenters, showed

A. Franklin (✉)
Department of Physics, University of Colorado, Boulder, CO, USA

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_12,

conclusively that this view was wrong. Within a very short period of time, the physics community accepted both his experimental results and his conclusion. He also showed that the previous experimental results, on which the view of exponential absorption had been based, were, in fact, correct. They had been misinterpreted.

The Exponential Absorption of β Rays

In 1902, Walter Kaufmann (1902) had demonstrated that radium emitted electrons with a wide range of velocities. A similar result also was found by Stefan Meyer and Egon von Schweidler (1899) and by Henri Becquerel (1900). Despite the evidence provided, the physics community did not accept, at this time (the first decade of the twentieth century), that the energy spectrum of electrons emitted in β decay was continuous. There were plausible reasons for this. Physicists argued that the sources used were not pure β-ray sources, but contained several elements, each of which could emit electrons with different energies. In addition, even if the electrons were initially monoenergetic, each electron might lose a different amount of energy in escaping from the radioactive source. This view was due, in part, to what was, in retrospect, an incorrect analogy with α decay. William Bragg (1904) had argued that each of the α particles emitted in a particular decay has the same, unique energy, as well as a definite range in matter. Physicists at the time thought, by analogy with the α particles, that the β rays would also be emitted with a unique energy.[2]

It was also believed that monoenergetic electrons would follow an exponential absorption law when they passed through matter. This was a reasonable assumption for physicists of that time. If electron absorption was dominated by the scattering of electrons out of the beam, and if the scattering probability per unit length was constant, then, as Bragg had pointed out, an exponential absorption law would follow.

Early experimental work on electron absorption gave support to such an exponential law and therefore to the homogeneous (monoenergetic) nature of β rays, particularly the work of Heinrich Schmidt (1906, 1907). Schmidt fitted his absorption data for electrons emitted from different radioactive substances with a single exponential or a superposition of a few exponentials. Figure 1 shows the absorption curve that Schmidt obtained for electrons from radium B and from radium C.[3] The logarithm of the ionization (a measure of the electron intensity) decreases linearly with the thickness of the absorber, which indicates an exponential absorption law. Each curve actually consists of two straight line segments, showing the superposition of two exponentials. Schmidt interpreted this result as demonstrating that two groups of β rays were emitted in each of these decays, each with its own unique energy and absorption rate.

There was, in fact, a circularity in the argument. Physicists believed that if the β-rays were monoenergetic, then they would give rise to an exponential absorption law. If they followed an exponential absorption law, then they were monoenergetic. This association of homogenous electrons with an exponential absorption law informed early work on the energy spectrum in β decay. This was the situation in

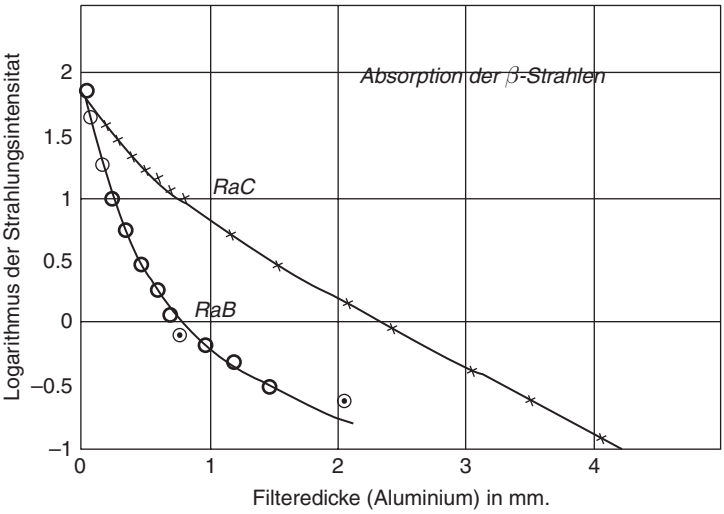


Fig. 1 Schmidt's result on the absorption of rays. The logarithm of the electron intensity (ionization) as a function of the absorber is a straight line, indicating an exponential absorption law (Schmidt 1906)

1907 when Lise Meitner, Otto Hahn, and Otto von Baeyer began their work on the related problems of the absorption of electrons in matter and of the energy spectrum of electrons emitted in β decay.[4] They first examined the absorption of electrons emitted in the β decay of several complex substances, uranium + uranium X (^{234}Th), radiolead + radium E, radium E alone, and radium. They found that the absorption of these electrons did, in fact, follow an exponential law, confirming the results obtained by Schmidt. They formulated the simple and attractive hypothesis that each pure element emitted a single group of monoenergetic β rays.

The Experiments of William Wilson

The evidential situation changed dramatically with the work of William Wilson. Wilson investigated what was, in retrospect, a glaring omission in the existing experimental program—the actual investigation of the velocity dependence of electron absorption (Wilson 1909). He noted that his "present work was undertaken with a view to establishing, *if possible*, the connection between the absorption and velocity of β rays. *So far no actual experiments have been performed on this subject* . . . . (p. 612, emphasis added)." There had been no real investigation of the issue. Wilson commented that, "It has generally been assumed that a beam of homogeneous rays is absorbed according to an exponential law, and the fact that this law holds for the rays from uranium X, actinium, and radium E has been taken as a criterion of their homogeneity (p. 612)." Wilson questioned that assumption.

Wilson used radium as the source of his electrons. He noted that Kaufmann had shown that radium emitted electrons with a wide range of velocities. Wilson, using a magnetic field, selected electrons within a narrow band of velocities—an almost monoenergetic beam—and investigated their absorption. He also varied that velocity and measured the absorption of electrons as a function of velocity. He stated his remarkable conclusion at the beginning of his paper: "Without entering at present into further details, it can be stated that the ionisation [the electron intensity] did not vary exponentially with the thickness of matter traversed. But, except for a small portion at the end of the curve, followed approximately a linear law (p. 613)." This result contradicted those of Schmidt and of Meitner, Hahn, and von Baeyer.

Wilson's results are shown in Fig. 2. The upper graph shows the ionization (not its logarithm) for various velocities as a function of absorber thickness. It is

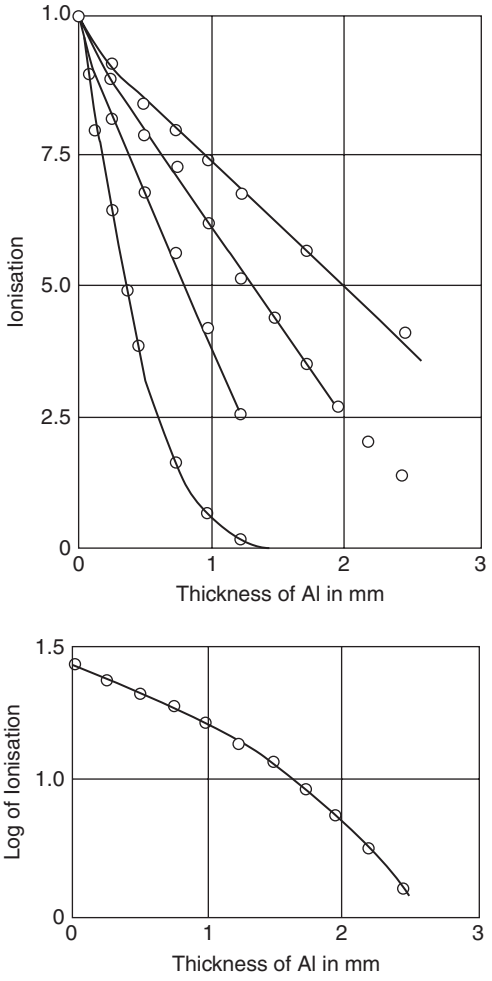


Fig. 2 William Wilson's experimental results. The *upper graph* shows the ionization, not its logarithm, as a function of absorber. It is a *straight line*, indicating a linear, not an exponential, absorption law. The *lower graph* shows the logarithm of the ionization as a function of absorber. It is not a *straight line* (Wilson 1909)

clearly linear, and not exponential. This is made clear in the lower graph in which the logarithm of the ionization is plotted against absorber thickness. If the law of absorption were exponential then this graph would be a straight line. It is not.

Wilson recognized that his result, which disagreed with all of those obtained previously, needed to be defended carefully and he did so. He reduced background effects which might have distorted his result, or, when that wasn't possible, he measured them so that they could be subtracted. He also showed that none of the effects that might have compromised his results were present. He calibrated his apparatus and obtained independent confirmation of his result using two different experimental apparatuses. He eliminated plausible alternative explanations of his result, and was left with the conclusion that it was correct. Wilson's results were credible.

How could such capable physicists like Wilson, Schmidt, and the trio of Hahn, Meitner, and von Baeyer reach such different conclusions about electron absorption? Wilson had shown that the absorption of monoenergetic electrons was approximately linear, whereas the others had found that electron absorption followed an exponential law. In retrospect, the simple explanation is that Wilson had actually measured the absorption of groups of monoenergetic electrons, each with a different energy, whereas the others had assumed that they were measuring the absorption of monoenergetic electrons when they were, in fact, measuring the absorption of electrons with a continuous energy spectrum. What makes Wilson's paper so fascinating is that he provided an explanation for these conflicting results. The other experimental results were not incorrect; they had been misinterpreted.

Wilson devoted a section of his paper to an "Explanation of the Exponential Law found by various Observers for the Absorption of Rays from Radio-Active Substances." He began

> Before entering into a discussion as to the meaning of the absorption curves obtained, it is preferable to try to explain why various observers have found that the rays from Uranium X, radium E, and actinium are absorbed according to an exponential law with the thickness of matter traversed. The fact that homogeneous rays are not absorbed according to an exponential law suggests that *the rays from these substances are heterogeneous*.
>
> (Wilson 1909, pp. 621–622, emphasis added)

Wilson then provided an explanation. He began with data from Schmidt's work that showed the ionization produced as a function of the velocity of the emitted rays. Schmidt had found a range of such velocities, but had not interpreted that result as indicating that the primary electrons were heterogeneous. He and others believed that they were emitted with a unique energy, but that they then lost energy by some unknown process. Wilson showed that the produced ionization curve varied with the amount of matter through which the electrons had passed and that the electrons lost energy in passing through matter (Fig. 3).[5] The figure shows the electron intensity as a function of momentum. Curves *a, b,* and *c* were obtained with thicknesses of aluminum of 0, 0.489, and 1.219 mm, respectively. Not only was the total ionization reduced, but the lower-velocity electrons were completely absorbed when the

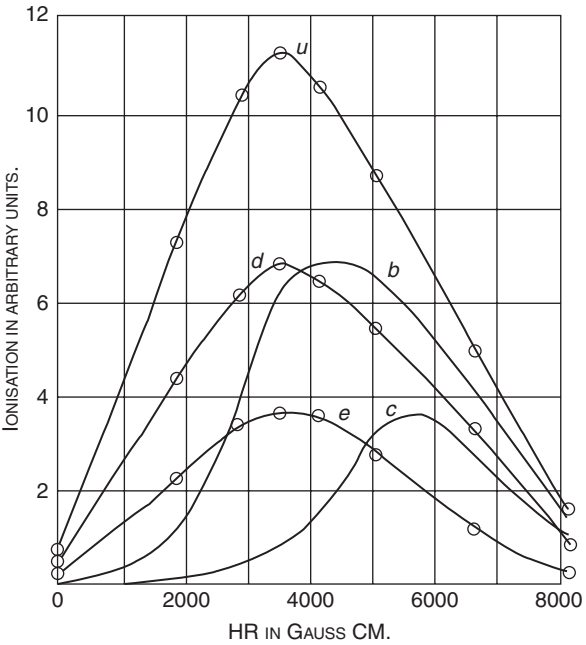


Fig. 3 Ionization as a function of momentum for different thicknesses of absorber. Curves a, b, and c are for aluminum absorbers of thickness 0, 0.489, and 1.219 mm, respectively, placed just under the electroscope. Curves d and e are for thicknesses of 0.489, and 1.219 mm placed before entering the magnetic field (Wilson 1909)

absorber thickness was increased. Curves *d* and *e* were obtained with the absorber placed before the electrons entered the magnetic field. If the electrons lose energy in passing through matter then curves *d* and *e* should be shifted to the left, as they are. Wilson calculated the absorption for various absorber thicknesses assuming that the electrons initially had a range of energies, that they lost energy in passing through matter, and that lower energy electrons were preferentially absorbed. He found that the total ionization produced by such electrons as a function of that thickness did indeed follow an exponential law. He concluded that, "It is thus clear that the exponential curve for the absorption of rays is not, as has been widely assumed, a test of their homogeneity, but that in order that the exponential law of absorption should hold, we require a mixture of rays of different types (pp. 623–624)."

Wilson had not, in this experiment, demonstrated that the energy spectrum of the electrons emitted in β decay was continuous. All he would have had to do was to measure the ionization produced as a function of electron velocity with no absorber present. He did not do so, perhaps because he was primarily concerned with the problem of absorption.

There was, however, existing evidence that disagreed with Wilson's result that the velocity of electrons diminished as they passed through matter. Schmidt (1907) had used the apparatus shown in Fig. 4 to investigate the constancy of the β-particle

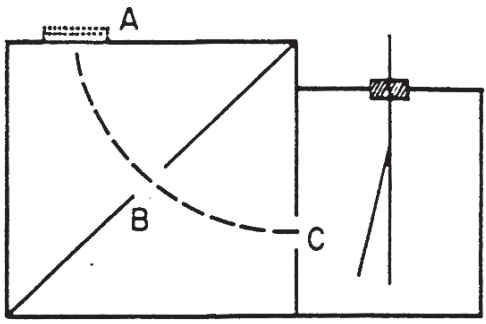


Fig. 4 Schmidt's apparatus for collimating a beam of β particles bent by a magnetic field into arc ABC (Schmidt 1907)

velocity. The β rays from a radium E source at A were bent by a magnetic field perpendicular to the plane of the paper so that they passed through a semicircular canal ABC and then passed into an ionization chamber. Schmidt adjusted the field strength to a value H_o, which resulted in the maximum ionization (i.e. the maximum number of β rays). He then placed aluminum foils between the radioactive source and the canal entrance. The β rays passed through the absorber. Once again he adjusted the field strength to obtain the maximum number of β particles. If the velocity had not changed in passing through the absorber, then, Schmidt believed, the field strength would again be H_o. It was. This cast doubt on Wilson's result that the β particles lost energy in passing through matter, a result he needed to explain why others had found an exponential absorption law.

Wilson argued that although Schmidt's experimental result was correct, his interpretation of that result was incorrect. Wilson had, in fact, already demonstrated that electrons lost energy in passing through matter (See discussion above and Fig. 3), although, as discussed below, the issue remained unresolved for a few years. He had also demonstrated that low-energy electrons were preferentially absorbed. Both points were necessary for his calculation that exponential absorption indicated that the electrons were not monoenergetic, but heterogeneous.

Once again Wilson provided an explanation of an incorrect interpretation of an experimental result.

> This experiment [Wilson's] also explains why the experiments of Schmidt apparently show no change in the velocity of the rays. According to the views expressed in this paper he [Schmidt] was dealing with heterogeneous rays and the position of the maximum should therefore move to the higher fields if the velocity of the rays does not change. [The lower energy electrons are preferentially absorbed.] The actual decrease in velocity, however, brings the maximum point back to practically the same position as before.
>
> (Wilson 1909, pp. 626–627)

The Initial Reaction of the Physics Community

Wilson's negative result on the exponential absorption of β rays received support from further work by Schmidt (1909). Schmidt inferred from his data that electrons did, in fact, change their velocity in passing through matter, confirming Wilson's

result. He also found that the electrons were not always absorbed exponentially. He did not, however, mention or cite Wilson's results. This was not the case in the paper by Hahn and Meitner (1909b) published in the same journal issue as Schmidt's paper. Hahn and Meitner argued that Wilson's results showed rather that the β-decay electrons were monoenergetic and that they did not lose energy in passing through matter. They also suggested that the energy spread in Wilson's electron beam was too large for him to draw any conclusions concerning monoenergetic β rays. Wilson (1910b) responded and argued persuasively for the homogeneity of his electron beam. He noted that his differences with Hahn and Meitner did not concern the correctness of their respective experimental results, but rather the interpretation of those results. Hahn and Meitner performed no further experiments on electron absorption and in a later account Meitner (1964, p. 6) remarked that they had realized that in order to say anything about the velocity of the electrons they had to use deflection in a magnetic field just as Wilson had done.

The question of whether electrons lose energy in passing through matter was immediately investigated and resolved in favor of Wilson. J. Arnold Crowther (1910) soon reported results on that very question. He noted that there already existed a considerable amount of indirect evidence on the subject, citing the work of both Wilson and Schmidt. Crowther investigated the question directly by measuring the velocity of electrons before and after they passed through an absorbing layer. He concluded that, "It is evident therefore that there is a small, but perceptible decrease in the velocity of the β-rays as they pass through absorbing media (p. 448)."

Crowther had shown that electrons lost energy in passing through matter. His results supported those of Wilson. Further support was provided by the experiments of J.A. Gray who, along with Wilson, was working in Manchester with Ernest Rutherford. To avoid difficulties arising from the decay of several elements in the same source, Gray used radium E (^{210}Bi), a single element source. On investigating the energy spectrum of radium E he found that "there was no sign of a set or sets of homogeneous β-rays. … The … magnetic spectrum as it may be called, shows no sign of bands, the spectrum being quite continuous (Gray 1910, p. 138)." His results showed, rather, a broad spectrum of electron velocities or energies. Gray measured the absorption of these electrons. He found that the logarithm of the intensity as a function of the thickness of the aluminum absorber "is practically a straight line," indicating exponential absorption. "[We] see that β-rays, which are very nearly absorbed according to an exponential law, are by no means homogeneous (p. 140)." This conclusion conformed to Wilson's view of exponential absorption.

Surprisingly, Gray did not emphasize the continuous energy spectrum of the electrons emitted in β decay, which seemed to be indicated by his results. One might speculate that this was due, in part, to concentrating on electron absorption. (Wilson, too, had not commented on the apparent continuous energy spectrum because of his focus on absorption.)

In a companion paper, Wilson (1910a) presented further evidence to support his view that electrons lost velocity in passing through matter. His results for various increasing absorber thicknesses are shown in Fig. 5. "It will be noticed that these maximum points move to the lower fields as the sheets of aluminum are interposed

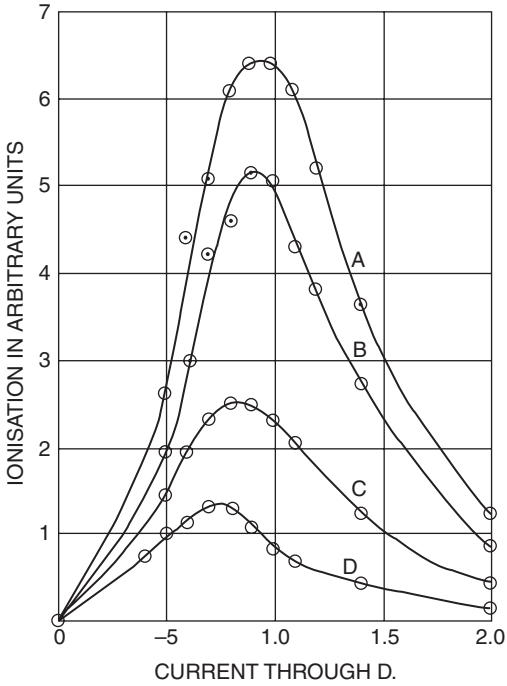


Fig. 5 Ionization as a function of velocity for increasing absorber thickness (A–D). The *thicker* the absorber the lower the maximum velocity (Wilson 1910a)

in the path, *proving conclusively that the velocity of the rays decreases by an appreciable amount as they pass through matter* (p. 145)."

Gray and Wilson (1910) showed the heterogeneity of electrons emitted from a thick layer of radium E. They remarked on the recent discussions and experimental evidence concerning the exponential absorption of electrons and on the decrease in the velocity of electrons as they pass through matter. They concluded that, "It follows as a necessary consequence of these results that β rays which are absorbed exponentially by aluminum are not homogeneous (p. 870)." They noted, however, that recent work by von Baeyer and Hahn had shown "that the β rays from several radioactive products possess a considerable degree of homogeneity." (This was a reference to the line spectra that had been found. See discussion below.)

