American Economic Journal: Applied Economics 3 (January 2011): 224–238
http://www.aeaweb.org/articles.php?doi=10.1257/app.3.1.224

Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments†

By Steven D. Levitt and John A. List*

The "Hawthorne effect" draws its name from a landmark set of studies conducted at the Hawthorne plant in the 1920s. The data from the first and most influential of these studies, the "Illumination Experiment," were never formally analyzed and were thought to have been destroyed. Our research has uncovered these data. Existing descriptions of supposedly remarkable data patterns prove to be entirely fictional. We do find more subtle manifestations of possible Hawthorne effects. We also propose a new means of testing for Hawthorne effects based on excess responsiveness to experimenter-induced variations relative to naturally occurring variation. (*JEL* C90, J24, J28, M12, M54, N32)

> The experiments started with, continued with, and ended with attention focused on one thing and one thing only, what people do. This was the new procedure and it was revolutionary, in the same way that Galileo's or Mendel's procedure was revolutionary in the science of their time.
>
> —C. W. M. Hart (1943)

When the National Research Council initiated a set of experiments at Western Electric's Hawthorne Plant in Cicero, IL in 1924, its objective was to answer a narrow question: does better lighting enhance worker productivity? The results of these experiments, however, have had a profound influence on research in the social sciences ever since. According to Ann B. Blalock and Hubert M. Blalock (1982, 72), to the surprise of the researchers, "each time a change was made, worker productivity increased .... As a final check, the experimenters returned to the original unfavorable conditions of poor lighting .... Seemingly perversely, productivity continued to rise." Spurred by these initial findings, a series of experiments were conducted at the plant over the next nine years. New empirical results reinforced the initial findings. Jonathan Freedman, David O. Sears, and J. Merrill Carlsmith (1981,

* Levitt: Department of Economics, University of Chicago, 1126 E. 59th St., Chicago, IL 60637, and National Bureau of Economic Research (e-mail: slevitt@uchicago.edu); List: Department of Economics, University of Chicago, 1126 E. 59th St., Chicago, IL 60637, and National Bureau of Economic Research (e-mail: jlist@uchicago.edu).Thanks to Sonia Jaffe, Andrew Hogue, and Colin Spitz for outstanding research assistance. Marc Nerlove, Joel Sobel, Charles Wrege, and three anonymous referees provided invaluable comments and suggestions. Financial support came from the National Science Foundation and the Sherman Shapiro Research Fund.

† To comment on this article in the online discussion forum, or to view additional materials, visit the article page at http://www.aeaweb.org/articles.php?doi=10.1257/app.3.1.224.

49) summarize the results of the next round of experiments as follows: "Regardless of the conditions, whether there were more or fewer rest periods, longer or shorter workdays … the women worked harder and more efficiently."

From these experiments, emerged the concept of the "Hawthorne effect," which is defined very broadly by the Oxford English Dictionary (OED) as meaning "an improvement in the performance of workers resulting from a change in their working conditions, and caused either by their response to innovation or by the feeling that they are being accorded some attention." These experiments are now recognized as some of the most influential in social science, helping to spawn (along with Frederick Taylor's 1911 *The Principles of Scientific Management*) the development of a new field of study—industrial psychology—and still influencing experimental research design today. As Milton Bloombaum (1983) notes, "In the history of science, certain contributions stand out as signal events in the sense that they influence a great deal of what follows. The Hawthorne Experiments exemplify this phenomenon in the field of industrial work…. "[1]

Although economists have historically devoted less attention to the Hawthorne effect than psychologists, this has begun to change as the use of randomization has become more prominent within the discipline (see, e.g., Esther Duflo, Rachel Glennerster, and Michael Kremer (2008) within the development context and List (2006) more generally). Within this line of research, issues such as randomization bias and effects of scrutiny remain of utmost importance when interpreting treatment effects. James J. Heckman (1991) provides a good discussion of the evidence on randomization bias, and Levitt and List (2007) point to the increased scrutiny associated with participation in laboratory and social experiments as one of the greatest threats to the generalizability of such studies. Economists are also beginning to explore the closely related issue of placebo effects (see, e.g., Anup Malani 2006).

