

---

# From Deception Trials to Control Reagents

---

## *The Introduction of the Control Group About a Century Ago*

---

Trudy Dehue  
University of Groningen

*This is the story of the remarkable psychologist John E. Coover, who, in the early 1900s, was the first to advocate the comparison of experimental and control groups as a methodological necessity. Moreover, the author raises the issue of why control groups were launched about a century ago, and why psychology was comparatively early in codifying group comparison as a methodological routine. In dealing with these questions, the author discusses the relations between turn-of-the-century science and society as well as between psychophysical research and educational experimentation. Furthermore, the mystery is solved of how Coover's rightful place in the received history of experimental controls could be taken by precisely the authors whom he criticized for the lack of controls.*

In 1872, Francis Galton, the British statistician and biometrician, published an article entitled "Statistical Inquiries Into the Efficacy of Prayer" (Galton, 1872). Arguing that the profit of piety should be assessed on the basis of its earthly revenues, Galton's article was a remarkable contribution to 19th-century science. So it was, however, for yet another reason. In making his case, Galton introduced a new research strategy. As he explained:

The principles are broad and simple. We must gather cases for statistical comparison, in which the same object is keenly pursued by two classes similar in their physical but opposite in their spiritual state; the one class being spiritual, the other materialistic. Prudent pious people must be compared with prudent materialistic people and not with the imprudent nor the vicious . . . We simply look for the final result—whether those who pray attain their objects more frequently than those who do not pray, but who live in all other respects under similar conditions. (Galton, 1872, p. 126)

Galton thus explicated a methodological design that was to become common practice in psychology and many other disciplines.<sup>1</sup> Nowadays, many researchers consider it to be evident that the evaluation of actions should best be done by comparing groups equal in all respects but the action concerned. Moreover, rather than drawing conclusions from post hoc comparison, as in Galton's example, ideally researchers themselves assign subjects to experimental and control groups.

In Galton's times, however, the method of comparing experimental and control groups did not yet belong to the codified routines. Historians have traced but a few incidental pre-20th-century examples of experiments with culti-

vated plants, laboratory animals, or diseased humans (Cochrane, 1976; Coleman, 1987; Marks, 1997). The cases with human subjects stem from medicine. Not a single case of experimental versus control group comparison with humans was found in psychology or social science (only some comparing individuals with themselves as in Triplett, 1898). In a historical introduction, the experimental psychologist Richard L. Solomon noted that "We were struck with the fact that the use of an experimental and a control group . . . is a very recent development" (Solomon, 1949, p. 137).

Solomon found the earliest instance in experiments published in 1901 by Thorndike and Woodworth (1901a, 1901b, 1901c). However, the authoritative historian of psychology Edwin G. Boring doubted that these experiments were a proper example. Thorndike and Woodworth did not look upon group comparison as a methodological necessity. They only casually referred to a single trial with a control group. Moreover, they did so after having discussed a range of experiments without controls (Boring, 1954, p. 584). Ever since Boring's reconstruction, historians of psychology have ascribed the first introduction of control groups to Solomon's next cited case of a 1908 British educational experiment. The earliest use of the word *control* in relation to human group comparison is located in American experimental psychology of the mid-1920s (Solomon, 1949, p. 138).

In the present article I amend this received history of the control group. I recount the use of "experimental and control reagents" in early 20th-century educational experiments by the Stanford psychologist John Edgar Coover. In

---

*Editor's note.* This article is part of the series *Snapshots of Psychology Circa 1900*, developed by Donald A. Dewsbury.

*Author's note.* I thank Donald A. Dewsbury, Rand B. Evans, and Laurence D. Smith for their helpful comments on a previous version of this article and Paul C. Nascimento for his help in checking unpublished manuscripts in the Rare Book and Manuscript Library of Columbia University, New York.

Correspondence concerning this article should be addressed to Trudy Dehue, Heymans Institute for Psychological Research, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands. Electronic mail may be sent to g.c.g.dehue@ppsw.rug.nl.

