

## **Appreciating Classic Experiments**

Douglas Allchin  
Editor, SHiPS Resource Center  
University of Minnesota  
Minneapolis, Minnesota

“The most beautiful experiment in biology” — that was how John Cairns described Meselson and Stahl’s work on the nature of DNA replication. What (one may surely ask) earned their study such remarkable praise? —And (just as important, perhaps) how can the practice of science — an experiment — be “*beautiful*”? Answering both questions ultimately shows how appreciation of historical experiments contributes to a fuller understanding of science. Here, I invite you to tour briefly this and three other classic experiments from the history of biology. Each holds valuable lessons for the student of biology. Each also illustrates how one may find such lessons in other great discoveries.

### **Meselson and Stahl and DNA Replication**

Meselson and Stahl gained renown for demonstrating in 1958 that DNA replicates semi-conservatively. Their experiment (described now in many introductory college textbooks and websites) tested Watson and Crick’s notion that when DNA replicates, its double helix splits and each strand serves as a template for its own new complementary strand. Meselson and Stahl’s achievement was twofold. First, they conceived how to label and identify the new versus parent strands. Second, they developed a method to separate the different forms of DNA resulting from successive replications. Labeling was done with isotopes, not based on their radioactivity, but on their different weights. Separation occurred in a density gradient, established with a heavy salt solution in a high speed centrifuge. Macromolecules of modestly different molecular weights would float (at equilibrium) at distinct levels in the gradient. The resulting bands at each generation were visually definitive: “perfect Watson-Crickery,” as celebrated by one researcher (Holmes 2001, p. 368).

The experiment was a paradigm of good practice in several ways. First, it captured a central theoretical question in a single experiment. The problem of DNA replication was certainly not new. Watson and Crick’s model had puzzled researchers for several years. Imagining possible events at the molecular level is relatively easy. Manifesting them in the lab is quite another thing. Sometimes, the molecular realm is revealed piecemeal, in clues and partial glimpses. Here, one well oriented probe sufficed. Second, Meselson and Stahl’s experimental design addressed all the alternative theoretical models of replication simultaneously. Failure to confirm one model would not lead to further tests exploring another. Third, the experiment was marked by laboratory expertise. Material skills matter as much as thinking. The results were “clean” and unambiguous.

Finally, the team also introduced a new method of wide scope. The technique of using heavy isotopes to differentiate macromolecules, once demonstrated, could be applied to many other studies. Meselson and Stahl's experiment thus exhibited creative arrangement of laboratory conditions, theoretical import, clarity, and craft skills, all while pioneering an important new method. Rarely do all such elements come together in one work. When they do, biologists justly celebrate.

“Beauty,” of course, is typically associated with works of art or design. Yet our aesthetic sense responds whenever form and function unite. Scientists thus come to regard some experimental designs as “elegant.” The method of observation and the conceptual interpretation complement each other fully, yet economically. Stahl himself later commented on the perceived beauty in his experiment: “It’s very rare in biology that anything comes out like that. It’s all so self-contained. All so internally self-supporting. Usually, if you are lucky to get a result in biology, you then spend the next year doing all those plausible controls to rule out other explanations; but this one was just a self-contained statement” (Holmes 2001, p. 429). To appreciate Meselson and Stahl’s experiment, then, is to understand how we justify the concepts inscribed in textbooks — that is, to understand (by example) the empirical foundation of scientific knowledge.

Appreciation may go even deeper, however. The final structure of an experiment, as a product, can hide the process that led to it. Classic experiments are prime opportunities to revive science-in-the-making. How did Meselson and Stahl *conceive* their novel experiment? How did they *create* its groundbreaking conditions? How do scientific discoveries *happen*, blind to the eventual outcome? Here, one may explore the *disciplinary and biographical contexts* of Meselson and Stahl’s efforts — all richly documented recently by historian Larry Holmes.

Matt Meselson and Frank Stahl met as graduate students in the summer of 1954 while at Woods Hole Biological Laboratory. Matt was a course assistant for James Watson himself. Frank was taking another course not available at his home institution. Stahl was drinking a gin and tonic under a tree. Watching from the main building, Watson remarked on his reputed fine lab skills. Meselson, curious, went to introduce himself. Frank had been considering a statistical problem requiring calculus. Several days later, Matt offered a solution, impressing Frank in turn. An enduring friendship developed. Before the summer was out, Meselson had mentioned a prospective study on DNA replication and Stahl had structured it experimentally.