Despite the prominence of the original Hawthorne experiments, when scholars have later re-analyzed the data, the results have not been as clear-cut as the original researchers claimed. Richard Herbert Franke and James D. Kaul (1978) were the first to carefully analyze data from what is known as the "first relay" experiment at Hawthorne, concluding that most of the variation in production rates could be explained by differences in variables such as managerial discipline and the amount of employee rest. Consequently, there was relatively little scope for "unmeasured changes in the human relations of workers." Jones (1992), again focusing on data from the first relay experiment, attempts to more directly measure the magnitude of any Hawthorne effects. He finds no evidence, either in the raw data or after controlling for other factors, to support the traditional interpretation of the Hawthorne data.[2]

[1] The effect, of course, has gone well beyond academic circles, as noted in Stephen R. G. Jones (1992), Thomas J. Peters and Robert H. Waterman, Jr. (1982, 5–6) write: "For us, the very important message of the research … is that it is *attention to employees*, not work conditions per se, that has the dominant impact on productivity. (Many of our best companies, one friend observed, seem to reduce management to merely creating 'an endless stream of Hawthorne effects')."

[2] Although it should be noted that Bloombaum (1983) challenged many of Franke and Kaul's (1978) conclusions. See Michiel A. J. Kompier (2006) for a discussion of why the mythology surrounding the original Hawthorne experiments has been so persistent.

Unlike the first relay experiments, no statistical analysis has ever been performed on the original illumination studies at Hawthorne, which predate the relay experiments that are the focus of Franke and Kaul (1978) and Jones (1992). It is the illumination studies that are the genesis of the popular claims that whenever light was manipulated, whether upward or downward, output increased. For example, Franke and Kaul (1978, 624) describe popular views of the data as showing that: "Inexplicably worker output … increased regardless of increase or decrease in illumination." Similarly, Blalock and Blalock (1982, 72) report that "each time a change was made, worker productivity increased…. As a final check, the experimenters returned to the original unfavorable conditions of poor lighting…. Seemingly perversely, productivity continued to rise." Interestingly, these statements have become urban legend and difficult to test because the illumination data are broadly believed to have been destroyed.[3]

In this paper, we provide the first statistical analysis of these nearly 90 year old data. We were able to locate microfilm relating to the illumination experiments in collections at the University of Wisconsin-Milwaukee and Baker Library at Harvard Business School. Our search for the original data was triggered by a sentence in Franke and Kaul's (1978) appendix that mentioned the University of Wisconsin-Milwaukee data archive and an index created by the reference librarians (Stanley Mallach and Steven Smith 1977). Indeed, going through that archive we found a number of documents relating to the illumination studies, including a graphical depiction of one room's data from the original illumination experiments, which appear to represent the means by which the data were originally recorded by the researchers. Further searching led us to the Baker Library at Harvard University, which contained graphical records for two additional rooms. These three data series are the basis of our empirical analysis. To the best of our knowledge, these illumination data have never previously been coded and statistically examined.[4]

Our analysis of the newly found data reveals little evidence in favor of a Hawthorne effect as commonly described, i.e., productivity rising whenever light is manipulated. A naïve reading of the raw data does produce such a pattern, however. This is only because all lighting changes occurred on Sundays, the only off day for workers. The empirical fact is that productivity is higher on Mondays than on Fridays and Saturdays. Output on Mondays is equally high, however, whether or not a lighting change occurs on that particular Monday. Thus, researchers seemingly misinterpreted the day of week effect with the Hawthorne effect.

We do uncover some weak evidence consistent with more subtle manifestations of Hawthorne effects in the data. First, productivity increases in the areas that were part of the experiments were much greater than for the plant overall. Second, out-

[3] Berkeley Rice (1982) laments this fact by noting that "the original [illumination] research data somehow disappeared." E. A. M. Gale (2004, 439) reinforces this notion by stating that "these particular experiments were never written up, the original study reports were lost, and the only contemporary account of them derives from a few paragraphs in a trade journal." Richard Gillespie (1991, 46–47) speculates that the reason the experiments were not written up was due to the fact that the findings did not assuage industry: "the final report would have recommended a basic lighting level of 7 to 10 foot candles … (which) … clearly failed to satisfy the electrical industry's expectations that the research would provide the scientific justification … for higher levels of illumination."

