

Science as Knowledge, Practice, and Map Making: The Challenge of Defining Metrics for Evaluating and Improving DOE-Funded Basic Experimental Science

M. Bodnarczuk



National Renewable Energy Laboratory
(formerly the Solar Energy Research Institute)
1617 Cole Boulevard
Golden, Colorado 80401-3393
A Division of Midwest Research Institute
Operated for the U.S. Department of Energy
under Contract No. DE-AC02-83CH10093

Prepared under task no. 32001000

March 1993

NOTICE

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

Printed in the United States of America
Available from:
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161

Price: Microfiche A01
Printed Copy A03

Codes are used for pricing all publications. The code is determined by the number of pages in the publication. Information pertaining to the pricing codes can be found in the current issue of the following publications which are generally available in most libraries: *Energy Research Abstracts (ERA)*; *Government Reports Announcements and Index (GRA and I)*; *Scientific and Technical Abstract Reports (STAR)*; and publication NTIS-PR-360 available from NTIS at the above address.

Science as Knowledge, Practice, and Map Making:
The Challenge of Defining Metrics for Evaluating and Improving
DOE-Funded Basic Experimental Science

Mark Bodnarczuk
National Renewable Energy Laboratory
1617 Cole Boulevard
Golden, Colorado 80401

Abstract

Industrial R&D laboratories have been surprisingly successful in developing performance objectives and metrics that convincingly show that planning, management, and improvement techniques can be value-added to the actual output of R&D organizations.¹ In this paper, I will discuss the more difficult case of developing analogous constructs for DOE-funded non-nuclear, non-weapons basic research, or as I will refer to it - basic experimental science. Unlike most industrial R&D or the bulk of applied science performed at the National Renewable Energy Laboratory (NREL), the purpose of basic experimental science is producing new knowledge (usually published in professional journals) that has no immediate application to the first link (the R) of a planned R&D chain. Consequently, performance objectives and metrics are far more difficult to define. My claim is that if one can successfully define metrics for evaluating and improving DOE-funded basic experimental science (which is the most difficult case), then defining such constructs for DOE-funded applied science should be much less problematic. With the publication of the *DOE Standard - Implementation Guide for Quality Assurance Programs for Basic and Applied Research (DOE-ER-STD-6001-92)* and the development of a conceptual framework for integrating all the DOE orders,² we need to move aggressively toward the threefold next phase: (1) focusing the management elements found in *DOE-ER-STD-6001-92* on the main output of national laboratories - the experimental science itself; (2) developing clearer definitions of basic experimental science as practice not just knowledge; and (3) understanding the relationship between the metrics that scientists use for evaluating the performance of DOE-funded basic experimental science, the management elements of *DOE-ER-STD-6001-92*, and the notion of continuous improvement.

¹ See Jerry D. Holmes and David J. McClaskey, "Improving Research Using Total Quality Management" published in *The Proceedings of the Juran Institute - IMPRO92*, November 11-13, 1992, Session 3B: Quality In Research and Development, p. 13 ff.

² See Mark Bodnarczuk, "A Conceptual Framework for Using DOE 5700.6C and the other DOE Orders as an Integrated Management System: The Fermilab Experience" published in *The Proceedings of the Nineteenth Annual ASQC National Energy and Environmental Quality Division Conference*, September 20-23, 1992, (Fermilab-Conf-92/154).

I. Roadblocks to Focusing on the Main Output of National Laboratories

Attempts to focus management elements like those contained in *DOE-ER-STD-6001-92* on the main output of laboratories run *by* scientists *for* scientists have been difficult because laboratory management and the scientists, engineers, and technicians who constitute the workforce of scientific practice have tended to "stiff-arm" attempts to implement them, claiming that these management elements have *nothing at all* to do with the conduct of basic experimental science. In what follows, I will discuss why I believe that this claim is fundamentally misguided and how the real issue is an organizational resistance to the cultural change that is occurring in DOE-funded laboratories. Juran summarizes the point nicely by stating that organizational change is always accompanied by an "uninvited guest" - the social consequence of the change.³ The social consequence becomes manifest as disruptions in the culture (habits, beliefs, attitudes, practices, traditions, status, values, etc.) and, consequently, the people who are affected develop strategies for resisting these perturbations. They believe that the price of change is too high, even though they frequently admit that the changes would probably be beneficial to the organization.

One type of defense strategy is to turn the application of these management elements into a "paper" exercise where huge volumes of documents (seen only by upper laboratory management bureaucrats) are developed. This approach gives the appearance of implementing the management elements, but allows upper laboratory management to hermetically "seal-off" the line organization from ever having to deal with these "bureaucratic" matters. Often these paper programs *cannot* be implemented because the documents either say nothing (they are too general and vague), or they say far too much (they are so prescriptive that they would grind a functioning laboratory to a halt). This strategy is extremely effective because sealing off the line organization almost guarantees that the management elements will never be focused on the main output of the laboratory - the experimental science itself. Other strategies for giving the appearance of applying the management elements include either viewing "implementation" as the act of writing one more layer of "interpretive" documents without ever actually doing anything, or by applying the management elements only to the support organizations.