Gray and Wilson then investigated the energy spectrum of electrons after they had passed through various thicknesses of aluminum. They found that lower velocity, or energy, electrons are absorbed more easily. "It will be noticed that the rays which produced the maximum ionization when no aluminum was placed under the electroscope are practically all absorbed by a thickness of 0.73 mm Al, while for rays corresponding to the higher fields [higher energy] appreciable quantities are still transmitted (Gray and Wilson 1910, p. 873)." Gray and Wilson also measured absorption curves for electrons of different energies directly. They found, once again, that the lower-energy electrons were more easily absorbed and that the absorption for such almost monoenergetic electrons was not exponential.

By 1911, Hahn, Meitner, von Baeyer, Gray, Wilson, Crowther and, no doubt, everyone else in the physics community were in agreement. Monoenergetic electrons were not absorbed exponentially and exponential absorption was not an indication that they were monoenergetic, but rather of a spread in energies.

The *coup de grace* was administered by Wilson. Wilson wasn't satisfied with only a calculation to show that other experimenters had misinterpreted their results on electron absorption. In subsequent experimental work he showed that an inhomogeneous beam of electrons was absorbed exponentially (Wilson 1912). He began with a monoenergetic beam of electrons and showed once again that it did not obey an exponential absorption law. He then modified the beam and made it heterogeneous by allowing it to pass through a thin sheet of platinum before striking an aluminum absorber. This resulted in an observed exponential absorption curve (Fig. 6), similar to the one he had calculated previously. He concluded:

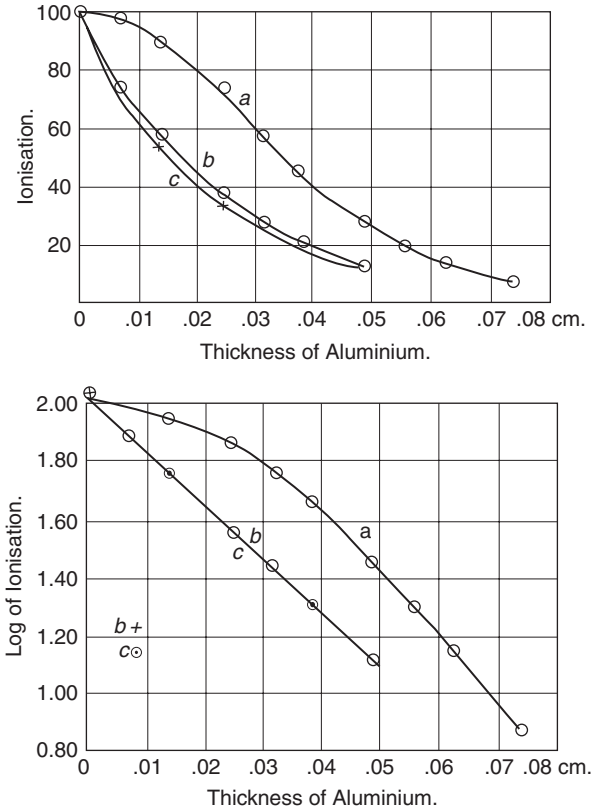


Fig. 6 Wilson's experimental graph showing the exponential absorption of a beam of inhomogeneous electrons. *Curves* b and c in the *upper graph* show the absorption of homogeneous electrons after they have passed through a platinum sheet rendering them inhomogeneous. *Curve* a shows the absorption of the homogenous electrons. In the *lower graph* the logarithm of the ionization is plotted. The exponential absorption is clearly shown for b and c (Wilson 1912)

> The fact that β-rays, initially homogeneous, are absorbed according to an exponential law after passing through a small thickness of platinum has been confirmed, and it has been shown that this is not due to mere scattering of the rays, but to the fact that the beam is rendered heterogeneous in its passage through the platinum.
>
> (p. 325)

Exponential absorption as a criterion for monoenergetic electrons was finished.

Line Spectra in β Decay?

At the same time as the above mentioned work on absorption was proceeding, Meitner, Hahn, and von Baeyer began to investigate the energy spectrum in β decay. They now used magnetic deflection, as had Wilson, to measure the electron energy (Fig. 7). Electrons emitted from the radioactive source S were bent in a magnetic field, passed through a small slot F, and then struck a photographic plate P. Electrons of the same energy would follow the same path and produce a single line on the photographic plate. The results showed a line spectrum and still seemed to

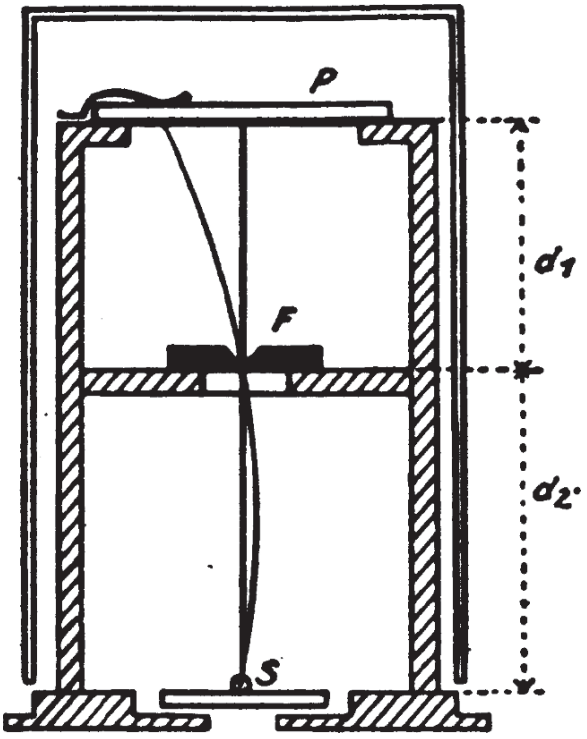


Fig. 7 The experimental apparatus used by Meitner, Hahn, and von Baeyer. The β rays emitted by the source S are bent by a magnetic field, pass through a slit at F and strike the photographic plate P (Hahn 1966)

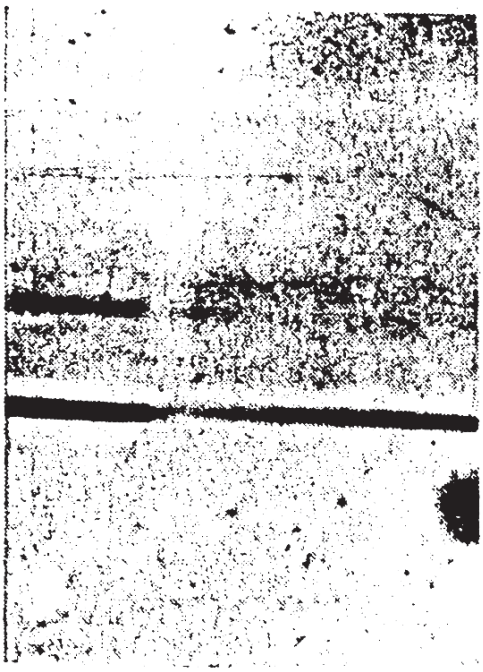


Fig. 8 The first line spectrum for β decay published by Meitner, Hahn, and von Baeyer. The two observed lines were thought to be produced by the two radioactive elements present in the source (von Baeyer 1911)

support the view that there was one unique value for the electron energy for each radioactive element. The best photograph obtained with a thorium source showed two strong lines, corresponding, the experimenters believed, to the β rays from the two radioactive substances present (Fig. 8). There were, however, some problems. There are, in addition, some weak lines in the photograph that were difficult to explain from the one energy line per element view:

> The present investigation shows that, in the decay of radioactive substances, not only α-rays but also β-rays leave the radioactive atom with a velocity characteristic for the species in question. This lends new support to the hypothesis of Hahn and Meitner …
>
> (von Baeyer et al. 1911a)

Further improvements to the apparatus, including stronger and thinner radioactive sources, improved the quality of the photographs obtained, but showed a complexity of electron velocities that made it difficult to argue for the Hahn-Meitner hypothesis. As Hahn later wrote, "Our earlier opinions were beyond salvage. It was impossible to assume a separate substance for each beta line (Hahn 1966, p. 57)." Still Meitner and collaborators retained the possibility that the observed inhomogeneity was a modification of an originally monoenergetic emission. "The inhomogeneity of fast β-rays can have its origin in the fact that the rays were initially emitted by the radioactive substance with unequal velocities. … It is more plausible to look for a secondary cause which renders inhomogeneous the emitted homogeneous rays. …

(von Baeyer et al. 1911b, p. 379)." This was a view shared by the physics community. Meitner and collaborators also conceded that the exponential absorption law "could not be a criterion for homogeneity of the radiation as Hahn and Meitner, in contrast to other scientists have assumed (von Baeyer et al. 1911b, p. 379)."

By 1911 it was clear that the energy spectrum of electrons emitted in β decay was quite complex. In a summary of work in the field, Rutherford reported that there were 29 lines in the spectrum of radium B plus radium C. Other spectra were even more complex. The general consensus in the physics community was that the energy spectrum of the electrons emitted in β decay consisted of a set of groups of electrons each with the same discrete energy, or electron lines. Although there were many lines present, the spectrum was not continuous. This evidence was similar to the discrete line spectra observed in light emitted by atoms and to the characteristic x-ray spectra obtained for atoms.

The work of Rutherford and collaborators (1914), made that view problematic. They surrounded their radioactive source with sheets of lead or gold that completely absorbed the primary electrons emitted. They found that, to within experimental error, the velocities or energies of the secondary electrons that emerged from the absorber was identical to that of the primary electrons emitted by the source. Because the primary electrons had been absorbed and removed from the beam, this suggested to the experimenters that the groups of energies observed, or the discrete energy spectrum, was not the energy spectrum of the primary electrons, but rather a secondary effect caused by the γ rays that were also emitted by the radioactive source.

Thus, there seemed to be a problem with the energy spectrum in β decay. Electrons seemed to be emitted with discrete energies, but that effect had been shown to be a secondary effect caused by γ rays. What then was the primary electron energy spectrum?

An answer was not long in coming. James Chadwick, who had worked with Rutherford in Manchester, had gone on to work with Hans Geiger in Berlin. In a letter to Rutherford he hinted at the solution. "We [Geiger and Chadwick] wanted to count the β-particles in the various spectrum lines of RaB + C and then to do the scattering of the strongest swift groups. I get photographs very quickly easily, but with the counter I can't even find the ghost of a line. There is probably a silly mistake somewhere (J. Chadwick, letter to Rutherford, 14 January 1914, Cambridge University Library)." Using Geiger's newly-invented counters they could not reproduce the line spectra found both by others and by themselves using photographic methods. This was not a failure of a new experimental apparatus, but rather a problem with the previous measurements, and is discussed below. Chadwick gave the details in a 1914 paper (1914). His apparatus is shown in Fig. 9. Electrons from the source Q pass through a slit and were both bent and focused by a magnetic field perpendicular to the paper. Only electrons of a certain velocity would pass into the detector T. The detector was either a standard ionization chamber or one of Geiger's new counters, in which the passage of a charged particle caused a large electrical discharge, which was then detected by the throw of an electrometer attached to the counter. Chadwick obtained the same results with both of his detectors (Fig. 10). Curve A is the number of β particles as a function of radius of curvature, or velocity, using the Geiger

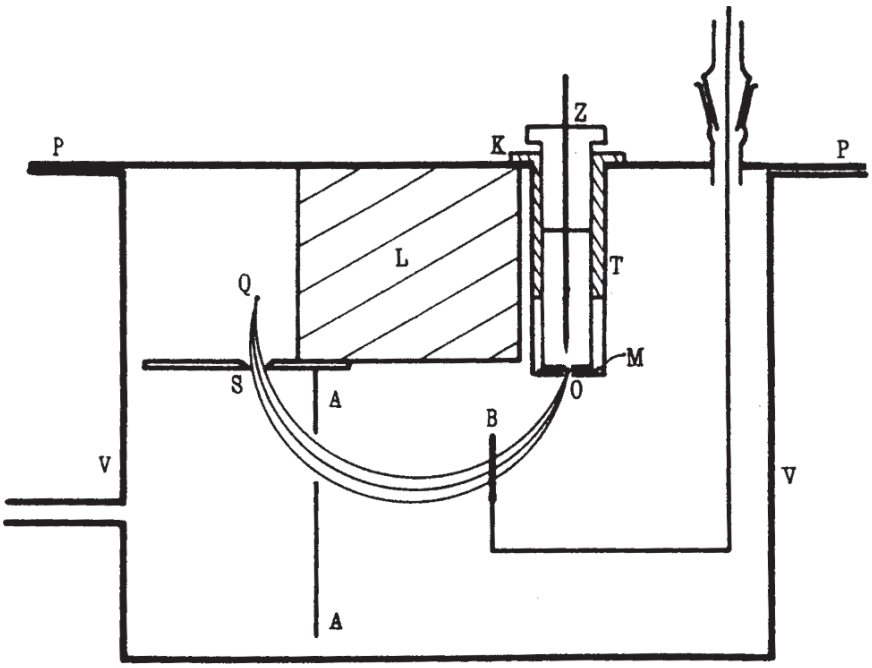


Fig. 9 Chadwick's experimental apparatus. Electrons are emitted from the radioactive source at Q and detected by the Geiger counter (ionization chamber) at O (Chadwick 1914)

counter. Curve B is the ionization produced as a function of velocity, detected with an ionization chamber. Both methods agreed, providing both independent confirmation and additional support for Chadwick's novel result. He had found four lines, identical to some found in previous spectral measurements, superposed on a larger continuous energy spectrum. As we can see the number of electrons in the lines is much smaller than the number in the continuous spectrum.[6]

How, one might ask, had the continuous spectrum been missed by all of the earlier experiments? It was due, in large part, to an artifact of the experimental apparatus. Rutherford had commented earlier that the photographic method could enhance presence of weak electron energy lines against the continuous background due to γ rays and scattered electrons. Chadwick remarked that the photographic effects of electrons with different energies was, in fact, not known and that therefore the photographic method could not be used to measure the relative intensity of various groups of electrons. He further noted that it was difficult to decide whether or not a continuous spectrum was superposed over the line spectrum. He also found that the intensity of the lines could be altered by changes in the development process of the photographic plates. Using a very slow development process he obtained a nearly black line against a clear background. Rutherford, however, offered a different explanation. In 1930 Rutherford commented that Chadwick had shown that the prominence of these groups in the photographs was due chiefly to

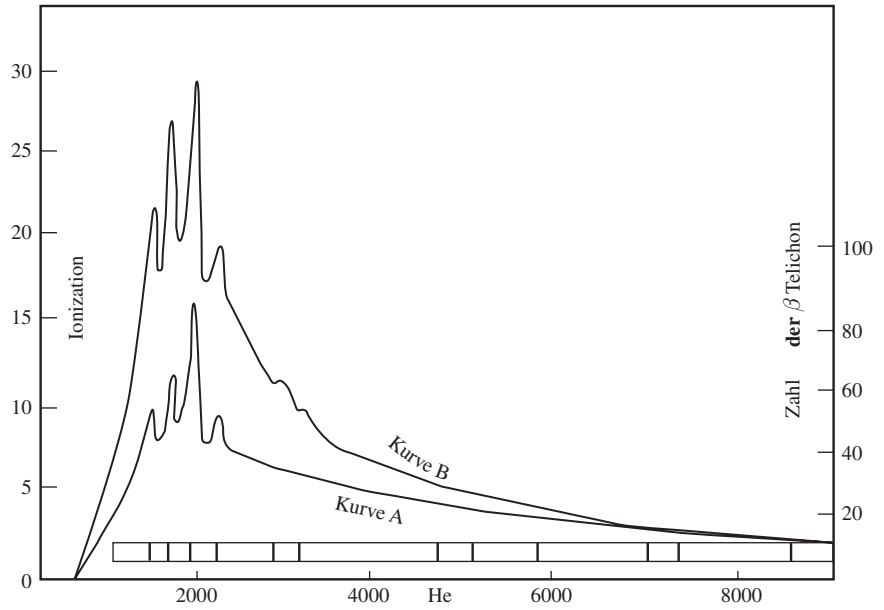


Fig. 10 Chadwick's results for the number of β rays as a function of energy. A few *discrete lines* are seen above a continuous energy spectrum. Kurve A was obtained with a geiger counter and Kurve B with an ionization chamber (Chadwick 1914)

the ease with which the eye neglects background on a plate. Whether it was due to a physiological effect or an artifact of photographic detection, the line spectrum was incorrect, although this was not recognized at the time. Chadwick had found that the energy spectrum in β decay consists of a very few lines superposed on a larger continuous spectrum.[7]

Despite the apparent decisiveness of Chadwick's experiment, not everyone within the physics community accepted the observed continuous energy spectrum as that of the primary decay electrons. In part, this was due to the fact that no other experimenter had replicated Chadwick's result, with either a radium source, the source that Chadwick had used, or with another radioactive element and, in part, because there was no theoretical explanation of the continuous spectrum. There were also, as discussed below, criticisms of Chadwick's experimental method.

Following a break in scientific activity caused by World War I, both experimental and theoretical work on the problem continued. Lise Meitner argued against the continuous spectrum on both experimental and theoretical grounds (1922a, b). She noted the complex nature of the β-decay spectrum which, in her view, contained many lines, some of which were made diffuse by the fact that the electron emitted lost energy in scattering from atomic electrons. She argued that Chadwick's experimental apparatus did not have sufficient energy resolution to resolve these lines and that this accounted for his observed continuous spectrum. She also noted that Chadwick's result had not, as yet, been replicated.

Later that year, Chadwick and Charles Ellis repeated Chadwick's original experiment and obtained the same result (1922). They considered three possible explanations for the continuous energy spectrum observed; (1) that it was due to electrons ejected by γ rays, (2) that it was due to electrons backscattered from the brass plate on which the radioactive source rested, and (3) that it was emitted as such by the radioactive atoms. "The first possibility is ruled out at once by the magnitude of the effect," which was too large to be caused by γ rays. The second possibility was eliminated by measuring the same spectrum for a source deposited on a very thin silver substrate. They found that only twenty percent of the effect could possibly be due to scattering from the brass plate. They also found that the ratio of the peaks to the continuous background was the same in both the silver and brass substrate experiments. That would be expected only if the original emitted spectrum was continuous. They concluded, "In our opinion these experiments strongly support the view that the continuous spectrum is emitted by the radioactive atoms themselves, and any theory of the β-ray disintegration must take this into account (p. 279)."

Meitner, however, argued that a quantized system such as an atomic nucleus was unlikely to emit such a continuous spectrum, citing her own previous work with Hahn and von Baeyer. The then recently proposed quantum mechanics required that an electron in an atom or the atomic nucleus can occupy only certain discrete energy states. Energy is released only when the atom undergoes a transition from one such state to another. The energy difference is also discrete and can take on only certain values. This accounts for the discrete line spectra of the light emitted by atoms, the Balmer series in hydrogen, for example. If the nucleus that emitted the electron in β decay was in one quantum state and the resulting nucleus was in another quantum state, then physicists believed that the electron emitted should also have a discrete, and unique, energy. This was indirectly supported by the fact that the γ rays emitted in radioactive decay had such a discrete spectrum, similar to that of the light emitted by atoms.

Ellis and William Wooster (1925) presented arguments against Meitner's suggested explanations of the continuous energy spectrum. They discussed several of Meitner's suggested mechanisms for the energy loss by the initially monoenergetic electrons including: (1) Compton scattering, the emission of recoil electrons of varying energy by the scattering of γ rays emitted by the nucleus from atomic electrons, (2) the emission of continuous γ rays by the electron as it passes through the intense electric fields of the atom after it is emitted by the nucleus, and (3) the scattering of the primary electrons from the planetary electrons of the atom.

Ellis and Wooster presented both evidence and argument against these possibilities and rejected all three. Compton scattering was rejected because it would have resulted in an incorrect energy spectrum for radium B and also could not explain the spectrum of radium E, which did not emit any γ rays. The absence of γ rays in the decay of radium E also argued against the continuous emission of γ rays as an explanation of the continuous spectrum. The third explanation, electron scattering, was rejected because it would result in the emission of several electrons in the β decay of a single nucleus and experiment had already shown that only a single electron was emitted in each decay.

Having eliminated all of the plausible alternative explanations of the phenomenon, Ellis and Wooster concluded,

> We are left with the conclusion that the disintegration electron is actually emitted from the nucleus with a varying velocity. We are not able to advance any hypothesis to account for this but we think it important to examine what this fact implies.
>
> (p. 860)

They also noted that there was, in fact, a direct test of whether the primary electrons lost energy as they escaped from either the atom or from the entire source.

> This is to find the heating effect of the β-rays from radium E. If the energy of every disintegration is the same then the heating effect should be between 0.8 and 1.0×10^6 V per atom and the problem of the continuous spectrum becomes the problem of finding the missing energy. It is at least equally likely that the heating effect will be nearer 0.3×10^6 V per atom, that is, will be just the mean kinetic energy of the disintegration electrons
>
> (p. 860).