[4] The illumination data were presented graphically. We extracted the information from the graphs and placed it into spreadsheet form. The raw data for all variables used in this analysis are available upon request.

put tended to be higher when experimental manipulations were ongoing relative to when there was no experimentation. Finally, productivity is more responsive to experimenter manipulations of light than naturally occurring fluctuations, consistent with the idea that it was not the light itself, but rather the manipulation of the light that mattered. Ultimately, however, the poor research design of the Hawthorne illumination treatments limits what can be learned.

The remainder of the paper is as follows. Section I provides a framework for thinking about the disparate definitions of the Hawthorne effect that are sometimes observed in the literature. Section II provides background on the illumination experiments conducted at the Hawthorne plant between 1924 and 1927. Section III provides the first rigorous analysis of the data generated by these illumination experiments. Section IV concludes.

I. What Do Researchers Mean When They Refer to a Hawthorne Effect?

One of the difficulties associated with testing for Hawthorne effects is the absence of a precise definition of the Hawthorne effect.[5] In this section, we outline the various possible manifestations of a Hawthorne effect.[6] For the purposes of explication, we will assume that the setting being analyzed is a randomized experiment where the nature of the experimental variation is increases or decreases in light.

There are at least three channels through which a Hawthorne effect might arise. The first channel is via the scrutiny and emphasis on process accompanying experimentation (e.g., signing of release forms, reading of instructions, etc.). With this channel, it is participation in the experiment, rather than the experimental manipulation itself (i.e., the increased or decreased light), that is affecting behavior. If this is the case and the same degree of scrutiny is given to the treatment and control groups, then such an effect would appear as a uniform increase in output across both treatment and control groups. To address this type of effect, researchers sometimes use two sets of control groups: a so-called "Hawthorne control group" that does not receive treatment, but mirrors the treatment group with respect to the levels of attention that are devoted to them and are made equally aware of experimental participation, as well as a second control group that is unaware that they are being experimented upon (e.g., John G. Adair, Donald Sharpe, and Cam-Loi Huynh 1989). Any differences in output between the "Hawthorne control group" and the other control group are attributed to a Hawthorne effect.

A second (but not mutually exclusive) source of the Hawthorne effect is driven by the experimental treatment itself. In our example, for instance, the repeated manipulation of lighting levels over time might provide reminders to the treatment group (but not to either the "Hawthorne control group" or the other control group) that they are under observation. Under this scenario, one might expect to observe spikes

[5] As evidence on this point, The Merriam-Webster dictionary offers a very different definition for a Hawthorne effect than the one cited in the OED: "the stimulation to output or accomplishment that results from the mere fact of being under observation." Along the lines of this second definition, is the use of so-called "Hawthorne control groups" in randomized trials in addition to the standard treatment and control groups.

[6] An earlier version of this paper presents a mathematical formalization of the ideas discussed in this section. This formalization is available upon request.

in output for the treatment group in response to each adjustment in lighting levels. This is the pattern described in the popular accounts of the illumination studies. Note that use of a "Hawthorne control group" may not completely capture this channel of the Hawthorne effect.

A third version of the Hawthorne effect can arise as a result of so-called "experimenter demand effects" whereby experimental subjects attempt to act in ways that will please the experimenter (Levitt and List 2007). If workers think that the experimenters wish to show that greater lighting increases output, the response of output to artificial, experimenter-induced variation in light is likely to be more exaggerated than the response to natural variation in light due, say, to weather conditions or the change of seasons. A simple experimental design with a treatment group, a control group, and a "Hawthorne control group" will, in this case, generate experimental estimates that exaggerate the true impact of naturally occurring changes in light.