A second type of strategy is for laboratory management to apply the management elements found in *DOE-ER-STD-6001-92* only to environment, safety, and health (ES&H) activities. This approach worked magnificently in the era of Tiger Teams because of the "tired" inspection mentality that was embodied in the "quality verification" criteria used by Tiger Teams. A third type of strategy is to claim that planning, management, and improvement techniques destroy the atmosphere of creativity that is so essential to the production of basic experimental science. I have described elsewhere how appropriate management boundaries are actually essential to fostering a creative atmosphere.⁴ The key is to develop asymptotic management boundaries within which there is a confined freedom

³ See J.M. Juran (ed.), *The Quality Control Handbook*, 3rd. ed, (New York: McGraw-Hill Book Company, 1979), p. 7-27 ff.

⁴ See Mark Bodnarczuk, "Assuring Both Quality and Creativity in Basic Research" published in *The Proceedings of the Seventeenth Annual ASQC National Energy Division Conference*, September 9-12, 1990, (Fermilab-Conf-90/65).

- where personnel are encouraged to have a critical and irreverent attitude toward the "status quo," accepting only those systems that have been closely scrutinized and survived the gauntlet of peer evaluation. Unfortunately, laboratory managers have had to develop all these (and other) strategies to protect working scientists from the inappropriate and overly prescriptive oversight of some assessors. But all too often, the strategies also conceal a lack of effective planning, management, and improvement techniques for the conduct of basic experimental science.

As I will describe below, the *NREL Management Program* attempts to focus the planning, management, and improvement techniques found in *DOE-ER-STD-6001-92* on the main output of the laboratory by integrating them with the elements of *DOE Order 5700.7C (Work Authorization System)* and then building these management elements directly into the field work proposals (FWPs) that define the experimental work. But before I turn my attention to these matters, I will discuss another roadblock to even beginning to think about basic experimental science in terms of the performance objectives and metrics - the tendency to define basic experimental science strictly as *knowledge*, not *practice*.

II. Characterizing Basic Experimental Science as Practice, Knowledge, and Map Making

In a previous paper, I characterized the dysfunctional tendency to focus *either* on performance objectives (products) or on the fine-grained details of the processes that lead to those products - processes that may or may not have salient causal effects on the outcome/products.⁵ I proposed instead an approach characterized by a *dialectic* tension between performance objectives (products) and processes, where the characteristics of the products powerfully constrain the process at every point and act as heuristics that point to problems back in the process. In this section I will describe why there is a dialectic tension between science-as-knowledge (product) and science-as-practice (process) even though many scientists define basic experimental science exclusively by its product - knowledge.⁶

To begin with, the tendency to define science exclusively as knowledge is typified by the heavy semantic burden that the word "science" and its derivatives have been forced to bear. While the word science does come from the Latin *scientia* (having knowledge), its primary meaning with reference to experimental science following the 17th century scientific revolution is the experience of *vexing* nature to understand its properties.⁷ For

⁵ See Mark Bodnarczuk, "The Application of 10CFR830.120 in a Basic Research Environment" published in *The Proceedings of the Eighteenth Annual ASQC National Energy Division Conference*, October 6-9, 1991, (Fermilab-Conf-91/95), p. 3 ff.

⁶ See Andrew Pickering, "From Science as Knowledge to Science as Practice" in Andrew Pickering (ed.) *Science as Practice and Culture*, (Chicago: The University of Chicago Press, 1992), pp. 1-26 and also David Hull, *Science as Process; An Evolutionary Account of the Social and Conceptual Development of Science*, (Chicago: The University of Chicago Press, 1988).

⁷ I am thinking primarily of Francis Bacon, see Francis Bacon, *The Works of Francis Bacon*, James Spedding, Robert Leslie Ellis, Douglas Denon Heath (eds.), (Boston: Taggard and Thompson, 1893); and Robert Boyle, see Robert Boyle, *The Works of the Honourable Robert Boyle*, (London, 1772).