The advantage of the proposed total-absorption, heating-effect experiment was that it avoided the criticisms that had been levelled at Chadwick's original experiment. These included possible differing energy losses by the decay electrons in leaving the radioactive source and the lack of sufficient energy resolution to resolve closely-spaced monoenergetic lines. Either of these effects would have the effect of transforming a line spectrum into a continuous spectrum. Later work would argue that these criticisms were unjustified.[8] At the time, however, these were important and plausible criticisms. Because the Ellis-Wooster experiment would measure the total energy emitted in β decay, both the energy of the decay electron and any energy deposited in the source, it was not open to one of the major criticisms of Chadwick's experiment. Similarly, the total absorption measurement was virtually independent of the energy resolution of the apparatus.

Ellis and Wooster wrote that they were engaged in performing this experiment, but suggested that it would be some time before they had definitive results.

One possible explanation of the continuous energy spectrum, and one rejected by Ellis and Wooster in 1925, was the possibility that energy was not conserved exactly in each β decay, but conserved only statistically in a number of such decays. Others, including Bohr, did not regard this possibility as so far-fetched. There was another possibility, one that Ellis and Wooster did not consider, that would ultimately prove to be correct. This was the suggestion that a neutral particle (the neutrino) was also emitted in β decay.

In 1927 Ellis and Wooster presented the definitive experimental result they had promised earlier (Ellis and Wooster 1927). It firmly established that the energy spectrum of electrons emitted in β decay was continuous. They did this by measuring the average energy of disintegration of electrons in the β decay of radium E, by measuring the heating effect produced by those electrons. If the energy spectrum really was continuous then the average energy obtained from the heating effect measurement would equal the average energy obtained by other methods, including ionization. If the energy spectrum was monoenergetic and the observed spectrum due to unknown energy losses, then the average heating energy measured should be at least as large

as the maximum energy measured in the continuous spectrum. For radium E the average and maximum energies were 390,000 eV (electron Volts) and 1,050,000 eV, respectively. Although Ellis and Wooster remarked that the measurement was quite difficult, they believed that they could easily measure such a large energy difference. "The experiment is difficult to carry out because large sources of radium E are not available and the heating effect is small, but owing to the great differences predicted by the rival hypotheses, it is possible to obtain a definite result (p. 112)."

They remarked that they had chosen radium E as their source of β-decay electrons because it was a radioactive source that produced no significant number of γ rays. Thus, the energy emitted was carried solely by the electrons. Noting that the average energy of disintegration could be obtained from the ionization measurements shown in Fig. 11, they continued,

> Now the average energy of disintegration can be measured by another method entirely free from any hypothesis, namely the heating effect of the β-rays. This is most simply done by enclosing a volume of radium E in a calorimeter whose walls are sufficiently thick to absorb completely the β-radiation. If the heating effect is now measured and divided by the number of atoms disintegrating per unit time, we obtain the average energy given out on disintegration. If this agrees with the value estimated from the distribution curve [Fig. 11], 390,000 V, then it is clear that the observed β-radiation accounts for the entire energy emission, and we deduce the corollary that the energy of disintegration varies from atom to atom
>
> (p. 111).

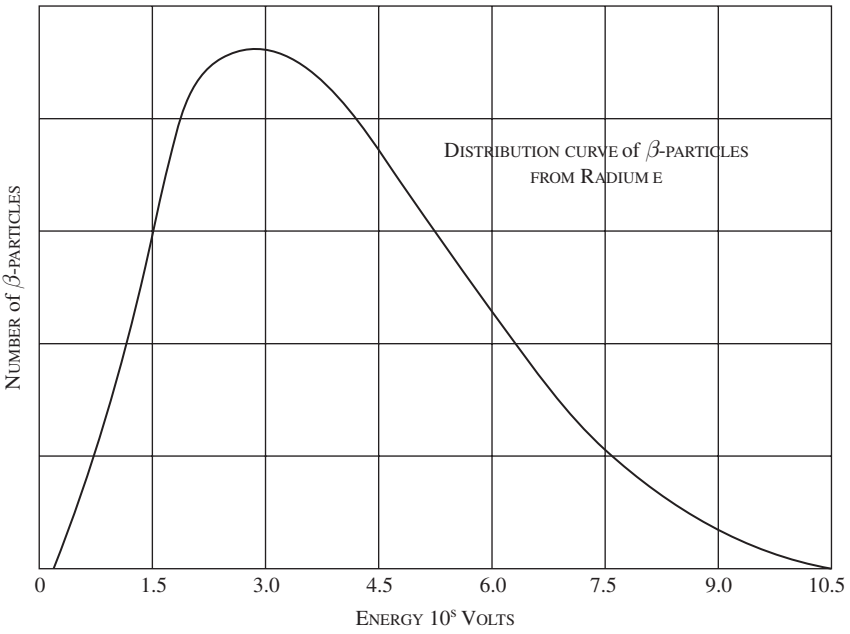


Fig. 11 The energy spectrum for electrons from Radium E measured by ionization. Radium E does not emit any γ rays (Ellis and Wooster 1927)

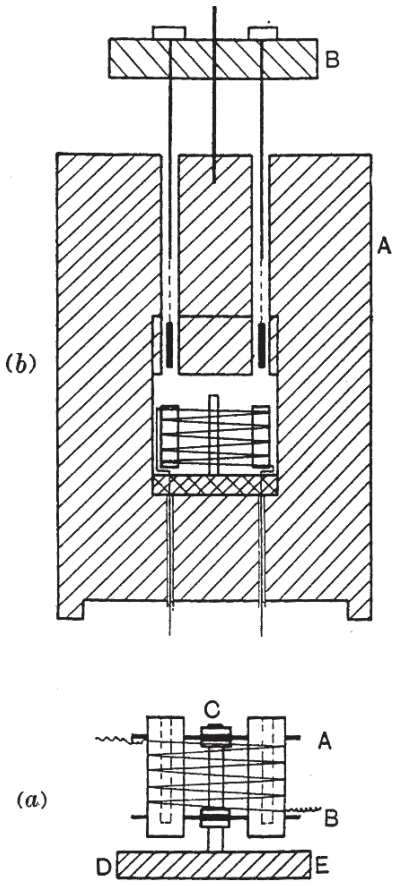


Fig. 12 The calorimeter used by Ellis and Wooster (1927)

Their experimental apparatus is shown in Fig. 12. The radium E source was deposited on a short platinum or nickel wire, enclosed in a brass case, that could be placed in, or removed from, a lead calorimeter, thick enough to absorb all of the β rays. The equilibrium temperature difference between the two calorimeters, obtained when the heat supplied by the radium E source was equal to the energy lost by the lead calorimeter, after about a time of about 3 min, was measured with a system of thermocouples, a device that produced an electric current when there was a temperature difference across it.

One further difficulty of the experiment was that the decay of radium E produces polonium, which is also radioactive, emitting an α particle. Thus, the energy deposited in the calorimeter was the sum of the energies from the β decay of radium E plus that of the α-particle decay of polonium. Although this was clearly a serious background effect it also provided an important element of their calculation of the final result. The lifetimes of radium E and polonium are 5.1 days and 139 days, respectively. From those quantities and the measured energy of each α-particle

decay, the average energy of each radium E decay could be calculated from the total heating effect.

It was absolutely crucial to determine the number of radium E disintegrations so that the average energy per disintegration could be calculated. The total absorption calorimeter precluded the counting of individual electrons, but Ellis and Wooster used the background due to the α-particle decay of polonium, discussed above, to determine the number of disintegrations. The total amount of energy due to the decay of polonium could be calculated from the measurement of the total energy due to both radium E decay and polonium decay, and the unique energy of the α particle from polonium decay could be measured. That determined the number of polonium decays, from which the number of radium E decays could be calculated. (Ellis and Wooster were unable to prepare a source that was initially completely free from polonium, but the amount of polonium initially present could be calculated.)

The final result obtained by Ellis and Wooster is shown in Fig. 13. The two curves show the total heating effect as a function of time as well as that due to polonium decay. The difference between them was the energy released by the decay of radium E. The measurements were taken over a period of 26 days and the value of the average energy of radium E decay calculated at various times. The average heating energy found was 344,000 ± 40,000 eV, in good agreement with the

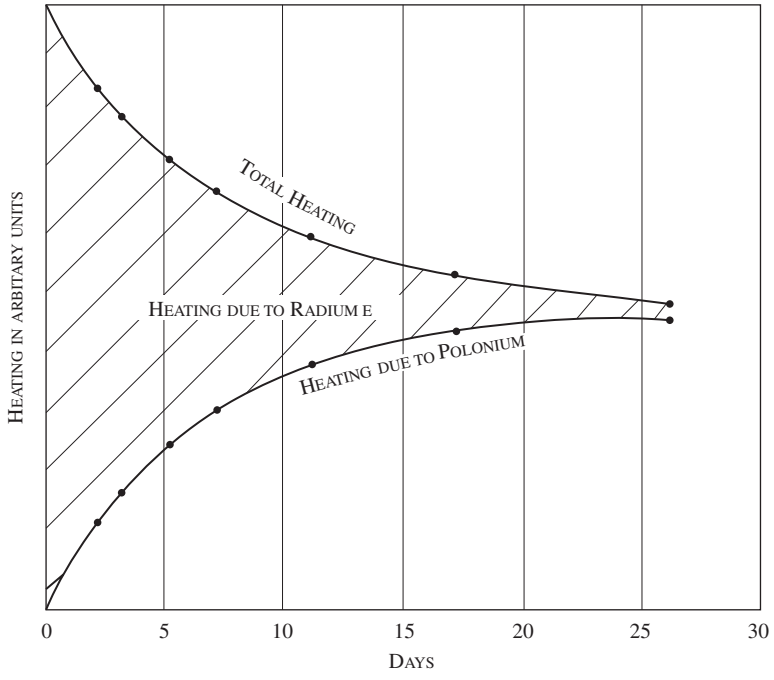


Fig. 13 The experimental result of Ellis and Wooster (1927). The decreasing total energy and the increasing energy due to the decay of polonium are shown

average value of 390,000 ± 60,000 eV obtained by the ionization measurement, and in marked disagreement with the value of more than one million electron volts expected for the monoenergetic energy hypothesis. These measurements were repeated with three other radium E sources, of varying strength, and consistent results found. In addition, if the experiment was, in fact, accurately measuring the energy of the electrons emitted in the decay of radium E, the heating effect calculated for radium E should follow an exponential decay with a period of 5.1 days. " . . . it is a most important confirmation of the accuracy of our experiments that this difference [the heating effect due to radium E] shows an exponential decay with a period of about 5.1 days (Ellis and Wooster 1927, p. 117)." (Fig. 14). The logarithm of the energy produced by radium E plotted as a function of time fits a straight line indicating an exponential decay with a period of approximately 5 days.

Ellis and Wooster concluded that, "We may safely generalise this result obtained for radium E to all β-ray bodies, and the long controversy about the origin of the continuous spectrum of β-rays appears to be settled (p. 121)." Meitner and Wilhelm Orthmann (1930) repeated the heating effect experiment with an improved apparatus and obtained an average energy per β particle of 337,000 ± 20,000 eV, in excellent agreement with that measured by Ellis and Wooster. Meitner wrote to Ellis, "We have verified your results completely. It seems to me now that there can be absolutely no doubt that you were completely correct in assuming that beta radiations are primarily inhomogeneous. But I do not understand this result at all (L. Meitner, letter to Ellis, cited in Sime 1996, p. 105)." Meitner was not alone. The energy spectrum of electrons emitted in β decay was continuous. It had taken 30 years, from the discovery of radioactivity by Becquerel to establish this fact. The question concerning the continuous energy spectrum in β decay had been answered, but the

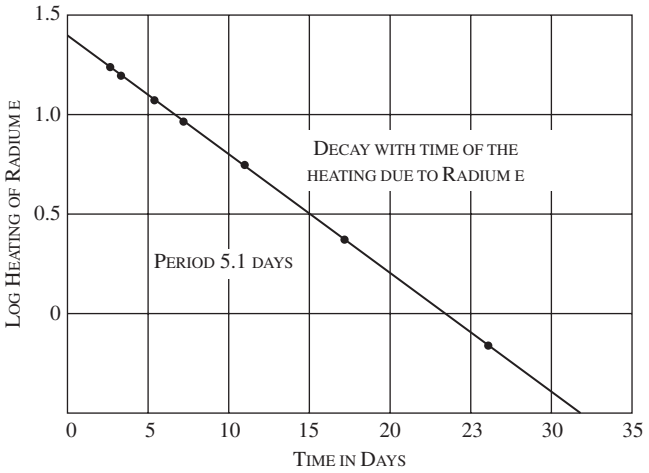


Fig. 14 The logarithm of the difference between the *two curves* in Fig. 13 as a function of time. If it is an exponential decrease due to the decay of Radium E then it should be a *straight line*. It is (Ellis and Wooster 1927)

difficulties were just beginning. No one knew why there was such a continuous spectrum. If β decay were a two-body process, as physicists at the time believed, then applying the laws of conservation of energy and of momentum required a unique energy for the electron emitted. This was clearly not the case. The conservation laws were under attack.

Discussion

Let us begin with what seems to be the most puzzling aspect of this history. Why did it take so long for the physics community to recognize that the energy spectrum in β decay was continuous after Chadwick's experiment had apparently demonstrated it? Why wasn't it accepted until after the experiment of Ellis and Wooster in 1927?

Perhaps the most important reason was that Chadwick's apparatus, as well as others at the time, measured the energy of the electron only after it had left the source, allowing for the possibility that the electron lost energy by some process in escaping from the radioactive source. Meitner proposed several possible mechanisms for that energy loss. Although the work of Chadwick, Ellis, and Wooster had argued persuasively against her proposed mechanisms, the possibility of some unknown mechanism remained. The total absorption calorimeter of Ellis and Wooster, which measured both the energy of the electrons and that deposited in the source, was not subject to the criticism that electrons were losing energy in escaping from the radioactive source, or that the experimental apparatus had insufficient energy resolution to resolve closely spaced monoenergetic lines. Using this different technique avoided those criticisms. Interestingly, no one, at the time, considered the possibility that energy might be escaping from the calorimeter. This is what was, in fact, the case. The very weakly-interacting neutrino carried away decay energy that was not detected by the calorimeter.

Another reason for the delay was the absence of any explanation for the continuous energy spectrum and the fact that a discrete spectrum seemed to be more in line with quantum mechanics. If the original nucleus was in a definite energy state, as was the final nucleus, and the final state consisted only of that nucleus and the decay electron, then the electron should have a definite energy. Some of the delay was also attributable to the slowing down of scientific work during World War I and the fact that Chadwick's result wasn't replicated until after the war.

The entire long and complex history of how the physics community came to accept the idea that the energy spectrum in β decay is continuous is one in which the perception that something had gone wrong, and its correction, played a central role. The key element was the disagreement, or apparent disagreement, between experimental results. That disagreement showed clearly that something was wrong. It should also be emphasized that this was the judgment of the participants at the time. They also agreed, after a time, about the resolution of that disagreement. With

one notable exception, the photographic detection of decay electrons, all of the discordant results were correct, one of each discordant set had been misinterpreted.

The story began with the plausible, but, in retrospect, erroneous, assumption, in analogy with α decay, that the electrons emitted in β decay were monoenergetic. Had Kaufmann and Becquerel not subscribed to this view and taken their initial observations of a wide energy range for β-decay electrons seriously, the history would have been far shorter. Their view was supported both by Bragg's argument that such electrons would be exponentially absorbed and by the experimental results of Schmidt and of Meitner, Hahn, and von Baeyer. Wilson's work on electron absorption disagreed with those results. One might speculate that an important difference between Wilson and Schmidt, Meitner, Hahn, and von Baeyer was that Wilson was investigating the subject of electron absorption, whereas the others were attempting to test the hypothesis of exponential absorption. Perhaps a strong belief in a particular view of a phenomenon may hinder a thorough investigation of the phenomenon or prevent the recognition of anomalous results.

A similar apparent discord was seen in the work of Schmidt and Wilson on the question of whether electrons lost energy in passing through matter. In both instances Wilson showed that all of the experimental results were, in fact, correct, but that one result in each pair had been misinterpreted. Wilson also demonstrated that monoenergetic electrons are absorbed linearly and that electrons do lose energy in passing through matter. Wilson both argued and experimentally demonstrated that exponential absorption showed that the electron energies were inhomogeneous. He did not, however, establish the continuous energy spectrum in β decay. Here we find an error of omission. Neither Wilson nor Gray considered the spectrum of decay electrons with no absorber present. No doubt this was because they were focused on the question of absorption.

Wilson's results led Meitner, Hahn, and von Baeyer to change their experimental technique and to use magnetic deflection of electrons with a photographic detector. Their results showed a line spectrum for electrons and supported the view that each radioactive element emitted a monoenergetic electron. Subsequent experimental work showed that the line spectra were far more complex and made the view of one element-one energy untenable. Chadwick's experiment disagreed with the line-spectrum result. His results, using ionization chambers or Geiger counters as detectors, showed weak lines superimposed on a large continuous spectrum. This experiment also had an internal disagreement. In contrast to his results obtained with ionization chambers or Geiger counters, Chadwick observed only the line spectrum with photographic detection. It was later argued that photographic detection tended to enhance the line spectrum and mask the continuous spectrum. As discussed above, even if the original energy spectrum consisted of discrete energies, a continuous energy spectrum might still be observed if electrons lost differing amounts of energy in leaving the radioactive source. Ultimately Ellis and Wooster showed that this was not correct and that Chadwick was right. The energy spectrum was continuous. It had taken 30 years.

Notes

1. $I = I_o e^{-\lambda x}$, where I_o is the initial intensity, λ is a constant depending on the absorbing material, and x is the absorber thickness.
2. If β decay were a two-body process then conservation of energy and momentum require that the electron have a unique energy.
3. The decay products of various elements were sometimes named with a letter or with a numerical suffix, and were later shown to be isotopes of other elements. Thus, radium B was an isotope of lead, ^{214}Pb; radium C was bismuth, ^{214}Bi; and radium E was ^{210}Bi.
4. Hahn and Meitner (1908a, b; 1909a, b; 1910), von Baeyer and Hahn (1910); von Baeyer et al. (1911a, b).
5. Wilson's curve was obtained with a radium source, whereas Schmidt had used uranium. Wilson also showed similar results for uranium.
6. This continuous energy spectrum is exactly what Wilson required in order to explain the observed exponential absorption law.
7. Later work showed that the line spectrum was due to one of two processes: (1) internal conversion, in which a γ ray is emitted when an atomic nucleus goes from a higher-energy excited state to at lower-energy state. That γ ray is then absorbed by an electron in the atom and ejected from that atom with the unique energy of the γ ray; (2) Auger electrons, in which the energy level left vacant by internal conversion is filled by an electron from a higher energy state. The discrete energy liberated is then absorbed by yet another electron in the atom, resulting in an ejected electron with a discrete energy.
8. Although energy loss in the radioactive source would be a problem for experiments in the 1930s and 1940s, that attempted to measure the shape of the energy spectrum in β decay it did not affect Chadwick's conclusion that the energy spectrum was continuous. (For details see Franklin (1990, Chapters 1 and 3) and Franklin (2005)). In addition, because there was, in fact, no line spectrum the energy resolution of Chadwick's apparatus would not have made such a line spectrum appear to be continuous.

References

Becquerel, H. 1900. "Sur la transparence de l'aluminium pour la rayonnement du radium." *Comptes Rendus des Seances de L'Academie des Sciences* 130: 1154–1157.

Bragg, W. H. 1904. "On the Absorption of Alpha Rays and on the Classification of Alpha Rays from Radium." *Philosophical Magazine* 8: 719–725.

Chadwick, J. 1914. "Intensitatsverteilung im magnetischen Spektrum der β-Strahlen von Radium B + C." *Verhandlungen der deutschen physikalischen Gesellschaft* 16: 383–391.

Chadwick, J. and C. D. Ellis 1922. "A Preliminary Investigation of the Intensity Distribution in the β-Ray Spectra of Radium B and C." *Proceedings of the Cambridge Philosophical Society* 21: 274–280.

Crowther, J. A. 1910. "On the Transmission of β-Rays." *Proceedings of the Cambridge Philosophical Society* 15: 442–458.

Ellis, C. D. and W. A. Wooster 1925. "The β-ray Type of Disintegration." *Proceedings of the Cambridge Philosophical Society* 22: 849–860.

