

A controlled experiment, everyone learns early in school, is the hallmark of good science. But what, exactly, is the hallmark of a controlled experiment?

For many people, a *controlled experiment* is, straightforwardly, one “controlled” by the scientist. If you can “control” the variables and produce a predicted result (based on your hypothesis), then you seem to have proved cause and effect. Scientific credibility seems to emerge from the power of demonstration and technology. However, this widespread view of the core meaning of *control* in science – this month’s Sacred Bovine – is mistaken (Boring, 1954, 1969). It thus fosters a misleading impression of what makes scientific claims reliable. Here, I journey through the history of the term *control* and show how the concept contributes to the distinctive nature of scientific reasoning.

○ A History of “Control”

The origin of the modern term *control* may seem surprising. Yet it is also informative. It comes from the French *contre-rolle*, translated as “counter-roll.” No one uses counter-rolls today. But historically (beginning at least by the 12th century), royal household expenditures were documented on paper rolls. Later, rolls recorded other business, trade, and tax accounts (Figure 1). In some cases, the transactions were recorded on *two* rolls simultaneously. The duplicate roll – the *counter-roll* – was kept independently by a trusted officer. The copy could later be consulted when auditing the accounts, allowing one to detect any illicit tampering with the original records. Basically, the counter-roll was a parallel reference copy for comparison, a method to identify and thereby limit errors. The method helped to guarantee reliability in accounting (see Hoskin & Macve, 1994, p. 76). Eventually, the person who managed the counter-roll became known as the *comptroller* – a term still used today for someone who manages and monitors the finances of a business or government body.

From the accounting context, the concept of regulating error was generalized. *Counter-roll* became contracted to *control*. In legal contexts, one person’s testimony could be used to check, or “control,” the testimony of another. Later, when Britain colonized India, they established a body in the late 1700s whose role was not to itself govern, but to oversee and to keep checks on how the East India Company operated. It was called the Board of Control. Again, the core concept of “control” was to guard against error through a second source.

The term entered science in the mid-nineteenth century. In Germany, it seems to have been used (although somewhat inconsistently) in agricultural field research (McManus, 2018). In France, veteran chemist Michel-Eugène Chevreul noted how simply observing a phenomenon was insufficient for making conclusions about what caused it. There were too many hidden uncertainties. One needed additional experiments to clearly demonstrate the cause. He cautioned:

Insofar as this cause has not been demonstrated true by a system of experiments, it is observation without control.

As an example, Chevreul cited Pascal’s 1648 effort to demonstrate the weight of air. One barometer was taken high up the Puy de Dôme, where it measured a thinner atmosphere and lower pressure. But was altitude the true cause? For comparison, a second barometer was kept all the while at the base of the mountain, where it showed no change. One needed to *compare* the *two measurements together* to validate the conclusion (Chevreul, 1850, pp. 73–74). For Chevreul, that extra observation, not obvious perhaps, was essential to ascertaining causes unambiguously.

In England, the first documented record of the term *control* (currently) seems to be in an 1873 letter from botanist and entomologist John Traherne Moggridge to none other than Charles Darwin. Moggridge was explaining his experimental work on seed germination – a topic that had engaged Darwin in the *Origin of Species* (1859, pp. 358–360). Having observed that harvester ants seemed able to suspend the germination of seeds, Moggridge was investigating the possible role of formic acid. Could it inhibit seeds from germinating? He was testing several different species. He wrote to Darwin:

Eleven tumblers were employed, ten containing acid or acid & water in the gallipot cover, & one, the control experiment, no acid.

Here, he used *control* in the sense of a counter-roll: as a parallel case for comparison. To test for the differential effect of just the acid, his control was simply “no acid.” (Note, too, how Moggridge used the term *control* as though Darwin was already familiar with it.) A month later Moggridge reported to Darwin again. In this letter he indicated more explicitly what he meant by “control”:

. . . the control sowing (that made to test the germinative power of the seeds, & in wh. no acid was used).



Figure 1. An early accounting roll (photo by the author).

Testing untreated seeds was just as important as testing treated seeds: to ensure that one did not unwittingly mistake the cause.

Later that summer the term *control* began to appear in Darwin's own correspondence. In a note to Joseph Hooker, he continued an ongoing discussion on insectivorous plants. Did the leaf glands of his tropical pitcher plant secrete the acidic fluid that collected at the bottom of the tube? Darwin sketched a possible experiment. First, wipe the leaf surfaces clean. Then try to elicit a secretory response by applying a small sample of fibrin – an animal extract that Darwin had used successfully to elicit a digestive response in another carnivorous plant, the sundew (*Drosera*). But then Darwin added:

As a control experiment you could stick in a bit of equally damp cotton or moss on another point.