<sup>1</sup> In fact, research on the efficacy of praying has been widely conducted in the 1980s and 1990s. Comparing equal groups proved to be not as simple as Galton imagined it (Sloane, Bagiella, & Powell, 1999).

a 1907 article (Coover & Angell, 1907), Coover advocated the design as a methodological necessity and as a critique of Thorndike and Woodworth (1901a, 1901b, 1901c). How did Coover fall into historical oblivion? Further, how could Thorndike and Woodworth become the avowed inventors of a procedure that according to contemporaries was in fact *lacking* in their work? At the end of this article, I present some grounded surmises on these issues. However, my main intent is not to rehabilitate Coover but to demonstrate that the emergence of experimental group comparison in psychology only seems “strikingly late” when one ascribes 20th-century aspirations to 19th-century practices. Furthermore, I discuss why psychology, compared for instance with medicine (Marks, 1997), economy (Morgan, 1990), or sociology and political science (Platt, 1996; Ross, 1991), was early rather than late in codifying group comparison as a methodological routine. To achieve these aims, some interrelated changes in science and society must first be discussed.

### From Aristocratic to Bureaucratic Social Policy and Social Knowledge

In the 19th century, social change and the creation of scientific knowledge were largely two separate spheres. Amelioration was predominantly a matter of local philanthropy and the production of social knowledge, an affair of scholarly gentlemen. Nineteenth-century statisticians generally focused on population universalities rather than human diversity. By calculating the averages of repeated measurements, they established the typical cases as intended by God or nature (Porter, 1986; Desrosières, 1998).

In turn-of-the-century Western countries, however, the common ideology of *laissez-faire* politics was countered with that of planned social change. Local charity was gradually replaced by centralized government. This process was to induce significant changes in both the *relation* between social policy and social knowledge-making and the *kind* of social knowledge produced.

Whereas the authority of 19th-century local officials depended on their personal trustworthiness, the rising class of bureaucratic officials had to justify its plans in terms of impersonal standards, procedures, and data. Thus, a large-scale alliance developed between social government and social research. Social scientists began to focus on population varieties rather than communalities and to produce technically useful data based on standardized methodological procedures. Heralded by Francis Galton, the former intellectual statistician was turned into a government statistician.

Under these new circumstances, a widespread demand rose for impersonal assessment of the efficacy of administrative actions. In the 1910s, the American engineer and sociologist F. S. Chapin argued that the differences made by social interventions should also be scientifically established (Chapin, 1917). Chapin conceded that he did not yet know how the effects of administrative policies could be assessed without arbitrariness. Simple before and after measurement would not do, he reasoned, as many other

events over and beyond the policy in question may influence the results.

In the meantime, however, American educational psychologists were dealing with this problem by comparing experimental and control groups. Coover, who first advocated this comparative group design, was a canny methodologist indeed. Yet, the design was not an instantaneous creation by a single individual. Just like conventions in society at large, the new research methodology developed gradually from preceding traditions.

### Control Trials in 19th-Century Psychophysics

Psychophysical experimentation began as a typical endeavor of mid-19th-century scholars interested in discovering the given design of human nature. In an attempt to understand the relationship between physical stimuli and mental sensations, they tried to establish what was the tiniest difference between two physical stimuli that could still be perceived. Experimenters lifted unequal weights or pricked their skin with two pins at varying distances to find the threshold of perception below which differences in weights or distances are no longer noticeable. The averages of series of measurements were taken as representations of general human functioning (Boring, 1942; Danziger, 1990).

The first psychophysical researchers often acted as their own subjects. However, from the 1870s pressure-point experimenters argued that awareness of the pins' distances in previous trials might give rise to expectations on the trials to come, which might bias perception. Moreover, it was argued, the sheer knowledge that there are two pins might incite the idea of perceiving a distance even if the actual distance is below the threshold of perception. For these reasons, it was decided that it takes two to experiment. An experimenter was to provide the stimuli to an “observer” in random orders as well as to randomly insert control or “deception” trials in which the observer's skin was secretly touched at one point only (this was the earliest use of randomization for experimental purposes; for a more extensive discussion of the present section, see Dehue, 1997).