Further complicating the empirical estimation of the Hawthorne effect is that, from a theoretical perspective, it is unclear whether the Hawthorne effect is a short-run or long-run phenomenon. Experimental manipulation may increase output in the short run, with productivity levels returning back to normal levels as workers habituate (see, e.g., Uri Gneezy and List 2006). Alternatively, if the productivity increases are due to process improvements, the effect of a one-time experiment may be long-lived.

As the preceding paragraphs make clear, Hawthorne effects may appear in different forms, ranging from short-run spikes associated with changes in treatment to long-run improvements for those participating in the experiments to excess sensitivity to variation that is experimentally induced. In our empirical work, we explore these manifestations.

II. Experimentation at the Hawthorne Plant

The illumination experiments were carried out at the Western Electric Company's "Hawthorne Works" factory located in the Chicago suburbs. Western Electric dispatched two engineers, George Pennock and Clarence Stoll to the Hawthorne plant from 1924 to1927 to oversee the experiments (Gale 2004). Charles Snow, a Massachusetts Institute of Technology instructor, assumed day-to-day responsibility for carrying out the experiments.

These experiments, which are summarized in Figure 1, are known as the "illumination experiments" because they varied the amount of light in the workplace to study how such variation influenced productivity. Figure 1 provides a timeline summarizing the various phases of the illumination experiments, noting for each department the periods during which light was varied experimentally, as well as the periods when output was recorded. Workers in these three departments were women who primarily assembled relays or wound coils of wire, and their output was measured as units completed per unit of time.

As Figure 1 shows, the first wave of the experiment lasted from November 1924 through April 1925 and involved all three departments. During the summer months there was ample natural light and light recording (and manipulation) was suspended, although output continued to be reported in all three departments. After the summer

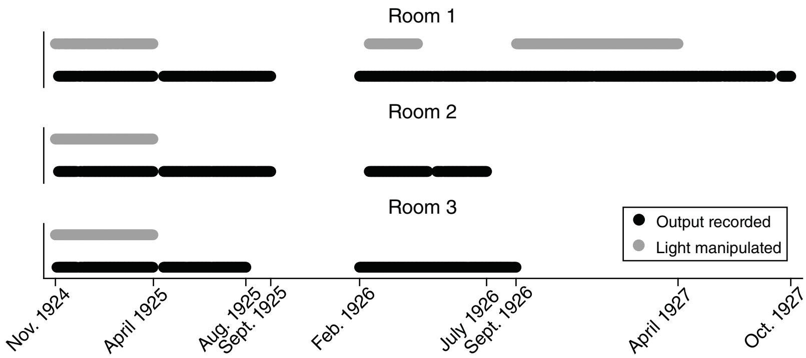

FIGURE 1. TIMELINE OF THE ILLUMINATION EXPERIMENTS

Notes: The illumination experiments took place sporadically between November 1924 and October 1927. Three separate rooms of employees participated. The lighter lines in the figure denote, by room, the time periods when active experimentation was taking place. The darker lines report time periods when output was recorded.

ended there was a period of time during which no data were recorded (not even output). In February 1926, light manipulation experiments resumed, but only in one of the three rooms. Although there was no experimentation, output was recorded in the other two rooms through the end of the summer of 1926. The fall of 1926 saw continued lighting experiments in the first room, but data from the other two rooms were no longer reported. The experiments culminated with the most radical experimental manipulations: workers in the first room were dispatched to a room where the windows were blackened and burlap covered the sky lights in order to achieve more precise control over lighting conditions. (see Elton Mayo 1933, 55–56 or F. J. Roethlisberger and William J. Dickson 1939, 14–18). April 1927 marked the end of the illumination experiments, although researchers continued to monitor output in the first room through October 1927.[7]

III. Illumination Data

The illumination experiments summarized in Figure 1 provide a wealth of data and exogenous variation. For example, Figure 2 shows how light varied over time in Room 1, where the primary experimental manipulation of illumination occurred. Figure 2 reports both the average levels of natural and artificial light over time. Natural light varies with the seasons and is not manipulated by the experimenters, except at the end of the study when the windows were blackened so there was no natural light. The original lighting system in place at the beginning of the experiments provided four foot candles of artificial light which was supplemented with

[7] Following on the heels of the illumination studies, a series of further experiments were undertaken. Western Electric brought in academic consultants, most prominently Mayo, in 1928. Experiments were carried out on workers whose jobs were to make relays, split mica, and assemble telephones. The experiments lasted until June 1932, when the women in the test room received their notices due to the poor economy (except the most exceptional worker, Jennie Sirchio, who worked in the office for a few months before being let go (Gale 2004)). This second series of experiments provided a wealth of data, summarized in Mayo (1933).