That is, Darwin indicated that the mere physical irritation of the leaf's gland might produce the fluid, rather than the chemical nature of a particular meat-like substance. One needed to rule out that possibility. The cotton was Darwin's suggested counter-roll: a way to check against an erroneous conclusion.

By 1875, Darwin had incorporated the terminology of “control” into his published work, in his book on *Insectivorous Plants*. In 1880, working with his son Francis, he continued to describe the various controls in his research, now on *The Power of Movement in Plants*. But the language had shifted subtly. In several places, the Darwins inserted a parenthetical phrase, as though to define *control* for a reader who may not yet have been familiar with the term. They repeatedly referred to their controls as “standards of comparison” (Darwin & Darwin, 1880, pp. 162, 163, 186, 525). And so they were. Controls are counter-rolls. But the Darwins' explanatory asides are valuable to us as a rare snapshot of a language evolving, at a moment when a new term has not yet made its way into common usage and so needs to be explicitly defined.

Many years later, the *Oxford English Dictionary* (OED) would cite *Insectivorous Plants* when it identified the earliest use of this meaning

of *control* in print. By that time, the definition entry was able to further characterize the meaning of *control* by referring to the structure of its experimental logic. A scientific “control,” it noted, embodied philosopher John Stuart Mill's “method of difference”:

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, of an indispensable part of the cause, of the phenomenon. (Mill, 1874, p. 280)

Such is a clear and concise expression of our modern concept of control – taught in virtually every science class. It is the scientific descendant of accounting's counter-roll.

The various experiments from Darwin's correspondence and publications may not seem particularly important. They are not monumental in scale. Nor do they seem to resolve great theoretical questions. They seem ordinary. Mundane. Even trivial. Yet that is precisely why they are significant. They indicate how this mode of comparative reasoning permeates all of science.

Namely, it is not enough to do an “experiment” – to intervene in nature – and observe what happens, as is widely believed in popular culture (again, this month's Sacred Bovine). Rather, to interpret causes and associative patterns effectively, one must consider differences. One needs multiple observations, parallel in all respects save one parameter (as Mill noted), to identify the key causal factors. To coin a brief motto, “Pair and compare.” That is the important modern lesson about the nature of science hidden in the medieval counter-roll.

○ Control without “Control”

The history of the term *control* helps convey the meaning of the concept (especially in contrast to popular misconceptions). At the same time, it hardly constitutes a comprehensive history of scientific *practice*. While the *OED* properly referred to Mill's logical structure, his principle did not (ironically) contribute directly to the origin or development of the concept of scientific control. Indeed, controlled experiments (by today's nomenclature) were common well before the term emerged to identify them as such. Darwin, for example, discussed “standards of comparison” with his colleagues decades before he came to use the term *control*. One can have a control (the counter-roll comparison) even without “control” (as a label). One thus finds renowned cases of controlled experiments scattered through history – many already familiar to biology teachers, and all ripe for student-centered inquiry activities.

One of the most celebrated cases (frequently recounted in biology textbooks) is Francisco Redi's *Experiments on the Generation of Insects* (1668). Redi addressed the then common belief that insects arose from decaying matter: “Hey, don't believe me! Do an *experiment* for yourself. Just leave out some raw meat or stale bread, and *observe*. In a few days, maggots will appear, as the foodstuffs vanish. Flies will swarm forth. The food transforms into new life—*spontaneously*. You can *trust what you see with your own eyes*.” Incredible, yes? Of course, with our modern knowledge of insects and eggs (and microbes), students readily see the flaws in such a “demonstration.” But the challenge is to craft the relevant evidence that exposes the

hidden cause. That was Redi's achievement. He compared open jars with gauze-covered jars. When insects could not access the meat, no life appeared. Error averted. That's the nub of science in this case: using a "counter-roll" for comparison and showing that an apparently plausible interpretation was, in fact, mistaken.

Another popular case is James Lind's (1753) experiments on scurvy (Carpenter, 1986; Brown, 2005). In the eighteenth century, sailors suffered from the debilitating disease of scurvy. Lind, a surgeon for Britain's navy, was aware that citrus fruits were reported as a remedy. For example, in 1534 Jacques Cartier had recorded that crew members on his voyage recovered from scurvy using the Native American custom of drinking a juniper-berry tea (Leland, 2007). But other treatments were proposed as well. Unambiguous evidence was missing. Lind sorted 12 ailing sailors into six pairs. While one group received two oranges and a lemon per day, the others received (variously) sea water, fermented cider, sulfuric acid, vinegar (also a sour acid), or a concoction of mixed herbs. Only the citrus fruits proved effective. Again, the alternative treatments were just as important as the successful treatment to justifying the conclusion. Through comparison, they showed that other factors (such as a merely acidic, fruity, or salty beverage) were not integral to the cure. Note again how the process of ruling out error was critical to the scientific conclusions.