Where had Meselson’s idea come from? Early in 1954 he had been working on problems on deuterium for his course with Linus Pauling on chemical bonds. He wondered whether organisms would live in heavy water (made with the hydrogen isotope). Later that spring, Jacques Monod gave a visiting lecture on the *de novo* synthesis of inducible enzymes. Meselson then imagined how to use the heavy isotope to label the new proteins. One could then separate new and old proteins by density in an appropriate solution, one floating to the top, the other sinking to the bottom. The core design for the later experiment on DNA was thus first developed in an entirely different context. A few months later, when Max Delbrück introduced him to his recent models of DNA replication, Meselson saw another application of his scheme — the one he shared with Stahl at Woods Hole in the summer of 1954.

But the route to the last run of the experiment in February, 1958 was hardly direct. Before long, 5-bromouracil replaced the heavy deuterium in their design. It would substitute directly for thymine in the DNA and provide a more dramatic weight difference. This, in turn, led the team to a second line of investigation on mutagenesis, which soon became primary. Finding a solution with an appropriate density involved trial and error. KBr? No. MgSO<sub>4</sub>? No. Ba(ClO<sub>4</sub>)<sub>2</sub>? No. CsCl? Perhaps. But at what concentration? In trying the new technique, they discovered to their dismay that centrifugation destroyed the homogenous density, creating a gradient instead. They had planned to separate the DNA in *discrete* layers! Fortunately, the gradient was gradual enough. The key molecules would still separate. Nonetheless, they explored electrophoresis as a possible alternative. In August 1957, Meselson saw an advertisement for the nitrogen isotope, <sup>15</sup>N. They had rejected it earlier, assuming the weight difference with DNA using <sup>14</sup>N would be too small to measure. The unexpected resolving power of the density gradient method now made it possible once again. Suddenly, 5-bromouracil was abandoned. The mutagenesis inquiry was set aside. The experiment so celebrated by history finally emerged. Note, too a host of practical challenges: competing for time on the centrifuge, finding the right spinning speed or chemicals to lyse the bacteria, getting theses written (on other topics), going on job interviews, etc. If the final experiment was simple, the process was anything but.

“Real” science hardly resembles the cookbook labs one frequently encounters in school classrooms. Nor is it the formulaic “scientific method” enforced on many science-fair projects. Science is a creative enterprise, filled with metaphoric thinking, chance encounters, false starts, tinkering and plain hard work. The final idealized “textbook” experiment may disguise how it developed from a convergence of contingencies. Appreciating Meselson and Stahl’s experiment — as much as appreciating science in general — includes understanding its convoluted history as well as the beauty of the product.

### **John Snow and Cholera**

Other classic experiments hold lessons, as well. Consider John Snow's demonstration in the 1850s that cholera was communicated via water. Snow's work, unfolding in successive layers, effectively underscores the nature and importance of controls, one of the fundamental tools of science. In particular, he studied a monumental “natural experiment” using data originally recorded for other purposes.

An epidemic of cholera swept London in late 1848. John Snow, a prominent physician (who attended Queen Victoria), noticed that the symptoms involved chiefly the intestines and that the patients responded to local treatment. He assumed some poison or toxin was ingested. Perhaps it was also discharged from the intestines. If some chemical process like fermentation amplified it, that might taint the water that others drank. Snow’s evidence was impressionistic, based on informal personal experience. By contrast, the prevailing theories blamed *miasmas*, or infectious airs. William Farr, Statistical Superintendent of the General Register Office, collected data more systematically. He recorded the various deaths and looked for associated factors. He discovered a quantitative law that linked cholera with the elevation of the soil. Farr drew on the statistical correlation to think about the underlying causation. Low-lying areas would foster putrefaction, he claimed, and produce an airborne material, *cholérine*, that caused the disease. Farr’s extensive

data seemed to yield firmer conclusions than Snow's.