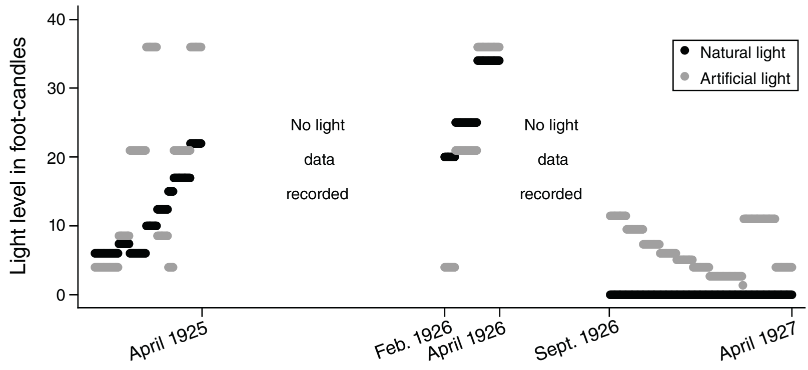

Figure 2. Variation in Artificial and Natural Light during the Illumination Experiments in the Primary Experiment Room

Notes: The work rooms were illuminated by both natural and artificial light. Fluctuations in artificial light represent experimental manipulations. Prior to September 1926, all variation in natural light was due to seasonal patterns; after September 1926, the windows were blackened and only artifical light was present.

natural light. The first year of light manipulation ranged from this baseline level up to 36 foot candles of artificial light. The second year's experiments used the same range of variation of artificial light, but the third year's experiments differed in that all natural light was eliminated.[8] In this last period, the experimenters started with 11.5 foot candles of artificial light and incrementally lowered light until it reached only 1.4 foot candles (for just one day) before being increased to 11 foot candles. The illumination experiments ended shortly thereafter.

For experimental purposes, all lighting changes commenced on Monday, so that facility employees would have ample time to alter the lighting arrangements on Sunday, which was the only day of the week that the factory was not in operation. According to Gillespie (1991), at the start of each new experimental period workers were made aware of the experimental changes. Of course, as discussed above, this very warning can have important effects on worker effort. Output is measured by room. There is neither recorded information on worker characteristics nor on the productivity of individual workers.[9]

A. *Short-run Output Fluctuations as Evidence for a Hawthorne Effect*

It is commonly asserted that output increased immediately in response to experimental manipulation of the work environment, regardless of whether light was

[8] As C. E. Snow writes in his unpublished report on the third year's illumination experiments, "This test differed from the two preceding tests.…The illumination supplied to the operatives was artificial only, while previously a mix of artificial and natural illumination had been used.…Close control of illumination intensity had always been unattainable in our previous tests due to unavoidable variation in natural illumination" (Snow 1927). In private correspondence, Charles Wrege also notes that the Hawthorne plant had its own power generator and experienced frequent fluctuations in voltage that affected the intensity of artificial lighting. Snow's unpublished report was graciously provided by Wrege. The report is contained in the Kheel Center for Labor-Management Documentation and Archives, Martin Catherwood Library, Cornell University.

[9] There is no mention of turnover of employees either in the published description of the illumination studies (Snow 1927), or in the unpublished documents that we have seen. The issue of turnover does appear to have had an important role in determining output fluctuations in the later "first relay" experiments (Franke and Kaul 1978).