Ellis, C. D. and W. A. Wooster 1927. "The Average Energy of Disintegration of Radium E." *Proceedings of the Royal Society (London)* A117: 109–123.

Franklin, A. (1990). *Experiment, Right or Wrong*. Cambridge: Cambridge University Press.

Franklin, A. (2005). The Konopinski-Uhlenbeck Theory of Decay: Its Proposal and Refutation. *Wrong for the Right Reasons*. J. Z. Buchwald and A. Franklin (eds.). Dordrecht: Springer: 209–228.

Gray, J. A. 1910. "The Distribution of Velocity in the β-Rays from a Radioactive Substance." *Proceedings of the Royal Society (London)* A84: 136–141.
Gray, J. A. and W. Wilson 1910. "The Heterogeneity of the β Rays from a Thick Layer of Radium E." *Philosophical Magazine* 20: 870–875.
Hahn, O. 1966. *Otto Hahn: A Scientific Autobiography*. New York: Charles Scribner's Sons.
Hahn, O. and L. Meitner 1908a. "Uber die Absorption der β-Strahlen einiger Radioelemente." *Physikalische Zeitschrift* 9: 321–333.
Hahn, O. and L. Meitner 1908b. "Uber die β-Strahlen des Aktiniums." *Physikalische Zeitschrift* 9: 697–704.
Hahn, O. and L. Meitner 1909a. "Uber eine typische β-Strahlung des eigentlicher Radiums." *Physikalische Zeitschrift* 10: 741–745.
Hahn, O. and L. Meitner 1909b. "Uber das Absrorptionsgesetz der β-Strahlen." *Physikalische Zeitschrift* 10: 948–950.
Hahn, O. and L. Meitner 1910. "Eine neue β-Strahlung bein Thorium X; Analogein in der Uran- und Thoriumreihe." *Physikalische Zeitschrift* 11: 493–497.
Kaufmann, W. 1902. "Die elektromagnetische Masse des Elektrons." *Physikalische Zeitschrift* 4: 54–57.
Meitner, L. 1922a. "Uber die Entstehung der β-Strahl-Spektren radioaktiver Substanzen." *Zeitschrift fur Physik* 9: 131–144.
Meitner, L. 1922b. "Uber den Zusammen hang zwischen β- und γ-Strahlen." *Zeitschrift fur Physik* 9: 145–152.
Meitner, L. 1964. "Looking Back." *Bulletin of the Atomic Scientists* 20: 2–7.
Meitner, L. and W. Orthmann 1930. "Uber eine absolute Bestimmung der Energie der primaren β-Strahlen von Radium E." *Zeitschrift fur Physik* 60: 143–155.
Meyer, S. and E. von Schweidler 1899. "Uber das Verhalten von Radium und Polonium im magnetischen Felde (I. Mitteilung)." *Physikalische Zeitschrift* 1: 90–91.
Rutherford, E., H. Robinson and W. F. Rawlinson 1914. "Spectrum of the β Rays Excited by γ Rays." *Philosophical Magazine* 28: 281–286.
Schmidt, H. W. 1906. "Uber die Absorption der β-Strahlen des Radiums." *Physikalische Zeitschrift* 7: 764–766.
Schmidt, H. W. 1907. "Einige Versuche mit β-Strahlen von Radium E." *Physikalische Zeitschrift* 8: 361–373.
Schmidt, H. W. 1909. "Beitrag zur Frage uber den Durchgang der β-Strahlen durch Materie." *Physikalische Zeitschrift* 10: 929–948.
Sime, R. L. (1996). *Lise Meitner, A Life in Physics*. Berkeley: University of California Press.
von Baeyer, O. and O. Hahn 1910. "Magnetische Linienspektren von β-Strahlen." *Physikalische Zeitschrift* 11: 488–493.
von Baeyer, O., O. Hahn and L. Meitner 1911a. "Uber die β-Strahlen des aktiven Niederschlags des Thoriums." *Physikalische Zeitschrift* 12: 273–279.
von Baeyer, O., O. Hahn and L. Meitner 1911b. "Nachweis von β-Strahlen bei Radium D." *Physikalische Zeitschrift* 12: 378–379.
Wilson, W. 1909. "On the Absorption of Homogeneous β Rays by Matter, and on the Variation of the Absorption of the Rays with Velocity." *Proceedings of the Royal Society (London)* A82: 612–628.
Wilson, W. 1910a. "The Decrease of Velocity of the β-Particles on Passing Through Matter." *Proceedings of the Royal Society (London)* A84: 141–150.
Wilson, W. 1910b. "Uber das Absrorptionsgesetz der β-Strahlen." *Physikalische Zeitschrift* 11: 101–104.
Wilson, W. 1912. "On the Absorption and Reflection of Homogeneous β-Particles." *Proceedings of the Royal Society (London)* A87: 310–325.

Part V
Surprise and Puzzlement

The Scent of Filth: Experiments, Waste, and the Set-Up

Christoph Hoffmann

Losses, Waste, and Filth

After two decades of vivid discussion, one thing seems to be indisputable: the 'life of experimentation' to which Ian Hacking once alluded is a rather happy and fruitful one. Notably, Bruno Latour has taught us that an experiment has to be understood as an 'event', which equally lets science, the scientist, and the scientific object 'acquire' a 'competence' they had not possessed before.[1] Hans-Jörg Rheinberger's definition of 'experimental systems' as 'machines for making the future'[2] is pointing in the same direction. Here, as in the case of Latour, experimentation is considered as a process of transgression, which simultaneously enriches both the research object and the researcher. For Rheinberger it is indeed 'the capacity of an excess'[3] that characterizes the particular economy of an experimental set-up in contrast to a mere technological process of reproduction. Nevertheless, as in every economy, experiments do not work without loss. Latour's own fairly unusual version of an experiment – Nicolas Baudin's survey of the Tasmanian shore at the beginning of the nineteenth-century[4] – immediately reminds us of the fact that not only did discoverers lose their life, like Baudin, but that sometimes whole expeditions disappeared on the oceans without leaving a trace of the competences they may have acquired.

Of course, in this case 'experimental loss' acquires a very literal meaning. But looking at more common sites of experimentation we will find that no biochemist or experimental physicist will be surprised by the fact that every time he or she works with an in-vitro-system or starts a set of precision measurements, contaminations and disturbances take place. More than twenty years ago Michael Lynch discovered the broad variety of such appearances 'as disclosure for the "unwitting" work of laboratory science'.[5] Artifacts, to use Lynch's collective term, served in this context as keys for observing and analyzing the processes of decision making and agreement in the course of laboratory research. My perspective is the opposite one: I am less interested in the troublesome impact such incidents can have on the activities in a

C. Hoffmann (✉)
Max Planck Institute for the History of Science, Berlin, Germany

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_13,

particular laboratory than in their potentially positive, productive agency for the research process. Contaminations and disturbances do not simply hinder or subvert the actual, real experimental work. As I argued elsewhere with respect to the enormous but unsuccessful efforts for precluding any unwanted impact that were made in the design of Physics Institutes at the end of the nineteenth century, it is exactly the other way around. Disturbances and contaminations are inescapable, intrinsic features of every experimental set-up, which offer a measure for evaluating its proper working.[6] Losses on the side of the experimental outcome that is wanted thus reappear as gains on the side of the experimenter's knowledge.

In the following section I develop in more detail the view that disturbances and contaminations do not automatically upset the researcher. Rather, they sometimes ease the mind. Such incidents in particular I propose to characterize as *experimental waste*. In general, 'waste' signifies the by-product of every productive process that necessarily appears even though nobody is interested in its appearance, except for the fact that it confirms the process of production itself. Speaking of waste has a second implication, which is connected to its obvious characteristic as a flexible category.[7] What aspects of the experimental outcome are qualified as waste, always depends on the actors' framework. What in one context may appear as experimental waste can, of course, in another context appear as a quite valuable research object that demands attention. Mostly, such dramatic changes go along with fundamental reversions in the background assumptions that organize the whole research framework. This applies, for example, to the episode from twentieth century biology that I present in the next section.

However, in certain cases it is rather the experimental situation itself that forces the 'reorientation'– in the sense of Hans-Jörg Rheinberger's contribution – both of the experimenter and of experimentation.[8] The third section of my paper will focus on a particular kind of disturbing incident, which can be characterized as *experimental filth*. Like the appearance of experimental waste, the appearance of experimental filth is closely connected to the technical circumstances of the experimental set-up. But unlike the appearance of experimental waste, that of experimental filth can not easily be attributed to common effects of the set-up or to certain malfunctions of the instruments in use. Experimental filth solely takes shape as something that is *not* in accordance with the well-known properties of the set-up. In other words, experimental filth lacks any positive definition, i.e. any definition in the available knowledge about the wanted or unwanted effects of a certain experimental set-up – and it is precisely for this reason that it may attract further attention.

In the broader context of this volume, the concepts of waste and filth can help specify two other concepts, namely, the concept of noise and Thomas Kuhn's notion of anomaly. Speaking of waste instead of noise not only underlines that such incidents are a regular by-product of the economy of every experiment. It emphasizes also that such incidents play a stabilizing role in the evaluation of the experimental outcome: waste results from the 'normal' shortcomings of the experimental operations; hence, waste is not an anomaly but part of the normal. In contrast, filth signifies an outcome that is characterized by the fact that it simply

cannot be related to the normal. Thus filth is quite the opposite of an anomaly: Because an anomaly always requires a normal against which it can take shape; filth is not an anomaly. Nor, of course, is it part of the normal.

Waste

The history of twentieth-century biochemistry offers a rather revealing example for the integration of a contamination into the economy of an experimental system. In the early 1950s the ultra-centrifugation of cell sap became a new, 'fractional' mode of representation in research on protein synthesis. In a number of runs at different speeds the original cell homogenate was separated into a microsome-rich sediment and a soluble enzyme supernatant, which contained, among other things, small amounts of RNA. From today's point of view this is not very surprising, but according to the actor's perspective RNA was only expected in the microsomal sediment of the system. Nevertheless, for the time being the discordance between biochemical background assumptions and observation did not lead to further inquiries; on the contrary, the RNA, which was found in the supernatant, started its career in protein synthesis research 'as a measure of a residual *contamination* of the soluble fraction with broken microsomal RNA.'[9]

Perhaps one or the other researcher felt uneasy with this 'contamination', at least after the whole research framework had changed from the biochemical 'synthesis paradigm' to the 'information paradigm' of molecular biology, which leads to the fact that the RNA-contamination or, shortly, the 'c-RNA' nowadays is better known as Transfer-RNA.[10] But at first the fact neither acquired the status of a Kuhnian anomaly, which does not agree with the dominant theoretical expectations in a given field of research to a given time,[11] nor did it acquire the status of a divergence in the sense that it violated a certain practical standard for the actors. Instead, by measuring the 'RNA-to-protein ratio of the non-microsomal fractions'[12] the 'c-RNA' was used as an indicator for the purity of the supernatant; that is, for the extent to which the technique of separation worked successful.

The way in which the contamination is handled here completely fulfills the above definition of experimental waste. From the actor's perspective, the 'c-RNA' was the result of the ultra-centrifugation – either because the process of separation remained incomplete or because some of the microsomes had been broken to pieces by the mechanical forces acting on them.[13] It was therefore considered as a necessary by-product of the experimental technique in use. And it was successfully set apart through its function as indicator for the purity of the supernatant fraction. Refining our definition, we can say that experimental waste forms the backside of an experimental system's 'excess'. In an experimental system, all those parts of its outcome are qualified as 'excess', which at least insinuate novelties and fall into the given research trajectory. Waste, on the other hand, is characterized by the fact that in the eyes of the researcher it does not promise any *scientific* profit. The attribute 'waste' can only be attached to those parts of an experimental outcome that for

the time being can not be combined to questions of research and do not point to a scientific problem. Excess always signifies the possible future of an experimental system, whereas too much waste sooner or later must lead to the collapse of the whole system.

Following Rheinberger's distinction between 'epistemic things' and 'technical objects' – i.e. the 'epistemic practices and material cultures', in which an epistemic thing is embedded – experimental waste has to be regarded as a technical and not as an epistemic event.[14] As already mentioned, the 'c-RNA' was considered by the actors as a by-product of the fractional procedure. In this respect the contamination seemed either unavoidable, if it resulted from the mechanical forces acting on the microsomes, or avoidable with the help of refined ultra-centrifuges, if it was only due to insufficient separation. While in the first case experimental waste appears as the consequence of a 'built-in' accident, to say it with Paul Virilio, that is, as a specific breakdown that goes along with the function of every apparatus or machine,[15] in the second case the contamination simply appears as a shortcoming that may perhaps be overcome in the near future. Nevertheless, in both cases the experimenter no longer has to speculate about the contamination. Either there is no remedy or the remedy has to be expected from the companies producing the ultracentrifuges.[16]

There is a second, even more profound way in which experimental waste has to be considered as a technological event. It is one thing that small amounts of RNA in the soluble parts of a cell homogenate appear to the actors as a contamination and it is another thing that this contamination could arrive at all on the scene of experimentation. Only with the new practice of ultra-centrifugation this kind of waste could emerge. I do not want to claim that the same thing may not happen in a different mode of biochemical representation. But the practice of ultra-centrifugal fractionation, which basically relies on the separation of soluble and insoluble parts of the homogenate, sharpens the awareness of the presence of RNA in the supernatant *as* contamination. This is exactly what subverts the purpose of the apparatus: complete separation of soluble and insoluble parts of a cell; a phenomenon that in those days implied complete absence of RNA in the supernatant. Therefore we may say that experimental waste is connected to and specified by the particular set-up in which it appears. It is shaped by the (both technical and theoretical) assumptions concerning what the set-up should guarantee in a certain experimental situation and by the observation that the result does not completely fulfill this promise.

My example illustrates that experimental waste has some unique qualities. Although it shows all the features of a disturbing impact, it has no intriguing or alarming character. From an epistemological point of view, experimental waste appears unproblematic. Neither does it call for further inquiries nor does it offer the possibility of new insights. Moreover, in a certain way, experimental waste eases the experimenter's mind. Confronted with the complex outcome of an experiment, he or she is allowed to sort out certain effects as waste due to technical circumstances. And in some cases – like the one just presented – the quantity of waste observed can even function as a measure for the success of the whole

experimental operation. Paradoxically, it seems that experimental waste is involved in the sharpening of what appears as the experiment's excess. Waste *as* loss here contributes to the stability of the experimental system insofar as the 'c-RNA' assures the experimenter of the reproductive coherence of the set-up. Not only does experimental waste comply fully with Michael Lynch's definition of 'positive artifacts' as a kind of 'standardized' feature of experimental practices,[17] but waste even plays a constitutive, positive role in demarcating the 'problematic', unclear and therefore interesting and attention-demanding part of the experimental outcome.

Neither Anomaly nor Phenomenon, nor Even Waste

Wilhelm Conrad Röntgen's discovery of the X-rays in 1895 very often figures as one of the most prominent proofs for the beneficial force of 'serendipity'.[18] It is here not the place to discuss at length the shortcomings of this notion. But it seems that 'serendipity' mainly covers up the embarrassing fact that sometimes researchers have completely failed to recognize the potential of the effects right in front of their eyes while others (or the same researcher on another day) more or less immediately appreciated what he or she was seeing. Indeed, a reader of Robert K. Merton and Elinor Barber's recently published manuscript on *The Travels and Adventures of Serendipity* (2004) can arrive at the conclusion that serendipity made its career as a narrative concept for reducing the very complex to one simple word. Instead of breaking up the apparently 'lucky event' into pieces it ties together the vague speculations about what happened to tell a convincing story of coincidence and the human, far too human factor in discovery. In this respect, a closer look at the details can provide a more satisfying picture. I will argue that with respect to the case of the X-rays what in retrospect appears as a lucky finding was in fact closely linked to the experimental situation, which became the scene of chance.

Drawing on a remark that Röntgen made in January 1896,[19] accounts of the discovery of the X-rays again and again emphasize its coincidental nature. As is well known, it was nearly the only comment that Röntgen made on the circumstances of the research work leading to his crucial observation. Even more problematic for today's historians of science is the fact that neither laboratory records nor any further notes from the bench have been preserved in the archives. However, according to the scant sources available the following situation can be imagined.[20] On the evening of the 8th of November 1895, Röntgen conducted some experiments with a cathode-ray tube in the Physics Institute of the University of Würzburg. On one of the laboratory-benches, or perhaps on the same bench as the tube, lay a piece of phosphorescent paper, which, as we will see, was commonly used as an indicator for the effect of cathode rays. Röntgen switched on the tube and the sheet of paper started to glimmer. Now it may appear that at the very moment in which the current passed through the tube, the discovery necessarily happened. As one of Röntgen's biographers concludes:

> 'I discovered by chance …', Röntgen will tell his audience weeks later. For certain, the gaze on the illuminated screen was pure chance, not a conscious search. But for sure it would have happened anyway on another evening. The pathway to discovery had already been prepared for a long time. It only needed the last few steps.[21]

Following this version, chance – inhabiting Röntgen's gaze – was only the accomplice of progress itself, which sooner or later forces nature to show its true character. No doubt this is neither a fair account of Röntgen's role nor of the role of the experimental set-up. A more refined explanation is offered by Thomas Kuhn, who suggests with explicit reference to 'accidental discoveries' like 'Roentgen and X-rays':

> It is probably the ability to recognize a significant anomaly against the background of current theory that most distinguishes the successful victim of an 'accident' from those of his contemporaries who passed the same phenomenon by.[22]

But in the case of Röntgen there could be no anomaly, because the investigation of cathode rays that Röntgen had initially explored following the footsteps of Philipp Lenard was far away from a state of mature theoretical understanding. On the contrary, it was first and foremost the debate of Lenard's experimental work, published since 1894, which stimulated the future framing of this research field in terms of the concept of the electron.[23] Furthermore we have to acknowledge that Röntgen's 'contemporaries' did not 'pass by the same phenomenon'. The considerable number of non-discoverers, who later on mourned their fate in public,[24] did not observe a phenomenon at all; at least if we admit that only those observed effects which are related unambiguously to the operations of the set-up can be called a phenomenon.

Perhaps the most instructive case is the one often quoted of Arthur Willis Goodspeed, a professor of physics at the University of Pennsylvania, who, in March 1896, proudly but some months too late published 'the *first* Röntgen picture' ever made.[25] With regards to its production, it is reported that six years before, at the end of February 1890, Goodspeed had been trying to fix the spark produced by the discharge of a powerful induction machine

> directly upon the sensitive plate, without any camera. Incidentally also the impressions of coins were obtained by sparking them when in contact with the sensitive film. After these experiments had been completed, a number of Crookes tubes were brought out and operated for the pleasure and amusement of Mr. W. N. Jennings, in connection with whom the work had been done.[26]

Afterwards the exposed plates were developed and it so happened that on one single plate 'very mysterious discs' showed up, which looked 'quite different' from the appearance of the coins on the spark pictures.[27] Not finding an explanation for the effect, Goodspeed put the plate aside until Röntgen's findings told him what had once been registered by the photographic process.

That ultimately 'Goodspeed finds himself a discoverer without discovery',[28] may have various reasons. But one thing is certain, namely, that the relation between the plate at hand and the ensemble of apparatuses and experimental operations that produced the effect was pretty blurred. Not only did the discharge experiments

and the experiments with the tube include very different apparatuses and processes, which all had to be considered. In addition, the effect was observed only days after the experiments had been carried out, so that the number of causes, to which it could have been attributed, was once more multiplied. Hence, the only thing Goodspeed noticed, was a 'mysterious' appearance, or, to say it with Michael Lynch, a 'situated distortion',[29] which, however detailed the representation on the photographic plate was, lacked any sharp, distinct, and replicable experimental definition. Thus he neither put away a phenomenon nor even a piece of waste – because for qualifying the appearance as waste Goodspeed must have been able to identify or at least to imagine its source.

Filth

In the case of Röntgen the situation is completely different. As already indicated he was originally engaged in the reproduction of a number of experiments on cathode rays, which had been published by Philipp Lenard in the spring of 1894.[30] It therefore seems likely that precisely the limited character of his efforts created a 'certain disposition for discovery'.[31] Given this, Röntgen became a discoverer by being not interested in observing novelties, while Lenard and other specialists in cathode rays always followed some particular goal that directed their attention to specific aspects. But it should be considered that this psychological advantage can only be crucial if the experimenter is able to evaluate the effects he or she is observing. Röntgen's reproduction of Lenard's experiments was an advantage primarily because through it he learned in detail what kind of effects he was to expect in a given situation and what kind he was not to expect.