Galileo, the act of *doing* an experiment (not the voyeuristic study of Aristotelian "texts") produced a kind of knowledge that could not be obtained in any other way.⁸

The published scientific paper is produced after-the-fact (downstream) of the creation of physical effects in the experimental apparatus (many of which do not occur in nature unattenuated).⁹ Viewed metaphorically, all the documenting and recording activities that go into the production of a scientific paper constitute the cartographic function of science - the map making function. In the same way that explorers have to physically travel through an unexplored region of the earth (drawing maps of that region as they go), the experimental scientist uses her experimental apparatus as an extension of her physical senses to explore uncharted physical domains, drawing experimental maps of the measurements and newly created effects and phenomena she encounters. If the experimental map is drawn accurately, it enables other experimenters to move directly to the frontier of scientific knowledge and continue mapping out physical effects that lay beyond that point. But because most experimental results are published as "pre-prints" and presented orally at conferences long before they appear in a professional journal, primary publications have *much less* to do with the transmission of scientific knowledge and *much more* to do with establishing priorities for discoveries or as valuable commodities that are traded back to the laboratory or funding agency in exchange for more resources to perform more experiments.¹⁰ It is the *practice* of science that most appropriately defines what basic experimental science *is*, not the experimental maps that are subsequently published in journals and traded back to funding agencies for additional resources.

Given the importance of scientific practice, it is somewhat enigmatic that scientists tend to remove references to human and social factors from their accounts of science, leaving the impression that the knowledge or "facts" contained in scientific publications rise above the organizational and human factors of laboratory life once the experiment is completed.¹¹ Part of this is due to the literary convention established by the editors of scientific journals that reinforces (or helps cause) this tendency to eradicate human factors,

⁸ Galileo Galilei, *Two New Sciences - Including Center of Gravity and Force of Percussion*, trans. by Stillman Drake, (Madison, WI: The University of Wisconsin Press, 1974), p. 66.

⁹ Hacking claims that certain effects (like the Hall effect) do not exist in nature outside of certain kinds of apparatus. While passing a current through a conductor at right angles to a magnetic field will produce a potential anywhere in nature, nowhere outside the laboratory (the technological devices we construct), is there such a pure arrangement. See Ian Hacking, *Representing and Intervening*, (New York: Cambridge University Press, 1987), p. 225 ff.

¹⁰ In a case study of high-energy physics experiments performed at Fermilab, I describe a number of distinct, yet interrelated, resource economies (each having its own commodity) within which high-energy physics collaborations must trade to obtain the resources needed to perform experiments. The resource economies and commodities are proton economics (high-energy protons from the accelerator), experimental real estate (a place to setup the apparatus), physicist economics (physics expertise), computing economics (on-line and off-line computational power), and physics economics (the publication that is traded back to the Director for more resources). See Mark Bodnarczuk, *The Social Structure of Experimental Strings at Fermilab; A Physics and Detector Driven Model*, Fermilab-Pub-91/63, March 1990, p. 2 ff.

¹¹ The tendency to leave out social factors has been documented by numerous authors like Daniel J. Kevles, *The Physicists: The History of a Scientific Community in Modern America*, (Cambridge, MA: Harvard University Press, 1987), p. xiv; Andrew Pickering, *Constructing Quarks: A Sociological History of Particle Physics*, (Chicago: The University of Chicago Press, 1984), p. 5 ff., and Bruno Latour and Steve Woolgar, *Laboratory Life*, (Princeton, NJ: Princeton University Press, 1986), p. 28.

laboratory organizational issues, social factors internal and external to the collaboration, human judgments and interpretations of data or theories, discussion of other scientists' perceptions of the skills of experimenters.¹² As important as these human and social factors are to describing what science is within the context of a national laboratory, they are *left out* of the scientist's retrospective account. The point is that if scientists leave all the human and social factors out of their descriptions of science, *no wonder* they can claim that planning, management, and improvement techniques have nothing at all to do with the conduct of basic experimental science.

When science-as-practice and science-as-knowledge are viewed dialectically, it is clear that the physical effects described in scientific publications *do not* rise above the human and material setting of laboratory life once they are published as experimental maps, but are inextricably bound to the experimental apparatus and the mundane work processes of the organizational infrastructure and collaboration who performed the experiment.¹³ Modern day experiments *do not exist* independent of the organizational infrastructures of the laboratories in which they are performed, and all aspects contained in the scientific publication must be ultimately reducible to work processes performed in (or associated with) the laboratory. An experiment/collaboration is not an "object" that has unambiguous boundaries and is entirely separable from the organizational infrastructure of the laboratory. For example, although experimentalists attempt to draw a firm line of demarcation around the "collaboration" and its activities for the sake of defining which names appear on the scientific publication, the fact is that numerous laboratory personnel often play crucial roles in experiments, and whether or not their names appear on the published papers is a socially negotiated matter that is decided by the personalities involved.¹⁴ As described in *DOE-ER-STD-6001-92*, there is no unambiguous "wall" between an experiment and the laboratory in which it is performed because some researchers may also be laboratory staff members, other researchers may be from industry, yet others may be "outside users."