Let the original stimuli be indicated as “S,” the responses as “R,” and the randomly applied control trials (in which the actual double stimulus was withheld) as “noS” Then the scheme of these experiments was something like S-R-noS-R-S-R-S-R-noS-R-S-R-noS-R-S-R-S-R (etc.). In later educational experiments, the randomly inserted control trials (noS-R) would be conducted with separate control groups. To understand how this transformation took place yet another aspect of psychophysical experimentation needs to be taken into account.

### Control Groups in Educational Psychology

After many trials, the skill of the experimenters improved, which artificially lowered the threshold of perception. Switching to the other arm or leg did not solve the problem,

because the undesirable practice seemed to transfer automatically to other parts of the body. This "transfer-of-training" phenomenon not only annoyed but also intrigued the researchers. Many of them became so fascinated that from a source of bias, transfer of training became a research topic of its own. At the end of the century, experiments were being conducted in European and American laboratories to establish transfer of training in all kinds of cognitive, sensory, and motor skills (Bray, 1928).

In these experiments, the established experimental scheme of S-R-noS-R-S-R-S-R-noS-R-S-R-noS-R-S-R-S-R had to be adapted. As trainings cannot be haphazardly undone, the usual design in which the stimulus was randomly given or withheld was no longer feasible. In the transfer-of-training experiments, first a response or ability was tested without preceding stimulus or training (*noS-R*), next the training was given (S), and subsequently the ability was tested again (R). The original scheme was limited to the particular "cutout" of (*noS-*) R-S-R.

Transfer experiments became an important resource in psychology's turn to the production of administrative knowledge and the nearly simultaneous introduction of the control group. At the turn of the century a debate blew up within the newly established circuit of educational administrators on the utility of teaching subjects such as Latin and formal mathematics. School administrators wanted to abolish such redundant subjects from the school curriculum, whereas proponents argued that "formal discipline" strengthens general mental capacities. This issue reminded some American psychologists of the psychophysical transfer phenomenon. The 1901 experiments by Thorndike and Woodworth (mentioned in Solomon, 1949, as the earliest instance of a control group) provide an early example of the use of psychophysical transfer experiments in the context of educational debates. Thorndike and Woodworth tested the educational profit of formal discipline in experiments according to the just-mentioned scheme of (*noS-*) R-S-R.

In Thorndike and Woodworth's (1901a, 1901b, 1901c) experiments the existence of transfer was *not* confirmed. For decades their findings were cited as the refutation of "a myth" and "a veritable bombshell into the educational camp" (Kolesnik, 1958, p. 7 and p. 34). Yet, soon after publication the experiments were severely criticized: The educational psychologist Coover from Stanford University rejected them for (among other things) lack of a proper control.

Coover was a student of the psychophysical experimenter Frank Angell, who gained his doctorate with Wilhelm Wundt at Leipzig in 1892. As a former school principal, Coover soon gave Angell's academic interests an instrumental twist. He took part in the debate on formal discipline, which in 1905 resulted in his master's thesis. In a 1907 article on the topic coauthored by Angell, Coover discussed Thorndike and Woodworth's studies as "rough experiments" of "very little value" (Coover & Angell, 1907, p. 330) and supported the notion of formal discipline. Moreover, the article amply explained why Coover had worked with not only experimental groups but also control groups. With a one-group design (R-S-R), he argued, one

cannot know for sure whether an improvement is indeed caused by the training. Results may be due to any factor operating between both tests or even to the first test, which may also act as a training:

In order to estimate the possible practice effect of the preliminary test and the interval between that and the after-test, a control experiment was carried along. In this three reagents were given tests . . . under conditions identical with the regular reagents. (Coover & Angell, 1907, p. 332)<sup>2</sup>

The idea, repeated throughout the article, was that alternative explanations of an effect can be ruled out only by comparing an experimental group with a control group that equals the experimental group in all respects but the training. With this comparative design the usual sequence of *noS-R-S-R* (i.e., a test without a training, the training, and next a test after training) was cut in two. The control group underwent the first part of *noS-R*, the experimental group the other half of S-R. In other words, the classical deception or control trials in which the stimulus was withheld were now applied to separate control groups.