Like almost every detail of Röntgen's research work the question of what kind of tube he really used has been amply discussed. In his first publication 'Über eine neue Art von Strahlen' Röntgen mentioned August Hittorf's tubes as well as Lenard's and William Crookes', but he did not specify which kind of tube was involved in the decisive observation.[32] However, we know for certain that Röntgen bought a cathode-ray tube like Lenard's as early as in the spring of 1894 and, as Albrecht Fölsing has convincingly argued, some of Röntgen's remarks on his initial findings make sense only with respect to this type of tube.[33] Compared with ordinary tubes, Lenard's apparatus had a decisive novelty. A little 'window' consisting of a piece of very thin aluminum foil was inserted in the glass body of the tube that allowed the rays to pass through. As Lenard emphasized:

> Thus the observation of the rays can take place outside the discharge tube in the open; they [the rays] can be investigated in any medium. But in particular the observation and the production of the rays thus become completely independent from each other; the circumstances of the former can be varied without varying those of the latter.[34]

And what is more, the design of the tube allowed Lenard to cover the whole tube except the window with a brass case, which protects the space of observation

'against the light and the electric forces of the discharge.'[35] Thus we have to recognize the trivial fact that the obscure glimmering in Röntgen's laboratory could not be caused by the flickering appearances of light, which, as is known, accompany the production of the cathode rays.

Nevertheless, we need to consider some more details in order to understand why this effect triggered the experimenter's attention. First of all, it is necessary to underline that a glimmering phosphorescent paper was nothing uncommon in a laboratory in which experiments with a cathode-ray tube were carried out. For Lenard (and not only for him) these sheets functioned as a test object in his exhaustive investigations of the intensity and propagation of the rays in different gases. Some observations in regular air mentioned right at the beginning of the report deserve our particular attention here. With these observations Lenard proved that the peculiar glimmer that 'bodies capable of phosphorescence' displayed when they were placed close to the window lost its intensity with increasing distance until it disappeared at a distance of 6–8 cm.[36] No less interesting are a number of improved observations to which Lenard points some twenty pages later. In order to measure the distances, a small moveable phosphorescent screen was placed in a glass body directly connected to the window. The result was that, when the glass body was filled with regular air, this time the cathode rays passed for hardly more than 2 cm.[37]

One might wonder why Lenard himself did not realize the obvious discrepancy with the experimental findings that he had mentioned earlier. Actually, if we admit that the field of research was still in a state of 'explorative experimentation' – and Lenard's work seems to be a good example for gaining conceptual ground through experimental variation and repetition[38] – it should make an even stronger difference whether an effect disappears in the same gas in a distance of six to eight or in a distance of 2 cm from the window of the tube. (And this was not the only surprising finding: for example, Lenard also noticed that photographic dry-plates 'are completely darkened in a few seconds even at larger distances'[39]) But the question is not why Lenard joined the illustrious party of non-discoverers. The question is why Röntgen's eyes became attracted by a glimmering piece of paper. One part of the answer is already given: of course, the effect had to be expected if the experimenter placed the paper directly in front of or at least close to the window. But from all we know, Röntgen did not hold the sheet of phosphorescent paper up to the window nor did he even hold it in his hands. It was merely lying around in the neighborhood of the tube. If we take into consideration that the tube was mounted on a support, it is more than likely that the paper even in the most conceivable case of proximity must have been situated in considerable distance to the window. Accordingly, if the paper was glimmering under these circumstances, this did not agree very well with the experiences presented in Lenard's report.

The next detail that has to be considered is once more a rather trivial one, namely the circumstance that the cathode rays can only pass the tube trough the window. This was the prerequisite of the whole experimental work and, of course, it was a prerequisite that did not need to be debated any further. Heinrich Hertz's investigations on the qualities of metal foil, which ultimately lead to the design of Lenard's tube, were precisely inspired by the problem that glass was not permeable for

cathode rays.[40] Nevertheless, even a fact that was so self-evident was recalled to the mind of Lenard's readers. Referring to the observation that an illuminated screen starts to darken at the very moment when the edge of a quartz plate touches the margin of the window, he concluded: 'the hitherto presumed idea that all effect only comes from the small window, through which the rays pass, is true.'[41]

None of the three details that I have just discussed: the covering of the tube, the very fast diffusion and absorption of cathode rays in regular air, and the impermeability of the glass-body for cathode rays, was in itself remarkable, nor were they in any way new. On the contrary, Röntgen could take all this from Lenard's publications and the reproduction of Lenard's experiments. These details only gain significance in the particular situation in Röntgen's laboratory on the evening of November 8th, 1895. However one last and indeed decisive detail has to be added: namely the fact that not a photographic plate but a little piece of phosphorescent paper was lying around in the neighborhood of the tube. In fact this circumstance was in two ways a prerequisite of the whole event: first because the glimmering of the paper signaled that an effect was taking place, and secondly because this effect could be observed and reproduced immediately. In contrast to the accidental darkening of a photographic plate, which in principle can be discovered only in retrospect and therefore, as the example of Goodspeed has taught, poses the problem of relating it to a particular experimental situation, the glimmering paper directly allowed to explore what was going on here.

Of course, an ultimate answer can not be given. But a closer look at the circumstances enables us to understand that Röntgen, observing the paper, neither stumbled over an anomalous experimental result, nor did he notice a mere disturbance. What he found was – to say it with Michel Foucault – a 'pure distance', an 'interstice',[42] which acquired its characteristic *as something quite different* against the technical properties of the tube and some earlier experiences made with this set-up. A glimmering phosphorescent paper clearly indicates to the experimenter that at this very moment an effect must be acting on it, but as discussed above this effect can neither be attributed to the appearances of light, which were caused by the cathode rays in the tube, nor to the cathode rays themselves.

To put it slightly differently: What Röntgen discovered was an effect where he would and could not imagine one. By this I do not want to suggest that the glimmering paper 'per se' – in a literal sense – was lying around in a place where it did not belong. We must instead consider that expectation is always dependent on an individual context. In this case, only the specificity of the set-up and the knowledge connected to it defined the glimmering paper as a paper glimmering, where a well-trained experimental physicist like Röntgen did not expect it to do so. Contrary to Kuhn's claim that 'the anomalous glow which provided Roentgen's first clue was clearly the result of an accidental disposition of his apparatus',[43] it was precisely the disposition of the paper in relation to the apparatus which caused the experimenter's attention. As Michael Thompson in his *Rubbish Theory* once pointed out: 'We only notice rubbish when it is in the wrong place',[44] i.e. the same effect observed with the paper right in front of the tube's window would not have provoked any surprise.

However, Thompson's claim has to be refined because actually waste or rubbish 'in the wrong place' does not exist. Either it is in the right place or it is no waste or rubbish, but something which is marked by a certain degree of uneasiness. In the case of Röntgen, this uneasiness was clearly dependent on the experience that the effect observed solely showed negative characteristics: In the beginning it stood for nothing more than what it did *not* represent. The proper name for such an incident has already been introduced: What Röntgen encountered was experimental filth. Filth is different from waste, but not identical with an anomaly. It is an incident that was not aimed for (like waste), but (unlike waste) it can not be credited to the disturbing powers of the set-up in the same way as the 'c-RNA' could be credited to the incompleteness of the separating process. Therefore filth starts to annoy the experimenter (like an anomaly); what appears can not be silenced. But the annoyance can not be limited to a particular aspect of the research object at stake (unlike an anomaly). The incident does not point to a theoretical framework to which it could be related as discrepancy or 'outlier'. It is not demarcated by an established normal background, against which it could acquire the characteristics of a well-defined irregularity. On the contrary, bestowing the scent of filth with properties means precisely discovering the horizon (the normal), which is experimentally confirmed by the effect. As we know, this is exactly what Röntgen pursued in the weeks before the publication of his findings in the last days of December 1895.

Summary

The two examples I have discussed have shown that if the historical and philosophical reflection evaluates the outcome of experiments only in the 'enriching' terms of excess it misses out on some fundamental conditions of experimental knowledge-production. At first glance, disturbing incidents can play quite prominent roles in the emergence of scientific knowledge.[45] As the example of the 'c-RNA' suggests, experimental waste may function as an indicator for the proper working of the experimental system and, by this, may contribute to the stabilization of the system's excess.[46] From the glimmering paper in Röntgen's laboratory, however, we can learn that effects may appear in the course of experimentation that neither can be explained 'away' as experimental waste nor do they show the features of well-defined anomalies. Filth does not enrich the given object of research with positive or negative properties. The scent of filth primarily marks an experience of sheer inconsistency, which might lead to the displacement of the whole enterprise (and in consequence to a new research object) or might disappear in the next moment. As a matter of fact not in every instance the uncovering of experimental filth provokes further profit. Sometimes it simply points to a source of disturbance, which had been ignored until then.

The argument of my paper underlines that conceptual accurateness is a requirement for understanding the contribution of disturbing incidents to the fruitful life of experimentation. To classify such effects simply as noise – as it is common usage

in the language of the laboratory – fails to acknowledge the fact that the appearance of noise can have quite diverging consequences. Of course one can differentiate between 'normal' and 'a-normal' noise, i.e. between noise that can be related to a certain source and noise that can not. But to speak here of a-normal noise masks the fact that sometimes it is exactly the lack of a normal that characterizes an effect as disturbing. Indeed, Kuhn's concept of anomaly loses its explanatory power in those situations in which the effect observed shows characteristics that are incompatible to the 'normal' objects of the field of research in which it appears.

The two concepts of experimental waste and experimental filth offer a more appropriate differentiation in this respect. Experimental waste always appears *inside* the given field of research and is evaluated as a divergence from what an experimental set up should do in a particular situation of experimentation. In turn, experimental filth appears *outside* the given field of research and is evaluated as a divergence from what can be expected from an experimental set-up in a particular situation of experimentation (in fact, the outside of a certain research agenda only becomes imaginable by such filthy incidents). In the first case the experimental operations result in less than what was wanted, in the second case in more than what was conceivable. To come to an end: In the total outcome of a particular experiment filth signifies an event of 'non-comprehension'[47] in its strictest meaning – namely an event that, although sharply defined, points to the complete absence of a reference system in which it can become comprehensible. Nevertheless, in one way or another both waste and filth contribute to the fruitful life of experimentation. But no doubt only the later stimulates additional research action.

Acknowledgments For assistance, comments and suggestions I am very grateful to Hans-Jörg Rheinberger, Caroline Welsh, Giora Hon, Friedrich Steinle, and above all to Jutta Schickore. The research was funded by the Max-Planck-Institute for the History of Science in Berlin.

Notes

1. Cf. Bruno Latour, The Force and Reason of Experiment, in: Homer E. Le Grand (ed.), *Experimental Inquiries. Historical, Philosophical and Social Studies of Experimentation in Science*, Dordrecht, Boston, London: Kluwer Academic Publishers, 1990, pp. 49–80, 65.
2. Hans-Jörg Rheinberger, *Towards a History of Epistemic Things. Synthesizing Proteins in the Test Tube*, Stanford: Stanford UP, 1997, p. 80. See also the whole of chapter "Experimental Reorientations".
3. Ibid., p. 183.
4. Cf. Latour, The Force and Reason of Experiment, pp. 53–59.
5. Michael Lynch, *Art and Artifact in Laboratory Science. A Study of Shop Work and Shop Talk in a Research Laboratory*, London, Boston, Melbourne, Henley: Routledge & Kegan Paul, 1985, p. 84.
6. Cf. Christoph Hoffmann, The Design of Disturbance. Physics Institutes and Physics Research in Germany, 1870–1910, *Perspectives on Science*, Vol. 9, 2001, pp. 173–195, 190.
7. Cf. Michael Thompson, *Rubbish Theory. The Creation and Destruction of Value*, Oxford: Oxford UP, 1979.
8. Rheinberger, this volume, p. 76.

9. Rheinberger, *Towards a History of Epistemic Things*, p. 145. For the practice of ultra-centrifugation see ibid., pp. 86–95.
10. Cf. ibid., p. 93.
11. Cf. Thomas Kuhn, *The Structure of Scientific Revolutions* (1962), 2nd. edition, Chicago, London: University of Chicago Press, 1970, ch. VI.
12. Rheinberger, *Towards a History of Epistemic Things*, p. 93.
13. Ibid.
14. Cf. ibid, pp. 28–31.
15. Cf. Paul Virilio, Der Urfall. (Accidens Originale), *Tumult: Zeitschrift für Verkehrswissenschaften*, Vol. 1, 1979, pp. 77–82, 77.
16. For the vital interaction between biochemists and instrument-makers in the beginnings of analytical ultra-centrifugation see Boelem Elzen, *Scientists and Rotors. The Development of Biochemical Ultracentrifuges*, Enschede: PhD, 1988.
17. Cf. Lynch, *Art and Artifact in Laboratory Science*, p. 91.
18. See David Stephen Halacy Jr., *Science and Serendipity. Great Discoveries by Accident*, Philadelphia: Macrae Smith Company, 1967, ch. 5; Gilbert Shapiro, *A Skeleton in the Darkroom. Stories of Serendipity in Science*, San Francisco: Harper & Row Publishers, 1986, ch. 1; Royston M. Roberts, *Serendipity. Accidental Discoveries in Science*, New York: John Wiley & Sons, 1989, ch. 21.
19. Cf. Albrecht Fölsing, *Wilhelm Conrad Röntgen. Aufbruch ins Innere der Materie*, München, Wien: Hanser, 1995, p. 225.
20. With all necessary caution I rely here on the account of H. J. W. Dam, A Visit to Professor Röntgen at His Laboratory in Würzburg. His Own Account of His Great Discovery (1896), in: W. Robert Nitske, *The Life of Wilhelm Conrad Röntgen. Discoverer of the X Ray*, Tucson/Arizona: The University of Arizona Press, 1971, pp. 126–137, 134.
21. Hans Leicht, *Wilhelm Conrad Röntgen. Biographie*, München: Ehrenwirth, 1994, p. 73.
22. Cf. Thomas Kuhn, The Function of Measurement in Modern Physical Science (1961), in: *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago, London: Chicago UP, 1977, pp. 178–224, 204.
23. Cf. Olivier Darrigol, *Electrodynamics from Ampère to Einstein*, Oxford: Oxford UP, 2000, pp. 300–310.
24. See, for example, Lord Rayleigh, Some Reminiscences of Scientific Workers of the Past Generation; and Their Surroundings, *The Proceedings of the Physical Society*, Vol. 48, Part 2, 1936, pp. 217–246, 240.
25. Arthur W. Goodspeed, The Röntgen Phenomena. A Few Early Results Obtained at the University of Pennsylvania, *Science*, N. S. Vol. 3, No. 63, March 13th, 1896, pp. 394–396, 396.
26. Ibid., p. 395.
27. Ibid.
28. Peter Geimer, Noise or Nature? Photography of the Invisible Around 1900, in: Helga Nowottny and Martina Weiss (eds.), *Shifting Boundaries of the Real: Making the Invisible Visible*, Zürich: vdf Hochschulverlag, 2000, pp. 119–135, 131.
29. Cf. Lynch, *Art and Artifact in Laboratory Science*, p. 100.
30. Cf. Fölsing, *Wilhelm Conrad Röntgen*, pp. 139–141.
31. Ibid., p. 126f.
32. Cf. Wilhelm Conrad Roentgen, Ueber eine neue Art von Strahlen. Vorläufige Mittheilung, *Sitzungsberichte der Würzburger physikalisch-medicinischen Gesellschaft zu Würzburg*, 1895, pp. 137–147, 137.
33. Cf. Fölsing, *Wilhelm Conrad Röntgen*, pp. 146f and 140f. Otto Glasser, Röntgen's first biographer, also assumed that the crucial observation took place with a tube like Lenard's (see Otto Glasser, What Kind of Tube did Röntgen Use when he Discovered the X-Ray?, *Radiology*, Vol. 27, 1936, pp. 138–140, 140).
34. Philipp Lenard, Ueber Kathodenstrahlen in Gasen von atmosphärischem Druck und im äussersten Vacuum, *Annalen der Physik und Chemie*, Vol. 287 (= N.S. Vol. 51), 1894, pp. 225–267, 226.

35. Ibid., p. 228.
36. Ibid., p. 229f.
37. Cf. ibid., p. 252f. See also Philipp Lenard, Ueber die Absorption der Kathodenstrahlen, *Annalen der Physik und Chemie*, Vol. 292 (= N. S. Vol. 56), 1895, pp. 255–275.
38. Cf. Friedrich Steinle, *Explorative Experimente. Ampère, Faraday und die Ursprünge der Elektrodynamik*, Stuttgart: Franz Steiner Verlag, 2005, p. 314f.
39. Lenard, Ueber Kathodenstrahlen in Gasen von atmosphärischem Druck, p. 237.
40. Cf. Heinrich Hertz, Über den Durchgang der Kathodenstrahlen durch dünne Metallschichten, *Annalen der Physik und Chemie*, Vol. 281 (= N. S. Vol. 45), 1892, pp. 28–32, 28.
41. Cf. Lenard, Ueber Kathodenstrahlen in Gasen von atmosphärischem Druck, p. 235.
42. Cf. Michel Foucault, Nietzsche, History, Genealogy (1971), in: Paul Rabinow (ed.), *The Foucault Reader*, translated by Donald F. Bouchard and Sherry Simon, London: Penguin Books, 1991, pp. 76–100, 85.
43. Thomas Kuhn, The Historical Structure of Scientific Discovery (1962), in: *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago, London: Chicago UP, 1977, pp. 165–177, 173.
44. Thompson, *Rubbish Theory*, p. 91f.
45. Cf. the contributions to the special issue "Secondary Matters: On Disturbances, Contaminations, and Waste as Objects of Research", *Perspectives on Science*, Vol. 9, No. 2, 2001.
46. Cf. Rheinberger, *Towards a History of Epistemic Things*, p. 80.
47. Rheinberger, this volume, p. 77.

In the Thick of Organic Matter

Ursula Klein

Introduction

Between summer 1806 and spring 1808 the French chemist Louis Jacques Thenard (1777–1857) performed a series of experiments with ethers. At the time, ethers had already a history. The properties of pure ether, prepared from spirit of wine and sulfuric acid, were first described in an article by the German chemist August S. Frobenius, published in the *Philosophical Transactions* in 1730. As Frobenius had commercial interests, he kept the method of preparation secret. Together with his English colleague Godfrey Hanckwitz he wanted to sell pure ether as a novel, most effective remedy. But immediately after Frobenius' publication several British, French and German chemists and apothecaries set out to reproduce pure ether, and in 1741 Cromwell Mortimer made the production process public.[1] Twenty years later, chemists and apothecaries had tested further possibilities of producing ether from spirit of wine and acids other than sulfuric acid, such as nitric acid, muriatic acid (today hydrochloric acid), and acetic acid.[2] Hence the article "éther," published in Pierre Joseph Macquer's famous *dictionnaire de chimie*, mentioned different ethers, such as "ordinary" or "vitriolic ether," "nitric ether," "acetic ether" and so on (Macquer 1766, 1:455–470).

By the end of the eighteenth century the two chemical teachers of Thenard, Antoine-François Fourcroy and Louis N. Vauquelin, undertook collaborative efforts to explain the formation of ordinary ether.[3] Since the operation yielded many byproducts of ether, and since sulfurous acid was one of the byproducts, many chemists believed that ordinary ether resulted from the oxidation of alcohol (contained in spirit of wine) by sulfuric acid; according to this understanding the sulfurous acid was the reduction product of sulfuric acid. This explanation was congruent with chemists' more general views about the action of strong mineral acids on organic substances. For example, in his chemical textbook Fourcroy pointed out that strong acids "engender a profound alteration of plant materials. When they are power-

U. Klein (✉)
Max Planck Institute for the History of Science, Wilhelmstraße 44, 10117 Berlin, Germany
e-mail: Klein@mpiwg-berlin.mpg.de

G. Hon et al. (eds.), *Going Amiss in Experimental Research*, Boston Studies in the Philosophy of Science 267, DOI 10.1007/978-1-4020-8893-3_14,

ful and concentrated, they always change the equilibrium of these materials' composition [and] denaturate them" (Fourcroy 1801–1802, 7:48).[4] He added that this denaturation often consisted in the oxidation of the organic materials, sometimes observable in their inflammation (ibid., 92–96). However in the case of alcohol and sulfuric acid, Fourcroy claimed, the reaction was not an oxidation. In their 1797 experiments, Fourcroy and Vauquelin had found eight different products of the reaction of alcohol with sulfuric acid: ordinary ether, sulfurous acid, carbonic acid, water, "sweet oil of wine," acetic acid, "oil building gas" (or "olefiant gas") and carbon. Based on their observation that these products were not produced simultaneously, the two chemists divided the operation into three phases, claiming that only the multiple byproducts of ether – among them sulfurous acid – were products of oxidation, which took place in the second and third phase of the operation when temperature was increased. By contrast, in the first phase and at lower temperatures, the ether was created, along with water and carbon, but without the simultaneous creation of sulfurous acid or other products of oxidation. Fourcroy and Vauquelin then explained the formation reaction of ether as a decomposition of alcohol, in which sulfuric acid withdrew hydrogen and oxygen from alcohol in such proportions to form water and the ether.[5] Thus, in accordance with the common understanding of the action of strong acids on organic substances, Fourcroy and Vauquelin explained the formation of ordinary ether as a decomposition; but they deviated from this common understanding inasmuch as they denied that it was an oxidation.