Earlier, I alluded to the claims of some scientists that planning, management, and improvement techniques have nothing at all to do with the effective conduct of basic experimental science. This claim can only be supported if one first extracts all references to human, social, and organizational infrastructure from the definition of what science is, *then* characterizes science exclusively as knowledge. If one can define the theoretical and experimental background against which an experiment is performed along with the measurement properties and characteristics of the experimental apparatus and the data recording techniques (which collaborations have to define for the publication), and then

¹² See Harry M. Collins (ed.), *Knowledge and Controversy: Studies of Modern Natural Science*, special issue of *Social Studies of Science*, vol. 11, no. 1, and also Trevor J. Pinch, "The Sun-Set: The Presentation of Certainty in Scientific Life," in *Social Studies of Science*, vol. 11, pp. 131-58.

¹³ See Hacking, *Representing*, pp. 149 ff., and Bruno Latour, *Science in Action*, (Cambridge, MA: Harvard University Press, 1987), p. 63 ff.

¹⁴ Nobelist Melvin Schwartz shows how tenuous these socially constructed walls are for large collaborations when he advocates divorcing some of the detector builders in modern high-energy physics experiments from the collaboration, then subdividing the remaining members of these mega-collaborations into distinct (smaller) collaborations that would develop their own research programs and compete for time using the detector. See Faye Flam, "Big Physics Provokes Backlash" in *Science*, 11 September 1992, vol. 257, p. 1470.

add back in descriptions of the human, social, and managerial processes that constitute the laboratory organizational infrastructure and the experimental collaboration, *then* the factors that constitute *science-as-practice* begin to re-emerge. Many of the human, social, and managerial elements are included in the elements of *DOE 5700.7C (Work Authorization System)* Attachment 3, which provides guidance for requesting DOE funding for FWPs. But sadly, the FWP has come to be viewed primarily as a budget document not the strategic planning and management tool that it was meant to be. The *NREL Management Program* builds the planning, management, and improvement techniques found in *DOE-ER-STD-6001-92* right into the FWP and attempts to utilize them as management plans defined by the scientist herself.

III. Defining Metrics for the Evaluation and Improvement of Basic Experimental Science

In this section, I will use the notion of a "metric" to indicate a standard of measurement, but I want to make the distinction between a direct metric (a close logical, causal, or consequential relationship to the things being measured) and an indirect metric (a collateral or circumstantial relationship to the things being measured). When possible, direct metrics are to be preferred because the data they provide more veridically characterize the thing being measured. In this section I will show how the metrics for evaluating science-as-knowledge and science-as-practice meet along the interface of the metric of *stability* because (as I described earlier), the knowledge contained in scientific publications is inextricably bound to the experimental apparatus and the mundane realities of scientific practice. It is at the interface of stability that one must define the relationship between the in-process direct metrics defined by scientists and the management elements in *DOE-ER-STD-6001-92*. I will describe why this type of in-process metric for science-as-practice cannot be used to evaluate science-as-knowledge until after an experiment is complete because evaluations of science-as-knowledge are *by necessity retrospective*. I will also describe the notion of continuous improvement for both science-as-knowledge and science-as-practice.

Although his work has been plagued by various philosophical problems,¹⁵ Thomas Kuhn's notion of a "paradigm" is still one of the most compelling conceptual frameworks for evaluating science-as-knowledge and science-as-practice.¹⁶ For Kuhn, a paradigm is a matrix of "consensus standards" for a scientific community that is constituted by a shared network of commitments to conceptual, theoretical, instrumental, and methodological

¹⁵ For a brief description of these philosophical problems see John Horgan, "Profile: Reluctant Revolutionary," in *Scientific American*, May 1991, pp. 40 and 49.

¹⁶ Other types of metrics are embodied in Alvin Weinberg's internal and external criteria that he claimed should be used to evaluate which types of science receive funding. His internal criteria should answer two questions. First, is the field ready for exploitation? Second, are the scientists in the field really competent? Weinberg claimed that these decisions could be made only by scientists. He identified three external criteria that could be decided by non-scientists: technological merit, social merit, and scientific merit. The criterion of scientific merit assessed the degree to which the knowledge produced by the discipline requesting funding contributed to its neighboring disciplines. See Alvin Weinberg, "Criteria for Scientific Choice" in *Minerva*, vol. 1, 1963, pp. 159-171.

ways of doing science.¹⁷ These "consensus standards" are articulated in the textbooks and journal publications (experimental maps) used to train new scientists in that field and a scientist gains status in the community to the degree that she can articulate the parameters of the paradigm and design theoretical and experimental puzzles that heuristically probe and test it in every conceivable way. In other words, when a collaboration claims to have created physical effects that have never before been produced in the laboratory, the parameters of the effect are compared to the consensus standards of the paradigm - to everything else we know - to test the validity of the claim. These tests are most convincing when they use methodologies or apparatus that rely on laws of physics unrelated to the first experimental design.

Metrics for Science-as-Knowledge

If one is evaluating whether an experiment has produced "good" physics, "bad" physics, or just "meat and potato" physics, one could use Irvine's and Martin's indirect metric of counting the number of citations that a particular publication receives, but this type of indirect metric is based on the inference that if a scientific publication contains important scientific results then it will be cited more frequently than one that does not.¹⁸ Whatever this type of metric tells us, it provides no direct indication of the quality of the content of the publication like the two types of direct metrics for science-as-knowledge that I will now discuss.