## Epilogue: Coover's Unacknowledged Impact

Control groups emerged only after the transformation of social science from a largely intellectual to a regulatory undertaking. The psychophysical roots of experimental psychology help to explain why the use of controls occurred comparatively early in psychology. Of old, methodological scrutiny was the hallmark of psychophysics, and psychophysical transfer research shared with educational experimentation (and much of later social research) the characteristic that actions are evaluated that cannot be haphazardly undone.

Moreover, the volunteers in laboratory research could comparatively easily be kept under control, as could the schoolchildren who subsequently became the principal subjects of educational experimentation. When it came to the evaluation of other administrative actions, not every social researcher was convinced that group comparison offered a feasible solution. Chapin, who kept looking for impersonal ways of assessing social legislations, never accepted it. In his view, humanitarian rather than scientific reasons should determine who is subjected to—or excluded from—particular social actions (Chapin, 1947). Other social scientists, however, such as the powerful political scientist Charles E. Merriam, attracted psychologists as methodological advisors (Ross, 1991) and used control groups in their own research projects (Dehue, 1997).

Space limits prohibit a further discussion of the ins and outs of the control group design (Dehue, 2000). The present article ends with the question of how Coover and Angell's article (1907) could have been omitted from the historical record. Thorndike and Woodworth's 1901 exper-

<sup>2</sup> Most likely, the term *reagent* was borrowed from Wundt, who sometimes used *Der Reagent* for the person reacting in an experiment (K. Danziger, personal communication, October 3, 1999). The later term *subject* reflects psychology's administrative turn (Danziger, 1990).

iments (1901a, 1901b, 1901c) indeed were an important preliminary step toward a new kind of psychological experimentation, and they did informally refer to a single trial with group comparison (Thorndike & Woodworth, 1901c, p. 558). However, it clearly was Coover who first advocated group comparison as an imperative methodological condition. How, then, could it happen that his rightful place in the annals was overlooked?

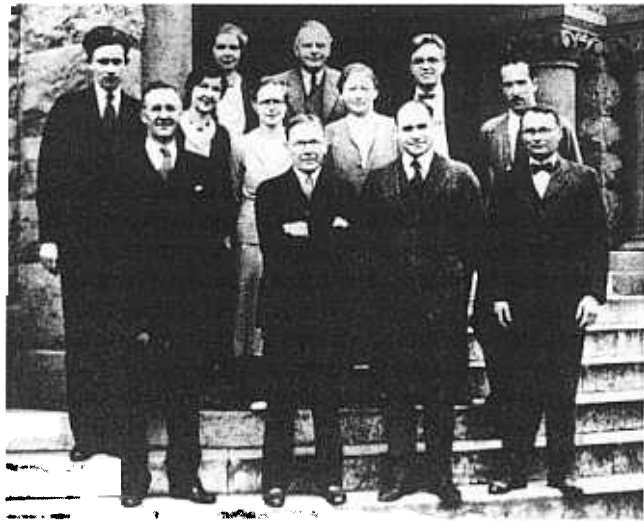
Solomon's article (1949) provides an intriguing clue. In his historical overview, he refers to some pages in Woodworth's renowned 1938 textbook *Experimental Psychology* (Woodworth, 1938). In these pages, Woodworth appears to advance group comparison in transfer research and to ascribe the historical introduction of the idea to the same authors as Solomon did (Woodworth, 1938, p. 178–181). Solomon seems to have borrowed his account directly from Woodworth!<sup>3</sup> Did Coover and his work never come to Woodworth's attention, or had Woodworth forgotten about the incident once he began writing *Experimental Psychology*? Even if this story took place in the present time, conclusive proof would be hard to find. Yet, there is circumstantial evidence.