Another epidemic appeared in 1853. Snow continued to pursue his notions about water. First, he famously mapped all the incidences. Using his "spot map," as epidemiologists do now, he targeted the water pump on Broad Street. Of course, water pumps were located all over London. A pump would likely be found in the midst of *any* contagion! Snow's reasoning was again only circumstantial. He needed to identify the source of water for *each individual* instance of cholera. This he did. With an assistant, he interviewed the household of every cholera victim. Snow's results were stunning — at least for those who now know that the cholera bacterium is indeed waterborne. Snow addressed all the presumed exceptions. Victims who lived closer to other pumps seemed to have drunk water from Broad Street — they preferred its taste, they went to school nearby, they visited from out of town, etc. Many who lived or worked nearby did *not* contract cholera. But they had also *not* drunk the pump's water: the workhouse had its own well, the workers at the brewery drank beer, etc. The apparent anomalies ultimately fit the pattern.

Farr and others were not fully convinced. They exhibited the skepticism often hailed as a hallmark of science. Snow had dramatically aligned all the cases with a common element. Yet other causes, such as local miasmas or soil chemistry, were still possible. Snow lacked effective controls.

Snow found an opportunity when cholera revisited London in 1854. By this time, buildings had their own plumbing. Pipes led to each house from different water companies. One could trace the source of the water. In one neighborhood affected by cholera, two water companies had competed for providing service. Individual homes, side by side — all sharing the *same* geographical features and *same* economic profile — had *different* water supplies. Here were the idealized conditions for a controlled study already in place! Today, the circumstances may highlight the meaning of 'control'. Many persons focus primarily on the *connotations* of the word "control" and thus (mistakenly) on the ability to manipulate variables in a lab. But the concept is fundamentally about comparison. The strategy is to isolate the causal effect of one variable using tandem observations. Philosopher John Stuart Mill had called it the "method of difference": consider conditions similar in all but one respect. Snow earns his renown for having discerned the relevant control in conditions that already existed: what is known as a *natural experiment*.

Snow did his best to identify the water company for each household. But the records were incomplete. He tried a water test (based on a chlorine precipitate — now known to have been unreliable due to tidal influxes of seawater). Snow then turned to the infection rates for the water companies as a whole. Here, he could not compare figures for just the district where cholera occurred. (Even the best studies may have limits!) But the numbers were nevertheless telling. Even though the ranges of the water companies did not fully coincide, the incidence of cholera among Southwark & Vauxhall's users was twenty times that of Lambert's. Snow's new findings were more persuasive (even while some colleagues cautiously remained open to evidence for other, perhaps supplemental factors). The evidence here was not just confirmatory. Alternative interpretations had been ruled out with just the right comparison, or *control*. The scope, detail and clarity with which Snow delineated the cause of cholera makes his work classic.

## Christiaan Eijkman and Beriberi

Experiments—even controlled experiments—have limits. Nowhere is this more evident than in Christiaan Eijkman's classic investigations of beriberi, a degenerative neurological disease prevalent in southeast Asia in the late 1800s. Eijkman shared the Nobel Prize in 1929 for the discovery of vitamins. Ironically, Eijkman at first rejected the very notion of vitamins and their role in causing beriberi. Why? How could Eijkman have contributed significantly to a discovery while misinterpreting the evidence?

Eijkman went to Java to study beriberi having just studied with Robert Koch, the pioneer of germ theory. Beriberi occurred frequently in prisons, insane asylums, military units and ship crews, strongly indicating contagion. Eijkman arrived, equipped with the new methods for isolating the suspected pathogen. Through a series of fortuitous accidents, Eijkman found that chickens suffered from a similar disease and that it was caused by a diet of polished (white) rice. Eijkman seemed to have pinpointed the source of the yet unidentified bacterium. Perhaps it entered the rice in the mills where the pericarp was abraded away? A diet of unpolished (red) rice, by contrast, offered a quick cure. The coating of the rice thus seemed to provide an anti-toxin or anti-bacterial agent. Not everyone accepted Eijkman's controlled experiments with chickens, comparing diets of polished and unpolished red rice, as a solution, however. Critics questioned whether the chicken's ailment was really beriberi. Assumptions about the model organism limited the claims.

Eijkman then turned to humans. He persuaded a prison where beriberi was found to change its diet from white to red rice. The beriberi decreased. But without a simultaneous control, one could easily imagine the epidemic ending for other reasons. (As critics did.) Eijkman then enlisted the support of the head of the Civil Health Department. As Snow did with cholera, they found a controlled experiment already in progress, in dozens of prisons across Java with different rice diets. The study was immense — over a quarter million prisoners (at least sample size was no limitation!). Even without proper statistical analysis, the figures were striking. Beriberi occurred among 1 in 39 prisoners with a diet of white rice, and 1 in 10,000 with red rice (and intermediate where the rice was mixed). The rigor of the study was reinforced by *supplemental* controls that aimed to exclude *other* possible, coincidental effects: ventilation (ruling out airborne germs), permeability of the floors (waterborne germs), age of the buildings, elevation and population density. None correlated with beriberi. The results inspired other researchers to conduct similar experiments. Ultimately, many institutions changed their diets and — due largely to Eijkman — beriberi decreased. Eijkman's investigations became classic.