The first type of direct metric is described by Peter Galison's dual metrics of directness and stability. By directness, he means activities that bring experimental reasoning another rung up the ladder of causal explanation like the measurement of a background that was previously only calculated or the separate measurement of two sources of an effect previously only measured together.¹⁹ By stability, he means all the experimental procedures that vary some feature of the experimental setup (including changes in the test substance, the apparatus, the arrangement, or the data analysis) and leave the result basically unchanged. The metric of stability is especially important when a collaboration claims to have created a physical effect that has not previously occurred in nature or in the laboratory. The stability, stubbornness, or robustness of the effect is defined by the fact that the collaboration has controlled for all known background effects and altered numerous parameters of the experimental apparatus to control for artifacts, and the proposed effect "just won't go away."²⁰ More importantly for this metric, other experimentalists can test the reality or artifactuality of the proposed effect by attempting

¹⁷ Thomas Kuhn, *The Structure of Scientific Revolutions*, 2nd ed., enlarged, (Chicago: The University of Chicago Press, 1970), p. 35 ff.

¹⁸ See J. Irvine and B. R. Martin, "Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators," in *Social Studies of Science*, 1985, vol. 15, p. 300, and also Steven Yearly, *Science, Technology, & Social Change*, (London: Unwin Hyman, 1988), p. 88 ff.

¹⁹ Galison, pp. 259-260.

²⁰ Galison, p. 234 ff. Latour's describes fundamentally the same notion in terms of the "trials of strength" that various effects and substances endure at the hands of experimenters in Latour, *Science in Action*, p. 74 ff.

to produce it using apparatus that rely on totally unrelated laws of physics.²¹ The fact is that each variation introduced into the experimental design makes it more difficult to postulate an alternative causal story that will satisfy all the observations because the effect is nested within ever more complex loops of experimental demonstration.²² William Wimsatt describes the theoretical equivalent, namely how our scientific theories and beliefs about the world are nested (generatively entrenched) within yet wider systems of beliefs. Theories and beliefs that are generatively entrenched are called "robust" because of their interconnectedness to the entire body of knowledge.²³

A second type of direct metric for science-as-knowledge is captured in the function of "crucial" experiments. While some recent philosophers of science have come to interpret the function of crucial experiments as absolute arbiters of scientific theories, this is not how it was conceived by the father of experimental science, Francis Bacon. As pointed out by Hacking, a more literal rendering of Bacon's Latin *-Instantiae crucis* - is instances of the "crossroads" or "finger posts," conjuring up the image of finger post signs that were set up where roads parted in order to indicate several directions.²⁴ In other words, crucial experiments are those that indicate any new (previously unknown) direction that yields new previously uncreated effects, phenomena, or more direct and stable measurements, not necessarily those that test competing theories. If a scientist performs an experiment that sends her down the wrong road (does not lead to new effects, etc.), this experiment may be both crucial and unproblematic because it produces knowledge of ways in which nature *will not* be vexed, and this is valuable information to be added to the overall cartographic file on that region of nature. The scientist simply retraces her experimental steps back to the misleading finger post, selects another direction, and continues on exploring and experimental map making.²⁵

While the principles of *continuous improvement* are normally associated with a manufacturing environment, designing experimental apparatus that will yield more direct and stable measurements is one crucial component of continuous improvement in experimental science-as-knowledge. Although it is common folklore to claim that experiments are repeated, in reality scientists almost never repeat the same experiment. Rather, follow-up experiments are almost always attempts to *improve* the directness and stability of some parameter of the same phenomenon, often using improved techniques for detection, data acquisition/monitoring, or data analysis. Elsewhere, I have performed a case study that shows how high-energy physics experiments are actually performed in "strings" where previous experiments and follow-up experiments meet at a transition-like interface where the first experiment is transformed into its follow-up progeny by the

²¹ For example, Hacking describes how interference, polarizing, phase contrast, direct transmission, and fluorescence microscopes can be used to discern the same basic microstructure using essentially unrelated aspects of light in Hacking, pp. 186-209 and Galison describes how neutral currents could be detected using bubble chambers or electronic counter detectors in Galison, pp. 135 ff.

²² Galison, p. 260.

²³ See William Wimsatt, "Robustness, Reliability, and Overdeterminism" in M. Brewer and B. Collins (eds.), *Scientific Inquiry and the Social Sciences*, (San Francisco: Jossey Bass Publishers, 1981).

²⁴ Hacking, p. 249.