To begin with, Coover was not particularly a man without a reputation. After completing a doctorate on formal discipline in 1912 (Coover, 1912), he received enormous sums of money from an endowment meant for psychological research. Both the story of this donation and Coover's methodologically astute experiments in parapsychology set many tongues wagging (Dommeyer, 1975; Mauskopf & McVaugh, 1980, pp. 253–254 and p. 285). Coover also figures in quite a few histories because of his extraordinary personality. His Stanford colleague Quinn McNemar remembered him as "owl-wise Coover" who would help with the most complicated logical problems provided one had the patience to listen to his endless elaborations. Lewis Terman, however, Angell's successor as the chair of the psychology department (see Figure 1) and a Thorndike and Woodworth ally, considered Coover as only "an irritant" and an "Angell holdover" whom he could not dismiss (McNemar, 1980, p. 316–317; Minton, 1988, p. 135).

If Woodworth hardly could have overlooked Coover, could he still have missed the 1907 article (Coover & Angell, 1907) or forgotten about it once he began to work on *Experimental Psychology* (1938)? Severe critique in a prominent journal is not likely to escape the attention of those targeted by it. Ignoring the issue of control groups, Ladd and Woodworth (1911, p. 569) very briefly mention Coover and Angell's article as a *confirmation* of the findings in Thorndike and Woodworth's 1901 experiments (which may be explained by Thorndike's argument that Coover and Angell misinterpreted their results; Thorndike, 1913, p. 408). Moreover, mimeographed drafts of *Experimental Psychology* were already available beginning in 1909 (Winston, 1990). In the preserved 1913 (Woodworth, 1913) and 1932 (Woodworth & Poffenberger, 1932) versions there is no mention of Coover and Angell nor, for that matter, of the indispensability of control groups or their early introduction by Thorndike and Woodworth.

**Figure 1**

Members of the Department of Psychology at Stanford University in 1932



Note. Quinn McNemar (furthest to the left), Lewis Terman (front, arms crossed), and John Coover (furthest to the right). Reprinted by permission of the Stanford University Archives, Department of Special Collections, Stanford University Libraries.

As it seems, Woodworth did not see a need for group comparison before the late 1930s, and when he did, he was well aware that Coover was the first one to discuss the procedure as a vital condition. Elsewhere in the 1938 version of *Experimental Psychology* (Woodworth, 1938), Woodworth overtly expresses his annoyance with Coover's 1912 doctoral thesis on formal discipline that also reports of experiments with control groups: "Do these results indicate transfer? . . . The results have been sometimes cited as evidence for positive transfer effect . . . but the reader must use his own judgement" (Woodworth, 1938, p. 198).

To all appearances, Woodworth deliberately kept the attack under wraps. Mentioning Coover in relation to the indispensability and history of control groups would have meant giving credit to an irritating man and moreover drawing attention to the lack of controls in experiments that were pivotal in establishing Woodworth's own career. One cannot deny the knack of his move to neglect Coover and simultaneously proclaim the criticized experiments as the introduction of the control group. Because Woodworth's "Columbia Bible" passed on this methodological requirement to generations of psychologists in America and abroad (Winston, 1990), Coover's story clearly demon-

<sup>3</sup> In fact, Thorndike, in his doctoral dissertation on experimental animal psychology, already described an occasional experiment in which animals are compared (Thorndike, 1898). Woodworth, in his history of experimental control (Woodworth, 1938), does not refer to this dissertation as an instance of group comparison without use of the word *control*, and neither do Solomon (1949) and Boring (1954).

strates that lack of recognition does not always testify to little impact.