Another engaging case is Jan Ingenhousz's (1779) experiments on photosynthesis (Nash, 1957; Magiels, 2010). This episode was, in many ways, a comedy of errors (*Sacred Bovines*, Sept. 2012). Joseph Priestley had first discovered that plants could "restore" the air fouled by animal respiration or combustion. But other investigators could not reproduce his results. Later, even Priestley's own replications failed. Priestley tinkered with one variable after another, leading him to conclude (ironically) that light and water alone were responsible, not plants. Ingenhousz then sorted it all out. Through many successive experimental "counter-rolls" (adding and subtracting various factors in parallel trials), he clarified Priestley's errors. Notably, Priestley had not recognized the role of microscopic green algae in his pump water. Plants were indeed important. Ingenhousz also confirmed the essential role of light. Originally, Priestley had failed to identify the relevance of a nearby window in his lab. Ultimately, Ingenhousz's discovery relied substantially on reasoning from the negative results of several experiments. Paradoxically, perhaps, negative findings can be an essential part of positive knowledge.

None of these great experimenters characterized their work in terms of "controls." But each clearly used the principle of the counter-roll. They discredited possible, but erroneous, conclusions by *comparing* results from parallel experiments. The comparison, not the lone demonstration, was key.

Teachers not uncommonly use these cases to illustrate experimental design. As the history of control suggests, however, students need to also develop a full understanding of the very *meaning* of control and the significance of "counter-rolls" and comparisons in scientific reasoning.

○ "Control" without Experiment

One might well imagine that conditions for an experimental "control" (reasoning from a metaphorical counter-roll) require actually exercising material control over the variables. That might explain,

in part, the popular misconceptions. John Stuart Mill certainly believed that applying the method of difference was congruent with "a method of artificial experiment." Nature is complex and it causes obscure, he contended. "It is very seldom that nature affords two instances, of which we can be assured that they stand in precise relation to one another" (Mill, 1874, p. 281). So, in the end, does "control" eventually reduce to the experimenter's power to control? Can one have "control" *without experiment*? Yes, such occasions do arise. They are called *natural experiments* (although without intervention one might well not call them "experiments" at all).

One stunning case emerged while tracking the cause of beriberi in Java in the 1890s (Allchin, 1996). In laboratory experiments, Dutch physician Christian Eijkman was able to show that chickens fed a diet of white rice fell ill. When fed unpolished, whole rice, they did not. But did these results apply to humans? Might the results have been a coincidence of local conditions? Eijkman enlisted Hans Vorderman, the Director of Public Health on Java. There, prisons varied in their rice diets. There were 100 prisons, housing nearly a quarter-million prisoners – quite a sample size. Data on the relevant comparison already existed. Vorderman collected the statistics on rice diet and incidence of beriberi for each prison. Sure enough, comparing the two forms of the same rice revealed a marked distinction. Yet the prevalence of beriberi in institutions (like prisons, navies, and insane asylums) still hinted at contagion by a microbe. So Vorderman collected further data: on the prisons' conditions relevant to possible modes of transmission. Were they densely populated? Were they well ventilated? Were the floors permeable? Were the buildings old? If these aligned with the incidence of the disease, then the conclusion about diet would be suspect. But none related to the frequency of beriberi. Rice diet was responsible. Soon, shifts in rice diet in institutions throughout Southeast Asia helped reduce the prevalence of beriberi. Unlike Lind, Vorderman did not create two groups with assigned diets. Rather, he found precisely the relevant data for comparison without such intervention. Yet his reasoning, checking against possible error from alternative explanations, was similar, and just as sound.

Another classic case of a natural experiment is John Snow's study of cholera in London in the 1850s (Snow, 1855; Johnson, 2006). Based on symptoms, Snow inferred that the disease spread through contaminated water. When an epidemic broke out in the Broad Street area, he immediately suspected the local pump. His research focused on who had drunk that water and, equally, who had not. Proximity to the pump was the strongest indicator. But there were many exceptions and it was incumbent on Snow to explain why. Through door-to-door interviews he showed that many locals who escaped cholera had relied on a different source of water. For those who lived farther away, he established that many had visited the area or sent for the water specially. In this way, Snow was able to show that in nearly all cases victims had drunk water from the now notorious Broad Street pump *while non-victims had not*. That *comparison* (including resolving the confusion of apparent exceptions) was the persuasive evidence. In a subsequent outbreak, the source of water was determined by the new water companies, in some cases with separate pipes leading down the very same street. That allowed Snow to eliminate local miasmas (foul airs) as a possible explanation. No experiment. But thorough, elegant "controlled" reasoning. The strategy of the counter-roll does not depend on setting up and

conducting a novel lab or field experiment. It is all about the comparison and regulating the possible errors in inference.