²⁵ I have purposely avoided the problems created by Kuhn's notion of the incommensurability of paradigms, i.e., the inability of experiments to adjudicate between competing paradigms. These problems are discussed in considerable detail in Hacking, pp. 65-74.

collaboration.²⁶ Each experimental configuration in the string displays a more complex iteration (improvement) of the original apparatus design but leaves the fundamental design of the experiment largely in tact. The case study also shows how these variations in experimental design are anchored to certain continuities in the physics goals, the apparatus configuration, and the scientists who perform the experiment. These continuities allow the experimenters to (1) avoid the physics uncertainties involved in designing and building entirely new experimental detectors, (2) continuously improve the directness and stability of the measurements, and (3) avoid the sociological uncertainties of securing resources from the laboratory or the DOE program office. I believe that *continuous improvement* in the smaller scale experimental "strings" that constitute research programs funded by "strings" of FWP's exhibit similar experimental and sociological continuities, with variances being due largely to the scale, cost, and complexity of the experiments.

Continuous improvement through experimental "strings" assumes that the experiments are being performed *within* a given scientific paradigm, or as Kuhn would describe it, it is the puzzle solving that typifies "normal science" where scientists extend and improve the scope and precision of the paradigm.²⁷ But as I will now describe, new paradigms are normally improvements for a scientific discipline. When multiple anomalies emerge and a paradigm cannot adequately accommodate them, a scientific discipline reaches a crisis state (revolution) where scientists begin abandoning their previous theoretical and experimental commitments in favor of a new paradigms that can (1) explain the anomalies, (2) explain a wider range of natural phenomena, (3) account for known phenomena with greater precision, or (4) account for phenomena that were previously unknown. While some of the previously held beliefs and experimental practices are replaced by various aspects of the new paradigm during revolutions, many of the beliefs and experimental practices and most of the experimental measurements become embedded in an improved conceptual framework that allows scientists to postulate new experimental problems and produce more reliable experimental "maps" of natural phenomena.²⁸

²⁶ See Mark Bodnarczuk, *The Social Structure of Experimental Strings at Fermilab; A Physics and Detector Driven Model*, Fermilab-Pub-91/63, March 1990, p. 14 ff.

²⁷ Kuhn, p. 24 ff.

²⁸ Kuhn uses the evidence of rewritten science textbooks and other associated science literature to claim that revolutionary scientific knowledge is not cumulative like the knowledge obtained over the course of normal science. For Kuhn, the emergence of new paradigms (and their resolution) does not bring science closer to some "truth" about the physical world, but rather like orthodox Darwinian theory, it is "The selection by conflict within the scientific community of the fittest way to practice future science." While I do not believe that paradigms are cumulative in the sense that all beliefs, practices, and measurements "survive" scientific revolutions and move science closer to some "truth" about the physical world, I do believe that paradigms show progress toward more robust and reliable representations of the physical world. On this view, the "mapping" or agreement between theoretical constructs and experiment data becomes more precise and reliable. See Kuhn, p. 52 ff on anomalies, pp. 140-143 on the non-cumulative aspects of revolutionary science, and pp. 170-172 on the analogy with orthodox Darwinian natural selection.

Metrics for Science-as-Practice

When a scientist proposes a program of basic research in an FWP he or she is normally standing on the frontier of knowledge beyond which no experimental maps have yet been drawn. As such, the scientist makes certain *knowledge claims* that are calculated hypotheses of the experimental path that must be taken and the expected results that may or may not be corroborated when the experiment is actually performed. Knowledge claims define what new knowledge the scientist believes he or she will obtain by performing the experiment. But the scientist must also make certain *practice claims* in an FWP where he or she defines planning and managing elements that can be thought of as strategic predictions of how that new knowledge will be obtained. This includes estimates of how long it will take to travel the experimental path, the types of experimental equipment that will be needed, how much on-line and off-line computing resources will be needed, how much engineering and technician support will be needed, and other costs and resources. If the scientist begins this experimental journey as planned and performs an experiment that turns out to be a *crucial* experiment (pointing in a different direction), the scientist is obliged to follow the data, and the requirements of *DOE Order 5700.7C* allow for this type of change in the direction of the research. Often, the practice claims in FWPs for experiments that are two or more years away are "place holders" because the knowledge claims are like moving targets that cannot be well defined. But the knowledge claims become increasingly refined and fine-tuned as they move toward the performance of the current year's research, and, consequently, the practice claims can be articulated and evaluated more precisely.