Around 1800, most French and German chemists substituted Fourcroy and Vauquelin's theory of the formation of ordinary ether for the older oxidation theory. Less agreement existed with respect to another question concerning ethers. In his chemical textbook, Fourcroy had presented detailed descriptions not only of the preparation and properties of ordinary ether, made with sulfuric acid, but also of similar operations performed with nitric acid and muriatic acid. Although he had painstakingly described the properties of the resulting ethers without ignoring their deviations from the properties of ordinary ethers, he asserted that the reactions of nitric acid and muriatic acid with alcohol did not differ significantly from the reaction of sulfuric acid with alcohol. "Ether is in itself an identical body," he summarized his credo; "[it is] a constant product of the decomposition of alcohol whatever the acid or the reagent may be that has formed it" (Fourcroy 1801–1802, 8: 175).

Fourcroy's assumption about the identity of ethers prepared with different kinds of acids was the main incentive for Thenard's repetition of the experiments in 1806. In August 1806, Thenard presented to the Paris Institute, the former Academy of Sciences, his first experimental results.[6] "The history of nitric, muriatic, acetic and phosphoric ethers is far from being illuminated," he declared right at the beginning of his presentation (Thenard 1807a, 74). Thenard's use of the term "history" in this experimental context is a curious historical fact that requires a brief explanation. Eighteenth-century and early nineteenth-century chemists indeed often used the term "history" in connection with descriptions of experiments, and most of the previous experiments with ethers fell under this category. "History" here meant the collection of facts about the ways of preparing substances as well as about their multiple perceptible properties and practical uses. Factual knowledge about

substances could be obtained in an experimental way, especially with respect to the chemical properties of substances and the methods of their production, or by observing practitioners' methods of production and practical uses of substances. As ethers were also produced by apothecaries and sold as remedies, it was, in particular, apothecaries' experience that contributed to their history. But other histories of substances, too, integrated the factual knowledge of both academic experimenters and artisanal practitioners. In so doing they continued a tradition of "experimental and natural history" that had been established during the seventeenth century, and whose most prominent spokesman had been Francis Bacon.[7]

Experimental history differed in many ways from its counterpart, which is much better known by historians and philosophers of science, namely, experimental philosophy. The literary style of experimental history appears much less methodical than the latter. But it did not, in fact, entirely lack organization. Comparable to natural history, the experimental history of chemical substances required their identification and classification.[8] This was also true for the ethers. Based on the observation that all ethers obtained from different kinds of acids were inflammable, volatile, and possessed an aromatic, sweet odor, Fourcroy had proposed in 1801 that the otherwise observed differences in the properties of ethers were unimportant for their identification and classification. Accordingly, he stated that the ethers made from different acids were only varieties of one and the same chemical species of ether. Thenard, however, questioned this classification, and his first series of experiments, which studied the ether made with nitric acid, fully convinced him that Fourcroy was mistaken. The repetition of the production of nitric ether and the careful examination of the various reaction products showed that pure nitric ether had not yet been isolated. That is, not even the question of the individuation of the ether had been solved. Furthermore, Thenard, who was a member of the *Société d'Arcueil*, was one of the first chemists to further develop the technique of quantitative analysis of organic substances. These analytical techniques were based on earlier experiments by Antoine L. Lavoisier as well as the Lavoisierian theory that organic compounds were composed mostly of carbon, hydrogen, and oxygen. In accordance with his analytical agenda, Thenard aimed to ground the identification of organic substances in knowledge of chemical composition. His further experiments with nitric ether and the quantitative analysis of this substance again showed that he was on the right track. They proved that nitric ether contained nitrogen and hence differed from the ether made with sulfuric acid, which contained only carbon, hydrogen and oxygen. Furthermore, from the kind of reaction products he concluded that the formation reaction of nitric ether must be different from that of vitriolic ether. Instead he assumed that an oxidation was taking place that decomposed the alcohol and thereby created the ether along with a large number of other products of decomposition.

However when Thenard went on to study the ether prepared with muriatic acid, he found an astonishing difference from the two kinds of ethers he had studied before. His unexpected results spurred him to repeat the experiment in order study the formation reaction of muriatic ether in finer, more quantitative detail. In the following I will show, first, how his repetitions and quantitative refinements of

experiments led to the stabilization and acceptance of the anomalous experimental result as an experimental fact. Second, I will show that Thenard's attempts to explain this anomalous fact along with the invisible processes in the formation reaction of muratic ether did not lead to an unambiguous result. Nevertheless despite this failure, they engendered new views on organic substances more general.

An Anomalous Experimental Result

The ether of muriatic acid was a special case in many respects. Although chemists had studied the chemical interaction between alcohol and muriatic acid for decades, around 1806 it was still contested that it was possible to produce an ether from these two ingredients. Hence Thenard's experiments first had to examine the latter problem. In his experimental report he went back to earlier attempts to produce muriatic ether by Guillaume François Rouelle, Pierre Joseph Macquer, and Antoine Baumé, which had yielded ambiguous results. But based on his own trials, he argued that the obstacle to the production of muriatic ether was merely of technical nature (Thenard 1807b). In consisted, for example, in the way of manipulating fire during the distillation; another technical problem was that the muriatic ether was a gas, which could easily escape the experimenter's attention. By means of his own procedure, which he described in fine detail, Thenard succeeded in isolating the gaseous ether. He also described in detail the procedure of its identification by measuring its physical properties and testing its chemical properties. In the test of the chemical properties of the ether he found that it was inflammable, and that its combustion yielded a large portion of muriatic acid; he detected the latter substance by its smell and the white precipitate (silver chloride) that was produced when it was introduced into a solution of silver nitrate. Based on this latter observation, the question arose of whether the muriatic acid was a chemical component of the muriatic ether. However, when Thenard set out to analyze the muriatic ether by means of the usual reagents, he obtained negative results. He could not prove by the usual analytical means that the muratic ether contained the muriatic acid as a chemical component. The solution of litmus did not turn red, which would have happened had the ether been mixed with muriatic acid. Furthermore strong alkalis such as potash and soda, which decomposed the known compounds of muriatic acid, did not have any effect on the ether, neither in its gaseous form nor when dissolved in water. Similar results were obtained with another reagent, a solution of silver nitrate. "Silver nitrate is a very sensitive reagent," Thenard's teacher Fourcroy had written in 1801, "for indicating and recognizing in mineral waters and in all liquids the presence and the quantity of muriats or muriatic acid that are dissolved therein" (Fourcroy 1801–1802, 6:335). But in the case of muriatic ether the test with silver nitrate failed, as did all of the other standard tests using reagents for detecting muriatic acid in the ether.

How was it possible that none of the usual reagents separated the muriatic acid from the muriatic ether, thus indicating its presence as a chemical component in the ether, and that nonetheless muriatic acid was released in large quantities by the ether's combustion? Thenard explicitly raised this vexing question (Thenard 1807b, 123), which

would have been a conundrum for any chemist of the time. Muriatic ether, he pronounced, is one "of the most singular compounds that can be created" (ibid.). Thenard left no doubt that this result was more than a single technical problem. It concerned established analytical methods, underpinned by the laws of affinity. With respect to the latter, he emphasized in particular the fact that fixed alkalis did not decompose muriatic ether. Fixed alkalis had the strongest affinity with acids hitherto known, and hence would replace any other substance from its combination with muriatic acid.

Thenard's experiments had led to an experimental anomaly, that is, a discrepancy between his expectations, which relied on collectively shared analytical methods and knowledge, and the actual experimental results. He tackled the problem by repeating the experiment in a quantitative way. The first question that had to be solved was whether the muriatic acid created in the combustion of the ether was actually released from the ether or, alternatively, stemmed from a different source (ibid., 123). That is, Thenard did not immediately accept the anomalous result as an experimental fact, but rather looked for possible mistakes in either the experimental procedure or the interpretation of the experiment. In a repetition of the preparation of muriatic ether he first measured how much of the muriatic acid used as an ingredient for preparing the muratic ether was consumed in the reaction. He found that from 450.937 g muriatic acid used as an ingredient, 122.288 g had been consumed (ibid., 126). Hence, Thenard concluded, the muriatic ether in fact contained the muriatic acid, or the elements of it, and the muratic acid observed in the ether's combustion stemmed, at least in part, from the muriatic ether, rather than any other source.[9]

But Thenard remained skeptical. In order to examine the question of whether *all* of the components of the muriatic acid set free in the combustion of muriatic ether were indeed released by this ether, he performed a quantitative combustion analysis of the ether in a glass tube under the careful exclusion of air. When he compared the volume of the muriatic acid consumed in the formation of the ether with the volume of the muriatic acid set free in this quantitative combustion analysis, he observed that 900 g muriatic acid consumed in the formation of the ether approximately equaled 896 g muriatic acid produced in its combustion analysis. This result demonstrated convincingly that the muriatic acid released in the combustion of muriatic ether stemmed both qualitatively and quantitatively from the ether. Based on this result, it was also clear that the failure of Thenard's original attempts to prove by means of chemical reagents that muriatic acid was a component of muriatic ether was indeed an anomaly. Shortly afterwards Thenard repeated the experiments in the presence of members of the Institute. Neither the academicians nor Thenard himself had any doubt about that the anomalous experimental result was a fact – they called it a "surprise" – which had far-reaching consequences:

> When, on February 18, I gave a lecture on the ether of muriatic acid at the Institute, all members of the Institute, Messrs. Berthollet, Chaptal, Deyeux, Fourcroy, Guyton, Vauquelin, Gay-Lussac, etc. considered the results to be very new; they were struck by the consequences which can be drawn. Mr. Proust, who is currently staying at Paris and for whom, in accordance with his wish, I repeated my experiments previously done at the Institute … completely shared the surprise, as well as the opinion of the French chemists.
>
> (Thenard 1807c, 135)

Why were the members of the Institute "struck" by the consequences of the experiment? What were these consequences? There were, first, consequences, mentioned above, that concerned the analytical methods and the laws of affinity. But there were other consequences too, concerning the existing views about organic substances. These latter consequences were connected with another experimental result, which I have not yet described so far.

I mentioned above that the production of the ethers of sulfuric acid and nitric acid yielded many different reaction products along with the ethers. Based on this observation, all chemists who had studied the reactions, including Thenard, interpreted them as complicated decompositions of the alcohol, differing only in their views about what kind of decomposition took place and whether the reacting acid was decomposed as well. This common understanding of etherification accorded with chemists' more general view that organic substances, as a rule, were very instable materials that decomposed easily into many products when mineral acids or other strong reagents were added. Organic substances even denaturated spontaneously when they were isolated from their natural site of origin, the body of plants or animals, yielding cascades of decomposition products.[10] Yet the experiment with alcohol and muriatic acid yielded only two reaction products, the gaseous ether and water. Contrary to the typical decompositions of organic compounds that were difficult to control and interpret, this reaction seemed to resemble the formation of an inorganic salt from a base and an acid. In the formation of inorganic salts, the base united with the acid, and simultaneously water was released. This kind of synthesis was one of the oldest, most familiar and simplest reactions in inorganic chemistry, but which chemists deemed impossible in organic chemistry. If the mechanism of the formation reaction of muriatic ether was indeed fully identical with the synthesis of inorganic salts, there was a clear consequence: the muriatic ether would contain alcohol and muriatic acid as integral components. This was a bold hypothesis at a time when chemists agreed that organic and inorganic compounds were utterly different kinds of substances. And one fact clearly spoke against this hypothesis, namely, the failure to detect the muriatic acid in the muriatic ether by means of chemical reagents.

As a possible solution to this problem, Thenard raised an important question. Was it possible that muriatic acid was not contained as a preserved component in the ether but decomposed into its elements when combining with the alcohol? In the Lavoisierian chemical system, muriatic acid, like other acids, was defined as a compound that contained oxygen and another hitherto unknown element or "radical." If muriatic ether contained the oxygen and the radical of the initial muriatic acid, rather than the undecomposed muriatic acid, the failure to trace the muriatic acid in the ether would have been "explained naturally" (Thenard 1807b, 133). Hence, Thenard proposed first that "from all of the hypotheses made so far only one can be accepted, namely that one which proposed that all of *the elements* of muriatic acid do exist in the ether and are *combined with the elements* of alcohol in the same way as *the elements of* water, carbonic acid, and ammonia, etc. exist in plant and animal materials" (ibid., 131f., my emphasis).

However, Thenard immediately questioned this hypothesis, since it created a new problem. This new problem concerned the collectively shared knowledge about the properties and chemical behavior of muriatic acid. As Thenard's teacher Fourcroy had observed in his chemical textbook, chemists had not yet succeeded in decomposing muriatic acid and examining its components (Fourcroy 1801–1802, 2:101). Given chemists' numerous earlier attempts to decompose muriatic acid into its elements and the failure of these attempts even with strong reagents, it was not very probable that a weak reagent, such as alcohol, which interacted with muriatic acid only under very specific experimental conditions, was able to decompose it into its elements. Based on this second consideration Thenard then came back to his first hypothesis that muriatic acid might be preserved as a whole in the muriatic ether, concluding as follows:

> If it is possible to prove that muriatic acid preexists as a whole in the gaseous ether we have created a compound *whose existence could not be predicted by theory.*
>
> (Thenard 1807b, 133, my emphasis)

Thenard's terminology reveals that he conceived his experimental results, and the hypothesis that could be concluded from it, as a deep conceptual problem. The "theory" at stake here concerned three issues: the apparent similarity of the formation of the muritatic ether to the formation of inorganic salts, along with the similarity of the constitution of muratic ether to the binary constitution of inorganic salts; chemists' analytical standards; and the laws of affinity, which underpinned chemists' analytical standards. The following excursus serves to clarify the discordance that Thenard perceived between the existing chemical knowledge and the results of his experiments.

Some Conceptual Preconditions: "Organic" Versus "Inorganic" Matter, and the Laws of Affinity

In the late eighteenth and early nineteenth centuries, chemists considered organic (or "plant and animal") and inorganic (or "mineral," "fossil") substances to be utterly different kinds of matter. "We conclude from the facts we have collected," Fourcroy wrote in 1801, "that vegetable organization must form compounds that are extremely different from those that constitute fossils, and that the chemical phenomena that the [former] compounds present must be entirely different from the results obtained by treating mineral substances" (Fourcroy 1801–1802, 7:36). When they were heated, organic substances, as a rule, decomposed into many different gaseous, liquid and solid substances. Chemists could further observe daily in their laboratories that even at normal temperatures organic substances underwent spontaneous transformations, such as fermentation and putrefaction. In 1804, the British chemist Thomas Thomson gave voice to the common view that spontaneous decompositions of organic substances constituted a profound and characteristic disparity between organic and inorganic substances:

> The most striking distinction between the substances belonging to the mineral kingdom and those which make a part of animals and vegetables, is the following: Mineral bodies show little or no tendency to change their nature; and when left to themselves, undergo no spontaneous decomposition; whereas animal and plant substances are continually altering; and when left to themselves in favourable circumstances, always run through a regular set of decomposition.
>
> (Thomson 1804, 4:442)

The interaction of organic substances with reagents, such as strong alkalis and acids, further corroborated chemists' belief that organic and inorganic substances belonged to worlds apart. In Fourcroy's words:

> Without any doubt, one must be ready to find great differences between organic materials and mineral or fossil material. In general, the first, which are much more complicated in their composition, are also much more subject to change through the action of these bodies [the reagents], and the changes which appear are thus far more diverse and far more difficult to grasp and explain than those which had been mentioned in the history of all the other [mineral] substances.
>
> (Fourcroy 1801–1802, 7:61f.)

As the decompositions of organic substances yielded many different reaction products, their complexity resisted analysis. But Lavoisierian chemists emphasized yet another difference between organic and inorganic substances that concerned the concepts of "chemical affinity" and "binary constitution." In the inorganic domain of chemistry, which had been at the center of the Chemical Revolution in the last third of the eighteenth century, chemists commonly held the view that chemical compounds had a binary constitution. The concept of "binary composition" meant that the chemical elements of a compound always grouped together to form two proximate components of a compound. Whenever an inorganic compound contained more than two different elements, the various elements would combine in such a way with one another that two proximate components of the respective chemical compound emerged.[11] The paradigmatic example for binary constitution was the constitution of salts. Since salts could be experimentally produced from an acid and a base, and since they could be decomposed again into these two ingredients, the acid and the base were considered relatively stable building blocks or the two proximate components of salts; yet in subsequent analyses both acids and bases could be further decomposed into chemical elements. The concept of binary constitution was closely related to the "laws" of affinity. Laws of affinity were represented in the eighteenth-century chemical affinity tables, the first of which was published by Etienne François Geoffroy in 1718 (Geoffroy 1718). As can be seen in the chemical affinity tables, chemists represented chemical affinity as a relationship between two different substances, and the laws of affinity, displayed by affinity tables, were regular relationships between two substances. Hence, chemical compounds had a binary constitution.

In 1787 the concept of binary constitution of chemical compounds became implemented in the new chemical nomenclature and system of classification of Lavoisierian chemistry, without being extended, however, to organic substances

(Guyton de Morveau et al. 1787). In his famous *Traité*, Lavoisier stated that the vegetable substances always consisted of three components, namely the elements charcoal, hydrogen and oxygen (Lavoisier 1965, 123). Considering the composition of vegetable acids, he explicitly rejected the view that they might be composed not directly from the elements but from water and carbonic acid, or an oil: "Though all these acids, as has been already said, are chiefly, and almost entirely, composed of hydrogen, charcoal, and oxygen, yet, properly speaking, they contain neither water, carbonic acid nor oil, but only the elements necessary for forming these substances" (ibid., 120).[12] In the same vein, Fourcroy stated in his chemical textbook that vegetable substances were ternary compounds consisting of carbon, hydrogen and oxygen (Fourcroy 1801–1802, 7:53). This view had consequences for the understanding of chemical affinities. Lavoisier stated clearly that in the case of organic substances all three or four elements constituting these substances had a mutual relation of chemical affinity. He further assumed that this relation was in a state of equilibrium only at relatively low temperature, such as the ordinary temperature of the atmosphere and of the living body (Lavoisier 1965, 120ff.). Hence the fact that organic substances decomposed easily was explained by their fragile equilibrium of chemical affinities. Around 1800, most chemists believed that organic substances required the permanent influence of living bodies to maintain the equilibrium of affinities between their elements. Many of them further tackled the question of whether the force that bound together the components of organic compounds was the same as that in inorganic compounds. Among them quite a few postulated a specific force of life that was different in kind from chemical affinity. Others, like Jöns Jacob Berzelius, postulated the existence of a modified form of chemical or electrochemical affinity within the organs of living beings. As this force was absent under the conditions of the laboratory, extracted plant and animal substances decayed easily. Putting it in the words of Berzelius: "From the peculiar and special modification of the electro-chemical properties of these elements, organic bodies in general constitute but feeble compounds, which often begin to undergo decomposition as soon as they escape from the influence of the organ in which they were produced. Almost all organic bodies are decomposed by the united influence of air, water, and heat. Their elements resume their ordinary electro-chemical modifications, and there finally results a number of binary or inorganic combinations" (Berzelius 1814, 329).