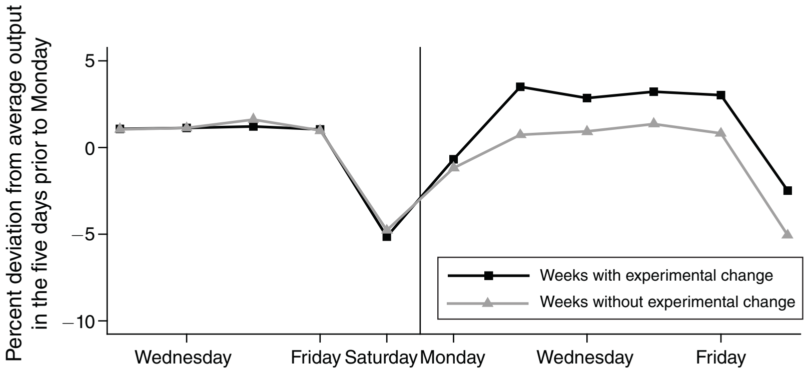

Figure 3. Comparison of Productivity between Weeks with and without Experimental Light Changes

Notes: This figure presents daily fluctuations in output per worker. The black line reflects weeks in which light was experimentally altered. For purposes of comparison, output fluctuations are also shown for weeks in which no experimental change occurred. All lighting changes took effect on Mondays, as denoted by the vertical line. Values in the figure are percent deviations from the average output in the five days preceding the Monday shown in the figure.

increased or decreased. Figure 3 plots average daily output in the five days preceding an experimental change and in the five days after a change to test this claim. As noted above, all lighting changes were instituted on Mondays. Because output varies systematically by day of the week, we also report the output patterns for the ten days surrounding Mondays in which there was no change in experimental conditions for purposes of comparison. For both lines, we have normalized output around the average productivity in the preceding five days. The gap between these two lines reflects a systematic response of output to variations in lighting.

As is apparent in Figure 3, output rises sharply on Mondays regardless of whether artificial light is altered. Output is no higher on the day of the lighting change, but is somewhat greater 2–5 days after an experimental change in the raw data—a pattern potentially consistent with a Hawthorne effect and the experimental data in Gneezy and List (2006). As we demonstrate below, however, in our regression specifications, the gap between these two lines remains, but is measured imprecisely.[10]

In order to more formally analyze these issues we estimate regression models:

$$Output_{rt} = \sum_{k=0}^{5} \delta k E_{rt-k} + \beta_1 t + \beta_2 t^2 + \lambda_r + X_{rt}\Phi + \varepsilon_{rt}, \tag{1}$$

where r and t index a room and time, respectively. E is a series of indicator variable reflecting whether there was an experimentally driven change in artificial light on each of the previous six days. We control for both linear and quadratic trends in output over time, as well as room fixed-effects. X is a vector of controls that includes weather conditions, indicators for the day of the week and the month, as well as

[10] Dividing experimental changes in lighting conditions into those that are positive (i.e., light levels are increased) and negative (light levels are lowered) yields patterns that are similar.

TABLE 1—IMMEDIATE RESPONSE OF OUTPUT TO EXPERIMENTAL VARIATION
REGRESSION ANALYSIS OF EXPERIMENTATION WITH LIGHTING ON OUTPUT
(*Dependent Variable: Daily Output per Worker* (*Baseline* = *100*))

	(1)	(2)	(3)
Day of experimental change	−1.021	−1.350	−1.282
	(0.81)	(0.80)	(0.75)
One day after experimental change	1.076	0.818	0.849
	(0.80)	(0.79)	(0.74)
Two days after experimental change	0.383	0.232	0.208
	(0.80)	(0.79)	(0.74)
Three days after experimental change	0.707	0.462	0.307
	(0.83)	(0.82)	(0.77)
Four days after experimental change	0.888	0.556	0.344
	(0.81)	(0.80)	(0.76)
Five days after experimental change	0.439	0.129	−0.051
	(0.81)	(0.80)	(0.76)
Linear time trend	0.025	0.065	0.052
	(0.00)	(0.03)	(0.03)
Quadratic time trend divided by 10,000	−0.034	−0.079	−0.020
	(0.03)	(0.29)	(0.28)
R^2	0.69	0.71	0.75
Includes controls?	Yes	Yes	Yes
Includes month-year interactions?	No	Yes	Yes
Includes room-month-year interactions?	No	No	Yes
p-value: Test of joint significance of experimental change dummies	0.252	0.702	0.858