In terms of defining *direct* metrics for science-as-practice, once the experimental design (conceived at the frontier of knowledge), procurements, installation, and overall configuration of the apparatus has settled down enough to actually perform the experiments, the crucial concern is to bring the operation of the experimental apparatus into a "steady state" in which all possible operational parameters of the apparatus are understood and functioning as designed. But at this stage of experimenting, the vast majority of the processes that occur in the organizational infrastructure of the laboratory in which an experiment is embedded have little or no causal efficacy on the outcome of an experiment. In addition, scientists construct organizational and experimental "protective belts" around their work to protect it from all but the most devastating laboratory perturbations - usually resource and funding related.²⁹ So how does one isolate the "vital few" activities and support organizational interfaces that can actually affect the output of an experiment? The vital few cluster along the interface between science-as-knowledge and science-as-practice, i.e., the stability of the experimental apparatus upon which the knowledge that will eventually be published is based. My claim is that of all possible laboratory activities, the vital few are defined as only those that exert direct causal effects on the stability of the apparatus and computing because only those can affect the outcome of the experiment itself. Experiments in laboratories should be viewed as a locus of

²⁹ One exception is the case of high-energy physics where the reliability, availability, and maintainability of the accelerators that deliver particle beams to experiments is crucial, and experiments cannot construct "protective belts" to protect them from these problems because they are totally dependent on the delivery of these beams to conduct experiments.

equipment and software, etc. around which a collaboration gathers resources and support in ever-widening circles, with each circle of resources exerting less and less causal efficacy on the experimental work. The vast majority of perturbations in laboratory activities never penetrate to the first few circles that surround an experiment, and those that do are normally absorbed by the innermost experimental and organizational protective belts constructed by scientists.³⁰

By way of illustration, one of the "vital few" components of the metric of stability is backgrounds and noise in the apparatus. If the goal of the experiment is to create new previously undetected phenomena, then the apparatus must physically isolate these properties and at the same time damp down all other non-salient effects that could present themselves as artifacts. While it is impossible to enumerate all conceivable backgrounds, Galison describes how the collaborators must demonstrate to themselves (and eventually to other scientists) that they have accounted for all known backgrounds either by constructing the apparatus to block them, or measuring and calculating those backgrounds and subtracting them from the data results.³¹ Having accounted for all known backgrounds in construction, measurement, and calculation, the vital component of the metric of stability is to reduce the variances between the parameters of the experimental design and the apparatus' actual performance, and to validate or calibrate the operational parameters of the apparatus against the values of physical phenomena that are well understood.

Other examples of the "vital few" that could affect the stability of an experiment if they are not done properly are things like (1) visually or computationally monitoring the apparatus to ensure proper operation of components like power supplies, gases, calibration of equipment, etc.; (2) ensuring that the proper materials, targets, and chemicals are being used; (3) ensuring that data rates are appropriate and that the data acquisition and software systems are functioning as designed and intended; (4) ensuring that the cartographic function of basic experimental science is carried out in scientific notebooks and that correct data are recorded; and (5) ensuring that the appropriate measurement uncertainty analysis is performed on experimental results that are reported in the literature. Only these types of things can seriously affect the production of scientific knowledge. They are a minority of laboratory activities. They are the only things that scientists are even interested in implementing management controls on, and even a casual reading shows that they are simply a laundry list of the management elements found in *DOE-ER-STD-6001-92*.

I already described how knowledge claims and the associated practice claims become increasingly fine-tuned as the scientist moves toward the current year's research, and, consequently, these practice claims can be articulated and evaluated more precisely. One would think that at this stage of the scientific process one could begin to evaluate whether or not the practice claims made in the FWP actually get cashed-out into the performance of the experimental work. In other words, one ought to be able to evaluate science-as-practice using the indirect metrics that constitute the majority of the practice

³⁰ One of the types of activities that are becoming more and more difficult to absorb are the costs and resources associated with regulatory compliance. Philip Ablest claims that "There is a growing questioning of the factual basis for federal command and control actions and the scientific competence of the regulators." Philip Ablest, "Regulatory Costs" in *Science*, vol. 259, January 8, 1993, p. 159.

³¹ Galison, pp. 2-3, and 255-257.

claims that scientists make in FWP's (time frames, milestones, equipment, computing, and human resource costs). Even though the direct and indirect metrics for science-as-knowledge and the direct metrics for science-as-practice are compatible with the management elements found in *DOE-ER-STD-6001-92*, I want to suggest that there are *cultural values* held by the scientific community that make the evaluation of the bulk of the practice claims made by scientists problematic.

The direct metrics for science-as-knowledge and science-as-practice can be defined and agreed upon only because scientists *value* more direct, stable measurements and crucial experiments. What a community values *defines* its most salient metrics because values constitute the criteria of choice.³² Previously I mentioned the enigmatic tendency for scientists to omit the mundane organizational and human factors of laboratory life from accounts of science once an experiment is completed. While scientists are forced to make practice claims about planning and managing their experimental work in order to obtain funding, most of them would prefer not to deal with the pedestrian realities of milestones, cost estimates for equipment, computing, graduate students, and especially deliverables like publications. In a scientific culture that mythologizes (and nostalgically exalts) the archetype of the solitary, irreverent, creative, rebel scientist who emerges from his laboratory only to reveal new secrets of nature, being a great theorist or experimentalist is coveted but being a great planner or manager of experimental work is viewed (at best) as a set of necessary skills that someone else should have so the scientist can go about the business of doing "real" scientific work.