## REFERENCES

- Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century-Crofts.
- Boring, E. G. (1954). The nature and history of experimental control. *American Journal of Psychology*, *67*, 573–589.
- Bray, C. W. (1928). Transfer of learning. *Journal of Experimental Psychology*, *11*, 443–467.
- Chapin, F. S. (1917). The experimental method and sociology. *Scientific Monthly*, *4*, 133–144.
- Chapin, F. S. (1947). *Experimental designs in social research*. New York: Harper.
- Cochrane, W. G. (1976). Early development of techniques in comparative experimentation. In D. B. Owen (Ed.), *On the history of statistics and probability* (pp. 1–26). New York: Dekker.
- Coleman, W. (1987). Experimental physiology and statistical inference: The therapeutic trial in nineteenth-century Germany. In L. Krüger, G. Gigerenzer, & M. S. Morgan (Eds.), *The probabilistic revolution* (Vol. 2, pp. 201–229). Cambridge, MA: MIT Press.
- Coover, J. E. (1912). *Formal discipline from the standpoint of experimental psychology*. Princeton, NJ: Princeton University Press.
- Coover, J. E., & Angell, F. (1907). General practice effect of special exercise. *American Journal of Psychology*, *18*, 328–340.
- Danziger, K. (1990). *Constructing the subject*. Cambridge, England: Cambridge University Press.
- Dehue, T. (1997). Deception, efficiency, and random groups. *Isis*, *88*, 653–673.
- Dehue, T. (2000). *Establishing the experimenting society*. Manuscript in preparation, University of Groningen, Groningen, the Netherlands.
- Desrosières, A. (1998). *The politics of large numbers*. Cambridge, MA: Harvard University Press.
- Dommeyer, F. C. (1975). Psychical research at Stanford University. *Journal of Parapsychology*, *39*, 173–205.
- Galton, F. (1872). Statistical inquiries into the efficacy of prayer. *Fortnightly Review*, *12*, 124–135.
- Kolesnik, W. (1958). *Mental discipline in modern education*. Madison, WI: University of Wisconsin Press.
- Ladd, G. T., & Woodworth, R. S. (1911). *Elements of physiological psychology*. New York: Scribner. (Original work published 1887)
- Marks, H. M. (1997). *The progress of experiment: Science and therapeutic reform in the United States, 1900–1990*. Cambridge, England: Cambridge University Press.
- Mauskopf, S. M., & McVaugh, M. R. (1980). *The elusive science*. Baltimore: Johns Hopkins University Press.
- McNemar, Q. (1980). Autobiography. In G. Lindzey (Ed.), *A history of psychology in autobiography* (Vol. 7, pp. 305–335). San Francisco: Freeman.
- Minton, H. L. (1988). *Lewis M. Terman: Pioneer in psychological testing*. New York: New York University Press.
- Morgan, M. S. (1990). *The history of econometric ideas*. Cambridge, England: Cambridge University Press.
- Platt, J. (1996). *A history of sociological research methods in America, 1920–1960*. Cambridge, England: Cambridge University Press.
- Porter, T. M. (1986). *The rise of statistical thinking*. Princeton, NJ: Princeton University Press.
- Ross, D. (1991). *The origins of American social science*. Cambridge, England: Cambridge University Press.
- Sloane, R. P., Bagiella, E., & Powell, T. (1999, February 20). Religion, spirituality, and medicine. *The Lancet*, *353*, 664–667.
- Solomon, R. L. (1949). An extension of the control group design. *Psychological Bulletin*, *46*, 137–149.
- Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Monographs*, *2*(4), 1–109.
- Thorndike, E. L. (1913). *Educational psychology: Vol. 2. The psychology of learning*. New York: Teachers College, Columbia University.
- Thorndike, E. L., & Woodworth, R. S. (1901a). The influence of improvement in one mental function upon the efficiency of other functions: I. *Psychological Review*, *8*, 247–261.
- Thorndike, E. L., & Woodworth, R. S. (1901b). The influence of improvement in one mental function upon the efficiency of other functions: II. The estimation of magnitudes. *Psychological Review*, *8*, 384–395.
- Thorndike, E. L., & Woodworth, R. S. (1901c). The influence of improvement in one mental function upon the efficiency of other functions: III. Functions involving attention, observation and discrimination. *Psychological Review*, *8*, 553–564.
- Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology*, *9*, 507–533.
- Winston, A. S. (1990). Robert Sessions Woodworth and the “Columbia Bible.” *American Journal of Psychology*, *103*, 391–401.
- Woodworth, R. S. (1913). *The problems and methods of psychology*. Unpublished manuscript, Columbia University, New York.
- Woodworth, R. S. (1938). *Experimental psychology*. New York: Holt.
- Woodworth, R. S., & Poffenberger, A. T. (1932). *Textbook of experimental psychology*. Unpublished manuscript, Columbia University, New York. (Original work mimeographed 1920)