Groping in the Dark

Thenard had started his experiments on ethers by tackling a question that belonged to the "history" of ethers. He had successfully demonstrated that his teachers' classification of the varieties of ethers as one chemical species was erroneous, and that the composition of nitric ether and muriatic ether differed from that of ordinary sulfuric ether. When he began to experiment on the formation of muriatic ether, he focused on experimental-historical questions as well as on technical problems of the operation. Yet the more he studied the formation of muriatic ether, its chemical properties and its

decomposition, the more he became entangled in conceptual problems. Thenard could not rest until these problems were resolved. For a couple of months, his experiments were driven primarily by conceptual issues. In a new publication, he mentioned that 3 months long he had been working for several hours daily studying the formation reaction of muriatic ether. He had repeated his earlier quantitative experiments on the consumption and release of muriatic acid, which confirmed his original results. The next problem he wanted to solve was the question of how alcohol behaved in the formation reaction of muriatic ether. Did all of its elements and the whole quantity of them unite with muriatic acid? Or was alcohol decomposed so that only one of its components became part of muriatic ether (Thenard 1807e, 344)? In this context, the origin of the second reaction product, water, was at stake. Either the water was synthesized in the formation reaction of muriatic ether – with the consequence that the alcohol was decomposed to some extent – or it was preexisting water that had been mixed but not chemically combined with alcohol, with the consequence that the alcohol was preserved in the formation reaction of muriatic ether. Again, Thenard was looking for a quantitative method to decide between the alternatives.

A month before, the Geneva chemist Nicolas Theodore de Saussure had published results of the first precise quantitative elemental analysis of alcohol (Saussure 1807).[13] Thenard drew upon these results, and additionally performed a series of quantitative analyses of muriatic ether aiming at comparing the quantitative composition of alcohol with that of muriatic ether. If muriatic ether contained the elements carbon, hydrogen and oxygen in the same proportions as alcohol, it was probable that alcohol combined with muriatic acid without losing water from its composition. As was usual at this time, the quantitative analysis of the ether turned out to be technically difficult. The results of the analysis and the subsequent comparison of the quantitative composition of the alcohol and the ether did not allow unambiguous conclusions. Nevertheless Thenard summarized: “considering everything, I am inclined to believe that alcohol, or its disunited elements, are part of the composition of muriatic ether” (Thenard 1807e, 346).

Based on the assumption that alcohol did not release water in the formation reaction of muriatic ether, Thenard then reconsidered the question of the interaction of alcohol and muriatic acid in the formation reaction of this ether. If neither alcohol nor muriatic acid was decomposed in this reaction the decisive, but “much more difficult” question again was:

> … in what manner the elements [of alcohol and muriatic acid] are combined in muriatic ether. Are hydrogen, oxygen and carbon disunited or united therein? Supposing they are [contained] in proportions necessary to form alcohol, are they in the state of alcohol? And supposing muriatic acid is a compound, is it contained as a whole [compound] or is it decomposed [into its elements]?.
>
> (ibid., 346)

Although Thenard had raised this question before, he was still groping in the dark. He saw the only way out in the repetition of his former attempts to decompose muriatic ether by means of different reagents and to trace the products of decomposition. For the first time he also took into account the time needed for the ether’s decomposition

as well as temperature. The results of his total of eighteen experiments were the following. Potash, ammoniac, silver nitrate and mercury nitrate did not immediately trace the muriatic acid in the ether. However, after a few days, or even months, the tests were positive. Thenard then tried to decompose muriatic ether by means of sulfuric acid, nitric acid, and nitrous acid. According to the laws of affinity, the affinities between these three mineral acids and salifiable bases were stronger than the affinity between the muriatic acid and salifiable bases; consequently they should displace muriatic acid from its union with a base. When Thenard performed the experiments at ordinary temperatures, muriatic acid was not released. But when he raised the temperature, a large quantity of muriatic acid was set free from the ether. Using oxygenated muriatic acid (later chlorine) as a reagent, Thenard even succeeded in extracting a large quantity of muriatic acid from the ether at normal temperature.[14]

In his long discussion of these results, Thenard reminded his readers of his two hypotheses. Either muriatic ether was a binary compound consisting of the two proximate components muriatic acid and alcohol, or it consisted of the disunited elements of these two compounds. Thenard carefully pondered the pros and cons of this alternative. Contemplating the consequences if alcohol and muriatic acid actually did preserve their integrity while recombining into the ether, he stated:

> … if the muriatic ether were a compound consisting of muriatic acid and alcohol, as some people do not fear to claim, it seems that *these two bodies had to unite like the acids and the alkalis*; and consequently had to neutralize themselves *as soon as they were in contact*, for they had to be viewed as having *a much stronger mutual affinity* than muriatic acid and potash and most of the other salifiable bases. Nevertheless, one knows that *they combine only with difficulty*, neutralize each other only little by little and by means of a light heat. … Presupposed it consisted of alcohol and muriatic acid, the decomposition of the ether had to take place immediately as the contact [between the two ingredients] is immediate and the same in each moment; *yet, it takes place only after a long time*.
>
> (ibid., 356f., my emphasis)

Thenard alluded to some "other people" who did not "fear" to claim that muriatic ether resembled inorganic salts and had a binary constitution. By this he presumably meant the pharmacist Pierre F. G. Boullay, who shortly before, on May 25, 1807, had given a lecture at the Institute on the very same issue – muriatic ether and its constitution – as well as on the ether of acetic acid.[15] In his lecture, Boullay had concluded unambiguously, based on very similar experiments as Thenard's, that muriatic ether had a binary constitution containing muriatic acid and alcohol as integral building blocks.[16] He had proposed that "the mode of etherification is the same for all different volatile acids, and that the ethers that result from their action are true compounds in the manner of salts, in which alcohol functions as a base" (Boullay 1807, 100). Furthermore, he had suggested dividing the known ethers into two classes: one containing the ethers produced with sulfuric acid and phosphoric acid, the other containing ethers produced with volatile acids, such as muriatic acid and acetic acid. The first class contained ethers resulting from the decomposition of alcohol by withdrawing water, whereas products of the synthesis of the alcohol and the acid were in the second class (ibid., 100f.).

Thenard, however, remained skeptical. He considered the time needed for the union of alcohol and muriatic ether, as well as for the decomposition of the ether, to be a very problematic issue that spoke against the assumption that there was a strong affinity between alcohol and muriatic acid. When E. F. Geoffroy designed his first affinity table in 1718, he took it for granted that the laws of affinity between two substances were generally valid. Later in the century, in particular Torbern Bergman argued against this view by considering the role played by temperature. In his table of affinities he introduced two classes of affinities, one displayed at very high temperatures, when operations were performed in the "dry way," and one displayed at lower temperature, when operations were performed in solutions or the "wet" way.[17] However, apart from Claude Louis Berthollet (Berthollet 1803), no chemist had taken into account theoretically the time needed for the formation of a compound, and for its decomposition. The question of how affinities related to time was still a very new one, in particular with respect to organic substances, and there was no ready-made theoretical clue to solve it. In the eyes of Thenard the results were so confusing that in the next step he decidedly rejected the hypothesis of a binary constitution of muriatic ether.[18] Instead he presented a bold atomic speculation about the reaction at stake, based on time. As alcohol needed time to be decomposed into its elements before it was able to combine with muriatic acid, he asserted, the formation reaction of muriatic ether was slowed down. For a similar reason, the decomposition of muriatic ether with reagents also required much time: "another arrangement of the molecules than the actual one needed to be achieved" (ibid., 358). Shortly afterwards Thenard published another experimental report on nitric ether (Thenard 1807f). Her had decomposed nitric ether with potash obtaining, after 36 h, alcohol, nitric acid, nitrous acid and acetic acid as decomposition products (ibid., 359). In a few steps he summarized his experimental results. He ignored nitrous acid, "supposing that one can extract from nitric ether no other materials than nitric acid, acetic acid, and alcohol" (ibid., 361). Then he created a link to his earlier experiments on the formation and decomposition of acetic ether (Thenard 1807d). In these earlier experiments he had shown that "acetic ether resulted from [the union of] all of the principles of acetic acid with all of those of alcohol without the formation of water or any other compound in the mutual reaction" (ibid., 157), and that "one easily transforms acetic ether into alcohol and acetic acid" (ibid.). Summing up these recent results, Thenard stated that "the ethers of muriatic, nitric and acetic acid are formed from these acids and alcohol, or the principles of these acids and alcohol" (Thenard 1807f., 361).

After many experiments, Thenard felt unable to solve the question concerning the similarity of the etherification to the formation of inorganic salts.[19] But at a time when the vast majority of chemists strictly separated organic from inorganic substances, he not only questioned this dichotomy but also presented experimental results and arguments that supported the opposite view. Together with Pierre F. G. Boullay he opened an avenue to a new understanding of organic substances, which remained on the collective agenda of French chemists in the decades to follow. This new understanding became even more visible in Thenard's subsequent experiments, which were no longer concerned with the vexing question of the binary constitution

of ethers and the similarity of etherifications to formations of inorganic salts, but rather examined more broadly the similarity of the chemical behavior of organic to inorganic substances. All of these new experiments focused on the questions of whether organic substances underwent chemical transformations other than decomposition and instead often united with reagents to form new chemical compounds. In order to answer these questions, Thenard not only performed a large number of experiments, but collected facts from many different areas. In so doing, his style of experiments again shifted to experimental history, a style which had stood at the beginning of his series of experiments on ethers.

Back to Experimental History

Thenard started his new experiments, which he reported to the Institute in November 1807, with a study of etherification by means of vegetable acids (Thenard 1809a). In the months before, when he had tried to solve a conceptual conundrum, he had presented lengthy reports on details of quantitative experiments and on various possibilities of hypotheses and interpretations. By contrast, he now switched to non-quantitative experiments and comparatively short reports in the style of experimental histories. A direct incentive for Thenard's new experiments on the ethers of vegetable acids was the series of experiments performed by the Swedish chemist-apothecary Carl Wilhelm Scheele in the 1780s.[20] Thenard first discussed the results of Scheele's experiments on the preparation of the ethers of acetic, benzoic, tartaric, citric and succinic acid. These experiments had been successful only in the cases of acetic and benzoic acid, and only under the precondition that a mineral acid was added. He further mentioned that Scheele had been able to decompose these ethers by means of potash, and that he had concluded from his experiments that these ethers contained the respective acids. But Scheele had not examined in fine detail the composition of the two ethers. In his own experiments, Thenard used as ingredients very pure benzoic, oxalic, citric, and malic acid as well as very pure alcohol. He then repeated all of these experiments, but this time adding a mineral acid to the mixture of alcohol and the vegetable acid. As a result, he found that all of the four vegetable acids yielded ethers, but only under the condition that a mineral acid was added. In a further series of experiments, Thenard succeeded in decomposing these ethers by distilling them with a solution of potash. As decomposition products he obtained alcohol and the vegetable acid, but no mineral acid. From these results he concluded: "There we are, these are new compounds of a vegetable acid with an alcohol" (ibid., 12). Assuming that "all vegetable acids behave with alcohol in the same way as the preceding acids" (ibid., 13), he further extended his series of experiments to study the formation of ethers with gallic and tartaric acid, and the subsequent decomposition of these ethers. As these experiments confirmed his expectation, he went on to perform similar experiments with animal acids.

Almost 3 months later, in February 1808, Thenard reported to the members of the Institute on the continuation of his experiments in an even more summerizing

style (Thenard 1809b). The next question to be examined was whether organic substances other than alcohol were able to combine with acids. In order to answer this question, Thenard gathered experimental facts by repeating several experiments that had previously been performed by himself or by other European chemists. His first experiment studied an oil produced from a mixture of alcohol and "oxidated muriatic acid" (later chlorine). A year before he had already performed this experiment, and concluded that the oil was a product of decomposition of alcohol by means of the acid (Thenard 1807d). But in light of his new experience he assumed that the oil might be a product of the union of oxidated muriatic acid (or, alternatively, of muriatic acid) and a yet unknown plant substance; he further assumed that this oil was fully analogous in composition to the muriatic and acetic ethers. His growing confidence in that analogy subdued any doubts arising from the presence of additional reaction products. Further experiments performed to decompose the oil with an alkali into muriatic acid and the plant substance, however, yielded ambiguous results. Even with strong alkalis only very small quantities of muriatic acid and of oxidated muriatic acid were produced, and the decomposition of the oil took a long time. Nevertheless Thenard now stated that the oil "contained" muriatic acid, and that the component muriatic acid is "intimately combined with another substance," still to be identified (Thenard 1809b, 26). Despite his failure to identify the second component of the oil, he also expressed his conviction that the unknown plant substance "neutralizes the acid in the manner of alkalis" (ibid.). Given Thenard's earlier reluctance toward premature conclusions from his experiments, this conclusion demonstrates strikingly how much his views about organic substances and their difference from inorganic ones had changed within the short time span of a couple of months.

The next substance Thenard studied was "artificial camphor," produced from a vegetable essential oil, namely, the oil distilled from turpentine, and muriatic acid. Thenard shared the belief of many chemists that "artificial camphor" was identical with the imported and expensive natural camphor, which was used as a remedy. Therefore this substance, which at the time was already fabricated and sold in large quantities (Cuvier 1827, 117), was also of peculiar technological interest to him. But he now doubted that the artificial camphor was a decomposition product of the essential oil of turpentine, as most chemists believed. In order to clarify this issue, he repeated its production following the recipes of H. H. Kind, Johann Trommsdorff and Adolf F. Gehlen. In the subsequent combustion analysis of his artifical camphor, he obtained muriatic acid (Thenard 1809b, 30). Based on this result, Thenard proposed that artifical camphor was the product of a simple union of the essential oil of turpentine with muriatic acid (ibid., 31). Against this interpretation stood the fact that in the preparation of artificial camphor a second, liquid reaction product was created that Thenard could not identify. The argument he nevertheless advanced in favor of his re-interpretation again gives testimony of his reinforced belief in the analogy between the reaction of inorganic, salifiable bases and acids, and organic "bases" and acids. "What brings me to believe that muriatic acid does not decompose the essence [essential oil of turpentine], and that on the contrary it combines with it, is [the fact] that this same acid evidently combines with all principles of

alcohol without creating additional products, and [the fact] that alcohol is a hydrogenated body just like the essential oil of turpentine" (ibid., 31). Here Thenard's explanation of the reaction relied almost entirely on reasoning by analogy.

Thenard went on to study further reactions between muriatic acid and other essential oils, namely the "essence of lemon" and the "essence of lavender," which were precious commodities used in the apothecary trade and in perfumeries. As a variety of the imported camphor could also be obtained from the essential oil of lavender, Thenard hinted at the possible practical, commercial consequences of his research. Perhaps, he proclaimed, the various kinds of natural camphor were also compounds consisting of an essential oil and an acid, and from this "great advantages" could result (ibid., 52). Without giving any details of his new experiments with essence of lemon and of lavender, Thenard reported the result that muriatic acid combined with these two kinds of essential oils, too.

The rest of Thenard's report is an excellent example for an experimental history – almost in the original, seventeenth-century Baconian sense of a collection of all kinds of facts from various different areas, including the arts and crafts and everyday life. Thenard first extended the class of base-like plant materials by adding the fatty oils, reminding his readers that it is well known that fatty oils yielded soaps not only together with alkalis, as could be observed daily in workshops, but also with sulfuric acid. After that came another commodity, tannin. In the case of tannin, Thenard stated that it was well known ("*comme on le sait*") that tannin could be precipitated by means of concentrated sulfuric acid; he now interpreted this precipitate as a compound consisting of tannin and sulfuric acid (ibid., 33). Relating this experiment to the tannin contained in oak apples, he remined his readers that it was known that oak apples yielded an acid when boiled, despite the fact that they were a neutral material. From this he concluded that the acid (gallic acid) contained in oak apples was neutralized by another substance, which "without doubt" was tannin. Similar reports were then given on five kinds of animal substances, namely casein, albumin, picromel, gelatine and urea. For example, in the case of casein, Thenard stated that it was "generally known" that casein combined with acids, which became obvious, for example, when milk coagulated after its mixture with an acid.

Thenard concluded this last series of experiments as follows:

> Hence, there are five plant materials and five animal materials that can combine with the acids. Three among them … neutralize the acids as well as the strongest alkalis. The other seven form compounds with the acids, which are themselves acids, similar to the metal salts and several salts of earths.
>
> (ibid., 39)

In the end, Thenard was fully convinced that chemists' traditional view that acids always decompose organic substances was wrong. Many organic substances, he now asserted, combined with acids to form new compounds. Thenard did not hesitate to draw further general conclusions for the future of organic chemistry: "Without any doubt, we will be able in the future to combine all of the other plant and animal substances with the acids" (ibid., 39). In an almost emphatic tone he proclaimed that his experiments were both "useful and important" (ibid. 39), and that they would

shed new light also on the behavior of organic substances in nature, that is, the living bodies of plants and animals. His report ended with the assertion that his new "general principle" is "capable of a great number of applications" (ibid., 41).

Conclusion

Experimental history was a pervasive style of experimentation in eighteenth-century and early nineteenth-century chemistry. Questions tackled under the auspices of an experimental history were, for example: how do apothecaries manufacture ethers? What are the properties of ethers? Are the varieties of ethers made from different acids significant from a chemical point of view, that is, are there different chemical kinds of ether? In their experimental histories chemists relied on a web of chemical concepts, such as compound, composition, analysis, affinity, binary constitution, and the demarcation of organic and inorganic compounds. But they never questioned these concepts in the context of experimental-historical investigations. Rather, in this context they used these concepts as heuristic tools. In so doing they took them for granted in a similar way as instruments are taken for granted when they are applied as working tools in experiments. In the vein of the experimental-historical style of experimentation, Thenard was not concerned with questions of affinity, binary constitution, and the relationship between organic and inorganic compounds. Instead, he asked questions concerning the techniques of preparing ethers, and their identification and differences. Yet when he studied the ether produced with muriatic acid, he encountered an anomalous experimental result. His repetition and quantification of the experiments convinced him that he had not made a mistake but rather discovered a true experimental anomaly, which he could not explain. His attempts to further study this anomaly drove him deeper and deeper into conceptual problems. At the conclusion of his experiments, which lasted for almost 2 years, Thenard, like the pharmacist Pierre F. G. Boullay, called into question one of the strongest beliefs of the chemists at the time, namely, that organic compounds would, as a rule, be decomposed in chemical reactions.

In the decades that followed, many European chemists accepted Thenard's and Boullay's proposition that the ethers of volatile acids were not products of decomposition but rather of the union of alcohol and an acid. The specific experimental anomaly, however, and the conceptual problem of the binary constitution of ethers, was not solved unambiguously in Thenard's experiments. Unlike Boullay, Thenard remained skeptical that the formation reactions of the ethers of volatile acids was the same type of reaction as the neutralization of inorganic salts; hence, that these ethers corresponded with inorganic salts in their binary constitution. Nevertheless, his experiments, and the conclusions he drew from them, undermined the existing dichotomy between organic and inorganic substances. Some 10 years after Thenard's experiments, leading chemists and pharmacists, most of them French, became convinced that the concept of binary constitution must be extended to, at least, some organic compounds, namely alcohol, ordinary ether and several other

derivatives of alcohol. To these chemists belonged Joseph Louis Gay-Lussac (Gay-Lussac 1815), Thenard's collaborator at the Société d'Arceuil, the Geneva chemist Theodore de Saussure (Saussure 1814), the Paris chemist Jean-Baptiste Dumas, Thenard's most famous pupil, and the pharmacist Polydore Boullay, the son of Pierre F. G. Boullay.[21] Not until the 1830s, when Friedrich Wöhler, Justus Liebig, Jean-Baptiste Dumas and Polydore Boullay performed new experiments, along with new work on paper with chemical formulae, did a larger number of European chemists agree on the extension of the concept of binary constitution from inorganic to organic compounds. They did so in conjunction with changes in many other practices and concepts, all of which contributed to a new type of organic chemistry that emerged between the late 1820s and the 1840s. In the early period of the establishment of the new culture of organic or carbon chemistry, the concept of binary organic compounds, and the practice of modeling by means of chemical formulae the binary constitution of organic compounds, was a significant part of the transformation process. But in the 1850s, the concept of binary organic compounds was up for grabs again. In this later period of organic chemistry chemists began to adopt quite different views about the invisible constitution or "structure" of organic compounds. Chemists then felt that the concept of a binary constitution of organic had been a profound conceptual error – yet this error was a constitutive part in the coming into being of the new culture of carbon chemistry that eventually made possible the new concept of molecular structure.