Eijkman concluded his studies on beriberi still believing that bacteria in the rice caused it. After all, all his evidence (from a decade of study) fit this conclusion. But here Eijkman was wrong. His successor in Java, Gerrit Grijns, demonstrated that beriberi was a nutrient deficiency. Other exclusively starchy diets caused beriberi. Other foods, such as the mung bean, cured beriberi. Revised comparisons exposed new possibilities. Eijkman's experimental categories had overlapped with another distinction: the absence (or presence) of an essential nutrient. Eijkman's controlled study had yielded a positive result, of great social significance. Yet it was also limited. Other unforeseen interpretations had not been excluded. For Eijkman, the very notion of a vitamin

had been a conceptual blindspot, outside his theoretical perspective. That gap between finding a solution and not finding an ultimate solution is key. The nature of the control determines the reach of one's conclusions. Appreciating experiments includes understanding their limits as well as their import.

### **Charles Darwin and Seed Germination**

Finally, consider Charles Darwin's investigation of the effect of salt water on seed germination. Darwin recognized that his novel theory of evolution was vulnerable to several criticisms. One challenge was to explain biogeographical relationships, such as the similarity of flora and fauna between the Galápagos Islands and the South American mainland. Birds could fly and colonize such islands. But what about plants? If species separated by vast oceans shared a common ancestor, as he claimed, how did they disperse? Seeds might well be carried by ocean currents. But could seeds float and survive such exposure to sea water? Was the conceptual prospect empirically warranted?

On April 7, 1855, Darwin wrote his friend and colleague, botanist Joseph Hooker, about beginning some seed-salting experiments. He asked Hooker which seeds might be easily killed. Hooker apparently gave them all no more than a week. At two weeks, therefore, Darwin wrote again, proudly reporting "a nice little triumph" in their ongoing survival. After two months, the soaking experiments continued. Darwin confided that he was worried the project would "turn into another barnacle job," referring to the eight years he had spent classifying the group of mollusks. In several months, however, the work was done and the germination results reported in the *Gardener's Chronicle*. In all, Darwin had tested 87 kinds of seeds. 64 had sprouted after four weeks. Later, he examined the floatation of dried versus green branches. Using 94 types of plants, 18 stayed afloat. Surviving an ocean voyage thus seemed plausible for an estimated 14 per cent of seed plants.

Darwin's experiment may seem far less grand than the other classic experiments I've described. Indeed, the basic design is so simple that children (even at the elementary level) are sometimes invited to echo Darwin's work. But the term "classic" is still appropriate. Darwin had posed a significant question in terms of his theory. Hooker, at least, seemed not to grasp its relevance at first. Instead, he urged Darwin to expand his study and identify how groups of plants varied in their dispersal ability. Darwin wrote back that he only wished to demonstrate the possibility of sea transport. (Even later, Hooker seemed unimpressed by the biogeographical consequences of Darwin's results.) Darwin summarized his seed work again in the *Origin of Species* profiling its significance in explaining the geographical distribution of plants (Chapter XI, pp. 358-360). Using information on average ocean current speeds, he expressed his results specifically in terms of the 28 days needed to cross the Atlantic. Darwin knew precisely how his experiments fit in answering a particular theoretical question. Nor could one take the outcome of the experiment for granted. Given Hooker's initial response, one cannot assume that the results were obvious (as they seem now). It was a genuine "test," with the potential to fail. Darwin's "modest" study also had considerable scope, in terms of the numbers of plants he tested — unlike the classroom replications recommended now. Darwin further consulted colleagues for their experience. He recognized that secure explanations relied on consilience in a vast spectrum of evidence.

In celebrating classic experiments, such as Meselson and Stahl's, one might easily imagine that great theories stand or fall on the basis of a single, simple experiment. But theories are conceptual networks. They generally rely on a large suite of studies, each providing a relevant empirical benchmark. Darwin's "modest" study on seed germination was theoretically significant, illustrating that experiments — even classic experiments — earn their meaning in the context of other experiments and concepts.