○ From Experiment to “Control”

Just as one can have “control” without experiments, one can have an experiment without “control.” That is, a scientist can explore or test various conditions by freely manipulating the variables. Those investigations can yield important discoveries – for example, by producing phenomena that were wholly unexpected or beyond current theoretical understanding. But that does not make them “controlled.” Not every experiment is a *controlled experiment*. Without a second test to check, the conclusions remain open to alternative explanation and possible error. What matters is the *comparison*, the *counter-roll*. As illustrated in the many cases above, identifying and ruling out possible sources of error is central to the reliability of science and thus to its public credibility. And that is why, ultimately, the unassuming counter-roll found a place in science, as the hallmark of controlled experiments – and of good science.

○ Acknowledgments

My deep appreciation to Janet Browne and Paul White for their assistance tracking documents and the biographical details surrounding Darwin’s use of controls.

References

- Allchin, D. (1996). Christian Eijkman and the cause of beriberi. In J. Hagen, D. Allchin & F. Singer, *Doing Biology* (pp. 116–127). Glenview, IL: HarperCollins. <http://shipseducation.net/modules/biol/beriberi.htm>.
- Boring, E.G. (1954). The nature and history of experimental control. *American Journal of Psychology*, 67, 573–589.
- Boring, E.G. (1969). Perspective: artifact and control. In R. Rosenthal & R.L. Rownow (Eds.), *Artifact in Behavioral Research* (pp. 1–11). New York, NY: Academic Press.
- Brown, S.R. (2005). *Scurvy: How a Surgeon, a Mariner, and a Gentleman Solved the Greatest Medical Mystery of the Age of Sail*. New York, NY: St. Martin’s Press:

- Carpenter, K.J. (1986). *The History of Scurvy and Vitamin C*. Cambridge, UK: Cambridge University Press.
- Chevreul, M.-E. (1850). Historie de la chimie. 8e article. *Journal des Savants*, 1850, 71–79.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, Or, the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.
- Darwin, C. & Darwin, F. (1880). *The Power of Movement in Plants*. London: John Murray.
- Hoskin, K. & Macve, R. (1994). Writing, examining, disciplining and the genesis of accounting’s modern power. In A.G. Hopwood and P. Miller (Eds.), *Accounting as Social and Institutional Practice* (pp. 67–97). Cambridge, UK: Cambridge University Press.
- Ingenhousz, J. (1779). *Experiments upon Vegetables*. London: P. Elmsley and H. Payne. Facsimile edition available online at https://archive.org/details/b30516638_0001.
- Johnson, S. (2006). *The Ghost Map*. New York, NY: Riverhead Books.
- Leland, T. (2007). Interpreting Native American herbal remedies. St. Paul, MN: SHiPS Resource Center. <http://shipseducation.net/modules/biol/native-herb.htm>.
- Lind, J. (1753). *A Treatise on Scurvy, in Three Parts*. Edinburgh: printed by Sands, Murray, and Cochran. Facsimile edition available online at <https://archive.org/details/b30507054>.
- Magiels, G. (2010). *From Sunlight to Insight: Jan Ingenhousz, the Discovery of Photosynthesis & Science in the Light of Ecology*. Brussels: VUB Press.
- McManus, A. (2018). Jutta Schickore’s “Contributions to a history of experimental controls” [blog]. <https://jhiblog.org/2018/03/28/dispatches-from-princetons-history-of-science-colloquium-jutta-schickores-contributions-to-a-history-of-experimental-controls>.
- Mill, J.S. (1874). *A System of Logic*. New York, NY: Harper & Brothers.
- Nash, L.K. (1957). Plants and the atmosphere. In J.B. Conant and L.K. Nash (Eds.), *Harvard Case Histories in Experimental Science, vol. 2* (pp. 323–435). Cambridge, MA: Harvard University Press.
- Redi, F. (1668/1909). *Experiments on the Generation of Insects* (trans. Mab Bigelow). Chicago, IL: Open Court. <https://archive.org/details/experimentsongen00redi>.
- Snow, J. (1855). *On the Mode of Communication of Cholera, 2nd ed.* London: John Churchill. Facsimile edition available online at <https://archive.org/details/b28985266>.