Notes

1. Among the chemists who first studied the production of pure ether were Cromwell Mortimer, Pierre Joseph Macquer, Guillaume F. Rouelle, and Antoine Baumé.
2. In modern terminology these substances were not "ethers" but "esters" and other kinds of organic compounds.
3. See Fourcroy and Vauquelin 1797.
4. All translations are my own, except where stated.
5. But ether was not simply alcohol minus water, since a portion of carbon contained in the alcohol precipitated at the same time (as a consequence of the fact that it was no longer saturated with hydrogen and oxygen). Based on the estimation that the mass of the precipitated carbon was greater than the masses of the hydrogen and the oxygen withdrawn from alcohol, Fourcroy and Vauquelin concluded that alcohol contained more carbon than ordinary ether (or that ether contained comparatively more hydrogen and oxygen than alcohol), and hence that ordinary ether was "alcohol plus hydrogen and oxygen" (Fourcroy and Vauquelin 1797, 209f.).
6. On Thenard's experiments on ethers see also Crosland 1967, 1981, 373–378.
7. On the tradition of experimental history and the experimental-historical and technological context of Thenard's experiments, see Klein 2005a.
8. On the problem of identification and classification of organic substances at the time, see also Klein 2005b.
9. As chemists at the time did not yet agree about the composition of alcohol, Thenard did not rule out the possibility that alcohol contained the hitherto unknown "radical" of muriatic acid. Hence he concluded first that the radical of muriatic acid stemmed "in part" from the initial

muriatic acid. Furthermore, as chemists assumed that muriatic acid contained oxygen, the possibility existed that its oxygen component stemmed from air.

10. On this view see also Klein 2003.
11. In the following, I mention only the meaning of the concept of binary constitution in the context of chemists' experimental and taxonomic praxis dealing with substances. "Binarity" and the adjective "binary" had a different meaning in atomic theories like that of John Dalton. In that context, "binary" could also mean "made up of two atoms." On the meaning of the concept of binary constiution in eighteenth-century and early nineteenth-century inorganic chemistry, and its extension to organic chemistry, see Klein 2003, 149–187.
12. On Lavoisier's stance on the binary constitution of organic compounds, see also Holmes 1985, 405; Melhado 1981, 124f., 139f.
13. Thenard mentioned that he had begun his series of experiments on February 21 (1807), finishing them on May 19. De Saussure read his article at the Institute on April 6.
14. If muriatic acid was set free as vapor, Thenard could smell it; alternatively he precipitated it with potash or silver nitrate. Alcohol was more difficult to trace, since it could not be precipitated. When it was separated as a distillate, its smell was sometimes concealed by the smell of undecomposed ether.
15. Thenard mentioned in his paper that he had performed experiments for more than three months after reading his first paper on muriatic ether, which was on February 18. He presumably completed his second experimental report by the end of May, that is, after Boullay had read his own report. This assumption is supported by Thenard's allusion to another person who assumed a binary constitution of muriatic ether. On Boullay's experiments see Klein 2005a.
16. According to Boullay, alcohol and muriatic acid were "constitutive parts" or "constitutive principles" ("*partie constituante,*" ibid., 97; "*principes constituans*", Boullay 1807, 98) of the ether.
17. On the history of affinity tables and Bergman's modification, see the study by Duncan 1996.
18. On the problems of authority concerning Boullay that were involved here see Klein 2005a.
19. In hindsight, esters are comparable with salts. However, whereas salts can be decomposed quickly because of their ionic bonding, esters have atomic bonds.
20. On Scheele's experiments see Partington 1961–1970, 3:233.
21. See Dumas and Boullay 1827, 1828. On chemists' views about the binary constitution of organic compounds, see also Brooke 1992, 1987; Klein 2003, 149–187.

References

Berthollet, Claude Louis. 1803. *Essai de statique chimique*. Paris: Didot.

Berzelius, Jöns Jacob. 1814. Experiments to Determine the Definite Proportions in Which the Elements of Organic Nature are Combined. *Annals of Philosophy* 4:323–331, 401–409.

Boullay, Pierre François Guillaume. 1807. Mémoire sur le mode de composition des éthers muriatique et acétique. *Annales de Chimie* 63:90–101.

Brooke, John Hedley. 1987. Methods and Methodology in the Development of Organic Chemistry. *Ambix* 34 (3):147–155.

Brooke, John Hedley. 1992. Berzelius, the Dualistic Hypothesis, and the Rise of Organic Chemistry. In *Enlightenment Science in the Romantic Era*, edited by E. M. Melhado and T. Frängsmyr. Cambridge: Cambridge University Press.

Crosland, Maurice P. 1967. *The Society of Arcueil: A View of French Science at the Time of Napoleon I*. Cambridge, Mass.: Harvard University Press.

Crosland, Maurice P. 1981. Thenard, Louis Jacques. In *Dictionary of Scientific Biography*, edited by C. C. Gillispie. New York: Charles Scribner's Sons.

Cuvier, Georges. 1827. *Rapport Historique sur les Progrès des Sciences Naturelles depuis 1789, et sur leur État Naturel, présenté au gouvernement, le 6 février 1808, par la classe des sciences*

physiques et mathématiques de l'Institut, conformement a l'arrêté du 13 Ventôse an X. Paris: Firmin Didot.

Dumas, Jean-Baptiste André, and Polydore Boullay. 1827. Mémoire sur la Formation de l'Ether sulfurique. *Annales de Chimie et de Physique* 36:294–310.

Dumas, Jean-Baptiste André, and Polydore Boullay. 1828. Mémoire sur les Ethers composés. *Annales de Chimie et de Physique* 37:15–53.

Duncan, Alistair. 1996. *Laws and Order in Eighteenth-Century Chemistry*. Oxford: Clarendon Press.

Fourcroy, Antoine-François de. 1801–1802. *Systême des connaissances chimiques: et de leurs applications aux phénomènes de la nature et de l'art*. 11 vols. Paris: Baudouin.

Fourcroy, Antoine-François de, and Louis N. Vauquelin. 1797. De l'action de l'Acide sulfurique sur l'Alcool, et de la formation de l'Ether. *Annales de Chimie* 23:203–215.

Gay-Lussac, Joseph Louis. 1815. Lettre de M. Gay-Lussac à M. Clément, sur l'analyse de l'alcool et de l'éther sulfurique, et sur les produits de la fermentation. *Annales de Chimie* 95: 311–318.

Geoffroy, Etienne François. 1718. Table des differentes rapports observés en Chimie entre differentes substances. *Histoire de l'Académie Royale des Sciences. Avec des Mémoires de Mathématique & de Physique pour la même Année, Mémoires*: 202–212.

de Morveau, Guyton; Bernard, Louis; Lavoisier, Antoine Laurent; Berthollet, Claude Louis and Fourcroy, Antoine François. 1787. *Méthode de Nomenclature Chimique. On y a joint un nouveau systême de caractères chimiques, adaptés à cette nomenclature, par MM. Hassenfratz et Adet.* Paris: Cuchet.

Holmes, Frederic L. 1985. *Lavoisier and the Chemistry of Life: An Exploration of Scientific Creativity*. Madison: University of Wisconsin Press.

Klein, Ursula. 2003. *Experiments, Models, Paper Tools: Cultures of Organic Chemistry in the Nineteenth Century*. Edited by T. Lenoir and H. U. Gumbrecht, *Writing Science*. Stanford: Stanford University Press.

Klein, Ursula. 2005a. Experiments at the Intersection of Experimental History, Technological Inquiry and Conceptually Driven Analysis: A Case Study from Early Nineteenth-Century France. *Perspectives on Science* 13 (1): 1–48.

Klein, Ursula. 2005b. Shifting Ontologies, Changing Classification: Plant Materials from 1700 to 1830. *Studies in History and Philosophy of Science* 36: 261–329.

Lavoisier, Antoine Laurent. 1965. *Elements of Chemistry in a New Systematic Order, Containing all the Modern Discoveries.* Translated by Robert Kerr from the French. 1789; with a new introduction by Douglas MacKie. New York: Dover Publications.

Macquer, Pierre Joseph. 1766. *Dictionnaire de Chymie, contenant la Théorie et la Pratique de cette Science, son application à la Physique, à l'Histoire Naturelle, à la Médicine et à l'Economie animale*. 2 vols. Paris: Lacombe.

Melhado, Evan M. 1981. *Jacob Berzelius: The Emergence of his Chemical System*. Madison: University of Wisconsin Press.

Partington, James Riddick. 1961–1970. *A History of Chemistry*. 4 vols. New York: St. Martin's Press.

Saussure, Nicolas Theodore de. 1807. Mémoire: Sur la composition de l'Alcohol et de l'Éther sulfurique. *Journal de Physique, de Chimie, d'Histoire Naturelle et des Arts* 64: 316–354.

Saussure, Nicolas Theodore de. 1814. Nouvelles Observations sur la composition de l'alcool et de l'éther sulfurique. *Annales de Chimie* 89: 273–305.

Thenard, Louis Jacques. 1807a. Mémoire sur les Éthers. *Mémoires de Physique et de Chimie de la Société d'Arcueil* 1: 73–114.

Thenard, Louis Jacques. 1807b. Deuxième mémoire sur les éthers: De l'éther muriatique. *Mémoires de Physique et de Chimie de la Société d'Arcueil* 1: 115–135.

Thenard, Louis Jacques. 1807c. Sur la découverte de l'éther muriatique. *Mémoires de Physique et de Chimie de la Société d'Arcueil* 1: 135–139.

Thenard, Louis Jacques. 1807d. Troisième mémoire sur les éthers: Des produits qu'on obtient en traitant l'alcool par les muriates métalliques, l'acide muriatique oxigéné et l'acide acétique. *Mémoires de Physique et de Chimie de la Société d'Arcueil* 1: 140–160.

Thenard, Louis Jacques. 1807e. Deuxième mémoire sur l'éther muriatique. *Mémoires de Physique et de Chimie de la Société d'Arcueil* 1: 337–358.

Thenard, Louis Jacques. 1807f. Nouvelles observations sur l'éther nitrique. *Mémoires de Physique et de Chimie de la Société d'Arcueil* 1: 359–369.

Thenard, Louis Jacques. 1809a. De l'action des acides végétaux sur l'alcool, sans l'intermède et avec l'intermède des acides minéraux. *Mémoires de Physique et de Chimie de la Société d'Arcueil* 2: 5–22.

Thenard, Louis Jacques. 1809b. Sur la combinaison des acides avec les substances végétales et animales. *Mémoires de Physique et de Chimie de la Société d'Arcueil* 2: 23–41.

Thomson, Thomas. 1804. *A System of Chemistry*. 2 ed. 4 vols. Edinburgh: Bell & Bradfute, & E. Balfour.

Epilogue

There are diverse ways of going amiss. Not all of them can best be characterized in terms of error—experimental error is too limited a technical concept for this purpose. The episode of Krebs's discovery of the urea cycle is a case in point (chapter, "Concepts from the Bench: Krebs and the Urea Cycle"). In the course of his research, Krebs reached a dead end. All his experiments had worked fine, but he was unable to answer his research question with the material and conceptual tools that were available to him. Obviously, something was going amiss, but he was not in error. In a similar vein, the investigation of glycogenesis and the development of in vitro techniques for analyzing subcellular particles (chapter, "Experimental Reorientations") indicate that the concept of error is too closely associated with "proven wrong" to offer a suitable characterization for the openness and indeterminacy of the actual situation in experimental research.

Going amiss comprises two distinct but related themes. The first theme concerns the occurrence of disturbances and the encounter of impediments in everyday research. Of course, as several papers in this volume point out, going amiss is a common feature of scientific experimentation. Scientists frequently go amiss while attempting to get it right. Sometimes they are successful in this endeavor—they get it right; and sometimes they are not—the project fails. But precisely because going amiss is an integral part of scientists' everyday research, we need to consider its function in experimental practice. Therefore, a comprehensive analysis of the process of knowledge generation in experiment must take account of the common phenomena of going amiss. Directing our attention to scientists' encounter with errors, dead ends, disturbances and the like sheds new light on the process of knowledge generation.

The second theme is the experimental result, its epistemic status, and its career. The concept of error is pertinent to the evaluation of the status of knowledge claims generated in experiments. Such an inquiry need not necessarily reflect the actor's perspective. Building on both the actor's appraisal and the analyst's assessment of the experimental situation, it seeks to draw out general features of the occurrence of error. A typology of error may demonstrate how those claims that proved wrong can be analyzed in terms of sources of error (chapter, "Error: The Long Neglect, the One-Sided View, and a Typology"). It aims at systematic analysis of possible sources of errors in research procedures and thus may go beyond the actor's methodological concerns.

The episode of the globule hypothesis (chapter, "Error as Historiographical Challenge: The Infamous Globule Hypothesis") brings to the fore the distinction between the two themes. It is insightful to reconstruct the microscopists' negotiations among themselves as a debate about whether that hypothesis was erroneous, and we may in hindsight conclude that certain versions of that hypothesis were indeed erroneous. We may analyze the nature of this error in terms of its possible sources. However, it is awkward to render the investigation of the microscopists' research practice in terms of error. Confusion reigned and the practitioners began discussing whether the methods they used were correct, effective, or faulty. Unlike error, the notion of going amiss is conducive to this kind of analysis.

Analytical Perspectives

To study the phenomena of going amiss, it is crucial to distinguish carefully between different historiographical perspectives in which these phenomena can be approached. At least five domains of analysis need to be recognized. First, there is the domain of scientific practice. How do scientists deal with the possibility that something is amiss in their experiments? The study of the missing rain (chapter, "Distinguishing Real Results from Instrumental Artifacts: The Case of the Missing Rain") traces in detail how late eighteenth- and nineteenth-century investigators tried to solve the height-catch problem. It describes how the researchers sought to establish whether they were led astray by an instrumental artifact, or by the intervening physical conditions; or, alternatively, whether the observed difference in rainfall was a real difference. The issue of learning (chapter, "Learning Without Error") gives another twist to this perspective, exemplifying how scientists address the problem of error as a research topic in the study of processes of learning.

Secondly, there is the domain of scientists' discourses about their actual research practice. Do they discuss the problem of going amiss when they communicate their research? If so, how much attention is given to this discussion? What terms do they use to report it? And how do they evaluate their situation vis-à-vis the possibility of going amiss? The case of the globules shows that in the early nineteenth century microscopists commented extensively in their publications on this problem (chapter, "Error as Historiographical Challenge: The Infamous Globule Hypothesis"). The practice of reflecting on obstacles in experimental research has changed considerably since the early nineteenth century. Thus, analyzing how the textual forms of reporting errors and other impediments change over time would be a fruitful project for future research.

Thirdly and closely related, there is the domain of scientists' appraisal of scientific results. How do they evaluate experimental outcomes? When and why do they decide that a particular claim to knowledge is erroneous? Does this evaluation change over time, and if so, how and why? Fizeau's ether drift experiment elicited a variety of responses to its result by the physics community, ranging from approval and silence to opposition (chapter, "Going Right and Making It Wrong: The Reception of Fizeau's Ether-Drift Experiment of 1859"). The account of these responses traces how they changed from complete agreement to general rejection.

Fourthly, there is the domain of retrospective conceptual analysis. How can we characterize the nature of erroneous results? What does it mean for a result to be erroneous? As mentioned above, one approach is to identify possible sources of error in an experimental arrangement. The classification of sources of error yields a typology that is designed to capture the epistemic structure of the experimental argument. This approach is called "probing experiment with error"; in this scheme error serves as a tool of inquiry to throw light on the elements that comprise the process of generating experimental knowledge (chapter, "Error: The Long Neglect, the One-Sided View, and a Typology").

The nomenclature that we offer in the introduction to this volume serves both the scholars who seek to uncover the actors' usage of terms related to going amiss and the scholars who devise their own terms to shed light on the nature and role of impediments in experimentation. The contributions illustrate various elements of the proposed nomenclature and give numerous indications of how it could be refined, modified, and expanded. For example, in some situations, organic chemists actually referred to their experimental outcome in terms of "surprise" (chapter, "In the Thick of Organic Matter"). In contrast, the distinction between "waste" and "filth" (chapter, "The Scent of Filth: Experiments, Waste, and the Set-Up 12) is clearly not an actors' distinction. It is introduced to characterize an important difference between two kinds of unwanted or unintended effects. The first kind stands for those side effects that are not only a necessary by-product of experimental practice, but also help assure the experimenter that his or her apparatus is in working order. The second kind represents those effects that cannot be interpreted within the accepted theoretical framework. Effects of the second kind may or may not bring about novel empirical insights.

In contrast to the nomenclature offered in the Introduction, a typology of errors does not aim to capture the intricacies of scientific practice. Rather, it provides, as we have indicated, tools for the analysis of erroneous results in terms of their different sources. The classification of computational errors that von Neumann outlined in his studies of large scale computing machines is a case in point (chapter, "Going Amiss in Experimental Research"). His typology reflects the functional principles of these machines. Moreover, the comparison of these errors with pathological functions is an instructive device for understanding the difference between artificial and living systems.

Finally, bringing together the various investigations from the different domains provides the basis for the overall epistemological evaluation of the role of going amiss in science. The productivity of going amiss now becomes apparent.

The Productivity of Going Amiss

The study of going amiss is productive in a double sense. First, the exploration of what was going amiss in experimentation is often an extremely important step in "making it right". Erroneous results may even stimulate the advancement of an

entire scientific field. To be sure, not all cases of going amiss are productive. When Dewar interpreted the anomalous results in the specific weight of gases as indicating a new form of Nitrogen, he maneuvered himself into a dead end and never managed to get out of it (chapter, "A Pioneer who Never Got it Right: James Dewar and the Elusive Phenomena of Cold"). However, most of the episodes related in this book demonstrate the productivity of going amiss. They further underline that there are different ways in which going amiss can be beneficial.

Going amiss can be productive by eliminating options. In the case of the investigation of liquefying Hydrogen, Dewar reached another dead end. However, this time his research created a background that was productive for the work of his contemporaries who had become acutely aware that this line of research did not lead anywhere. Going amiss can also be fruitful as a motivation to re-direct research. Thenard encountered an anomaly in his experiments—something happened that should not have happened given the standard concepts of his time (chapter, "In the Thick of Organic Matter"). The encounter took him by surprise and re-directed his attention from the technical concerns of his experimental agenda to novel conceptual questions. This re-direction of attention initiated a new line of research that resulted in the formation of an entirely new discipline.

Both Krebs and Faraday exhausted their conceptual tools in the course of their respective research pursuits (chapters, "Concepts from the Bench: Krebs and the Urea Cycle" and "How Experiments Make Concepts Fail: Faraday and Magnetic Curves"). This led them to retreat and re-assess the experimental situation. In both cases, only the reconceptualization of the experimental situation made it possible for them to advance. In the case of the β decay, a controversy arose among different groups of investigators about whether the spectrum of the decay was continuous or discrete (chapter, "The Spectrum of β Decay: Continuous or Discrete? A Variety of Errors in Experimental Investigation"). The various groups could not reach an agreement. The ongoing controversy motivated Ellis and Wooster to rethink the experimental approach and thereby free it from all disputed hypotheses. Here it was not a reconceptualization of the experimental situation but its replacement with an altogether different experimental approach that made it possible to resolve the controversy and establish a new physical fact. All these episodes illustrate that for the scientific researcher and indeed for the whole community the comprehension of going amiss can be productive in many ways.

Secondly, the inquiry into going amiss is a profitable resource for historians and philosophers of science. In the case of Fizeau's ether drift experiment, the contemporaneous actors' positive response is as uninformative as the present-day negative assessment (chapter, "Going right and making it wrong: The reception of Fizeau's ether-drift experiment of 1859"). However, in hindsight, one can trace the gradual transformation from an unproblematic experiment to an experimental result that today is no longer of interest to the practitioners. This case is instructive because it can give clues as to how an experiment may or may not enter the canon of a discipline. Examining the dynamics that underpins this process of exclusion gives historians and philosophers of science insight into the way in which scientific claims coalesce into knowledge. By contrast, the episode of the globules is instructive because

the investigation of the scientists' discourses about the possibility of going amiss reveals previously neglected driving forces for scientific advancement, in particular, the productivity of the proliferation of conflicting experimental results (chapter, "Error as Historiographical Challenge: The Infamous Globule Hypothesis").

This volume places the phenomena of going amiss in the limelight. Understandably, scientists do not wish to be remembered for their errors, confusions and dead ends; they want to succeed, not to be detained by going amiss. Still, they cannot avoid in their daily routine the possibility of going amiss which comes in a variety of unforeseen ways. The volume captures the many ways in which experimenters may go amiss. At the same time, it argues that historians and philosophers of science may find the phenomena of going amiss rich and productive. This should not come as a surprise since going amiss is part and parcel of the practice of experimental research. What is surprising, however, is the fact that so far there has not been an attempt to address this unwieldy theme in a systematic way. Our volume aims to fill this lacuna and at the same time to turn the methodological analysis of impediments and hurdles into a rich tool for better understanding the process of generating knowledge in experimental research.

Giora Hon
Jutta Schickore
Friedrich Steinle