### **Beyond A Few Classics**

Judging Meselson and Stahl's experiment, among all experiments in biology, as *the* most beautiful surely invokes personal perspective. Indeed, in 2003 the American Institute for Biological Sciences invited nominations of others' favorites (see the June, 2004 issue of *BioScience* for the results). The range of experiments to appreciate is hardly limited. Each seems to contain an instructive story. I hope that my examples indicate how not only classic experiments but virtually *any* historical experiment might be an occasion for learning more about science and the nature of science.

Many such experiments appear in introductory college textbooks: Mendel's study of inheritance in peas or Kettlewell's field work on selection in peppered moths. Some may describe Harvey on the circulation of the blood, Morgan on sex-linked inheritance, Hamner and Bonner on photoperiodism in cockleburrs, or Lorenz on imprinting in greylag geese, to name just a few. But the textbooks often cast these episodes as sidebars. The format itself may suggest to the naive reader that such experiments are peripheral to, rather than constitutive of science. Likewise, the typical brevity may promote an oversimplified caricature of science. The cases profiled above illustrate, I trust, the value of broadening such perspectives. Effective teachers, of course, regard textbooks as resources, not a final curriculum. With a dash of creativity and participation by students, and access to a good library or the internet, one can begin to recreate great historical episodes of science in the classroom. The science becomes more vivid and more human. At the same time, students deepen their skills in thinking experimentally and thinking critically about results. Biology comes alive by appreciating classic experiments.

### **Selected Bibliography**

*BioScience*. "Biology's Most Beautiful." (June, 2004 issue).

Carpenter, Kenneth. *Beriberi, White Rice and Vitamin B: A Disease, A Cause and a Cure*. Berkeley, CA: University of California Press, 2000.

Darwin, Charles. *On the Origin of Species*. London: John Murray, 1859. Available online (<http://www.literature.org/authors/darwin-charles/the-origin-of-species>).

Eyler, John. "The Changing Assessments of John Snow's and William Farr's Cholera Studies." *Präventivmedizin* 46(2001): 225-232. Available online (<http://www.ph.ucla.edu/epi/snow.html>).

Holmes, Frederic Lawrence. *Meselson, Stahl and the Replication of DNA: A History of the Most Beautiful Experiment in Biology*. New Haven, CT: Yale University Press, 2001.

Meselson, Matthew and Franklin W. Stahl. "The Replication of DNA in *Escherichia coli*." *Proceedings of the National Academy of Science U.S.A.* 44(1958): 671-682.

Snow, John. *On the Mode of Communication of Cholera*. London, 1855.  
Available online at The John Snow Site (<http://www.ph.ucla.edu/epi/snow.html>).

### **Classroom Case Studies**

The College Board. *The Spanish Flu and Its Legacy*. New York, NY: College Board Publications, 1999.

Hagen, Joel B., Douglas Allchin and Fred Singer. *Doing Biology*. Glenview IL: Harper Collins, 1996.

Allchin, D. "Of Rice and Men." SHiPS Resource Center, 2001. <http://ships.umn.edu/modules/eijkman1.htm>.

### **Web Excursions**

- ◆ Meselson-Stahl & DNA Replication  
Annotated animation of the experiment.  
<http://www.dnafb.org/dnafb/20/concept/index.html>, click on "Animation"
- ◆ John Snow & Cholera  
The complete original 1855 publication, maps, biography and more.  
<http://www.ph.ucla.edu/epi/snow.html>
- ◆ Christiaan Eijkman & Beriberi  
1929 Nobel citation.  
<http://www.nobel.se/medicine/laureates/1929/press.html>
- ◆ Charles Darwin & Seed Germination  
Seed activity [PBS "Evolution" series].  
<http://www.pbs.org/wgbh/evolution/educators/teachstuds/pdf/unit2.pdf>
- ◆ *For more links, visit AP Central:*  
<http://www.apcentral.com>

### *About the Author*

Douglas Allchin takes pride in having taught AP Biology, 1981-1985. He subsequently earned his M.S. in Evolutionary Biology and Ph.D. in Conceptual Foundations of Science, both from the University of Chicago. He continues historical work and philosophical analysis on error and disagreement in science, while championing the use of history and philosophy of science in science education. He edits the SHiPS Resource Center ([ships.umn.edu](http://ships.umn.edu)).