In some ways, the tendency to devalue the planning and managing aspects of science is driven by the fact that most performance claims found in FWP's or other experimental proposals do not seem to contribute meaningfully to the production of science-as-knowledge, and an evaluation of these metrics gives no indication of the quality of the basic experimental science. Once the publication has been written, no one in the scientific community asks about how much the experiment cost, whether it was completed on time, how much scientists exceeded their allotted computing time etc., so the successful achievement of a scientist's practice claims is trivialized as a value that is not to be sought by the very best and brightest scientists. Unfortunately, this attitude is modeled by senior professors for graduate students, which ensures that the problem will be perpetuated.

Indirect metrics for science-as-practice can only be defined when the majority of scientists (and funding sponsors) agree that they need to adopt a new philosophy that does not *automatically accept* missed milestones, increased requests for computing, and the inability to estimate other experimental resources as an *inevitable* aspect of doing basic experimental science.³³ Only when these practice claims are factored into the scientific community's evaluation of a scientist's *competence as a scientist* and only when funding agencies develop methods for teasing apart the difference between problems that are intrinsic to basic research and those that result from poor planning and management

³² A classic account of the role of values and norms in the scientific community is found in Robert K. Merton, *The Sociology of Science; Theoretical and Empirical Investigations*, Norman W. Storer (ed.) (Chicago: The University of Chicago Press, 1973).

³³ This is point 2 of Deming's 14 point management philosophy. See Henry R. Neave, *The Deming Dimension*, (Knoxville, TN: SPC Press, 1990), p. 293 ff.

techniques, will we be able to use the practice claims defined in FWPs (time frames, milestones, equipment, computing, and human resource costs) as indirect metrics that will enable us to improve the conduct of basic experimental science. Adopting this new philosophy will not improve the *quality* of science-as-knowledge, but it will probably improve the *quantity* of science-as-knowledge by making the practice of science more efficient and effective.

Given the fact that there are no salient differences between the management elements found in *DOE-ER-STD-6001-92* and the direct and indirect metrics for science-as-knowledge and the direct metrics for science-as-practice, I see no fundamental reason why many of the continuous improvement methodologies developed by Juran (Continuous Improvement Projects) could not be tailored, then applied to these problematic practice claims. Scientists and laboratory managers could form cross-functional teams (composed of scientists, engineers, technicians, etc.) that would more closely examine why scientists miss the milestones that they have defined for themselves. Following the standard method of storyboarding, the team could analyze the symptoms of the missed milestones, formulate theories about why they were missed, test these theories, and identify root causes. Was the problem attributable to a limitation that nature imposed on the experiment? Was it impossible to push the technologies involved any further? Was the problem an inevitable part of the pedagogic process of obtaining knowledge by actually *doing* an experiment and could not be avoided? Or was it a systemic problem where a spokesperson had no authority to make her collaborators come through on their commitments? Was it a lack of planning or management on the part of the principal investigator or the inability to stop introducing new parameters and changes into an experimental design that should have been fixed? Was the problem due to the lack of supervision of a graduate student by his or her senior professor or laboratory manager? Problems imposed by nature may be unavoidable, but the practice of science would certainly be improved if most of the other problems were solved.

Epilogue

The history of science is replete with examples of how some unknown (or improperly understood) aspect of an experimental apparatus led to problematic data results that were only realized retrospectively.³⁴ The problem with any type of *preventative approach* is that *during* the process we have only indirect (or intuitive) metrics that justify the cost (in terms of time and money) associated with doing them, and it is impossible to predict how many errors or problems *would have* occurred without this kind of up-front planning. When scientists demand that laboratory managers demonstrate the "value-added" of the management elements in *DOE-ER-STD-6001-92* up-front, all we can do is appeal to history - to those instances where something did go wrong due to poor planning or management. In an analogous sense, we possess the same type of indirect (intuitive, retrospective) metrics for justifying the cost (in terms of time and money) associated with

³⁴ The turn of the century episode with the French scientist Blondlot and N-rays (Latour, *Science in Action*, p. 73 ff), the incident of Fermilab experiment E-1A's "alternating currents" as described in Galison, p. 174 ff. and the much celebrated recent incident of Fleishman's and Pond's "Cold Fusion," are all incidents where a part of the experimental apparatus was not thoroughly understood.

performing basic experimental science itself. Much like planning and managing, it is impossible to predict how many discoveries would have been made if we had increased our up-front commitment to this type of esoteric research. When congress demands that scientists demonstrate the "value-added" of basic research up-front, all scientists can do is appeal to history - to those instances where something was discovered and eventually was developed into new technological processes or products. In a world that demands instant demonstration of the effect of an activity on the "bottom line" before that activity will be supported, the preventive approach to problems will never be popular and will always have to justify its existence by appealing to history.