Notes: The unit of observation is daily output per worker in a room, normalized to equal 100 in the baseline period. Entries in the table are regression coefficients with accompanying standard errors. All specifications include linear and quadratic trends, weather conditions, day and week indicators, day-before-holiday and day-after-holiday indicators, defective inputs, variations in output-checking, and room fixed effects. Specification (2) also controls for room-month-year interactions. The number of observations is equal to 1,502 in all specifications. Full regression results are presented in the online Appendix.

indicators for days before and after a work-canceling holiday, and days in which the inputs were denoted as defective. We also include the frequency with which output is checked and whether workers are located in their original room in the vector of controls. In some specifications we add month × year interactions to provide more flexibility in the way that output fluctuates over time. In other specifications we include room × month × year interactions so that treatment effect identification comes solely from comparisons of observations in the same room, month, and year.

The coefficients of primary interest from the estimation are presented in Table 1. Column 1 includes all control variables except the month × year and room × month × year interactions, which are added in columns 2 and 3, respectively.[11] Standard errors are shown in parentheses. The bottom row of the table

[11] Estimated coefficients on these control variables are not presented in Table 1 for parsimony, but full results are presented in the online Appendix. The trend in output over time is positive, but concave. The estimated coefficients on the control variables are intuitive, and imply that output responds positively to light, negatively to high temperatures, and is low after holidays, when the inputs used are defective, and when the workers are moved to a new room.

reports the p-value for rejecting the null hypothesis that the sum of the indicator variables for recent experimental changes in lighting.

None of the coefficients in any of the three columns is statistically significant. The point estimate is negative on the first day light is changed and, in almost all cases, positive on the following days. The combined impact is positive, but we cannot reject that the sum of the coefficients is equal to zero, as shown in the last row of the table. The claim that output immediately and positively responded to experimentally induced changes in lighting is not supported by the data.[12]

B. *Longer Run Fluctuations in Output as Evidence of a Hawthorne Effect*

Even if productivity does not respond immediately to manipulations in lighting, over a longer time horizon a Hawthorne effect might manifest itself as an unusually large increase in productivity in those rooms subjected to experimentation, potentially followed by a decline in output when the experiments are halted.

Figure 4 shows how output varied by week in each of the three rooms. In the first year, experimental manipulation occurred in all three rooms with no well-defined control group. Productivity changes in the three rooms follow a very similar pattern in the first year of the study. Output initially rises sharply and steadily, reaching a peak roughly 10 percent above the baseline productivity prior to the start of the experiment. Using information from annual company reports from the 1920s, we estimate that output per worker for the plant as a whole increased 1.4 percent per year over the period the illumination experiments were active. As crude as this analysis is, these large productivity increases relative to the rest of the plant nonetheless stand as some of the most compelling evidence in support of a Hawthorne effect that we are able to muster from the data.

The decline in output after experimentation is temporarily halted in April of the first year, while also superficially consistent with a Hawthorne effect that fades with time, appears instead to merely be due to seasonal fluctuations in output. That same decline is observed in later years, even for rooms that are not being actively experimented upon.

The second round of experiments arguably provides more useful variation than the first round because only room 1 was subjected to experimental variation in that period, which is bracketed by the third and fourth vertical lines in Figure 3. Consequently, rooms 2 and 3 serve as quasi-control groups for room 1, although it should be stressed that there is no evidence that any sort of randomization was involved in the choice to continue experimentation in room 1, but not in rooms 2 and 3. Room 1 experienced large productivity gains during this second phase of experimentation (with output peaking around 20 percent above the pre-experimental baseline), but interestingly, room 2 (and to a lesser extent room 3) also saw substantial increases in output over this interval despite the fact that those workers were no longer included in the experiment.

[12] Similar conclusions are reached when we limit our sample only to the period when one room was being experimented on and the other two rooms were not—a source of variation that conforms more closely to the notion of treatment and control.

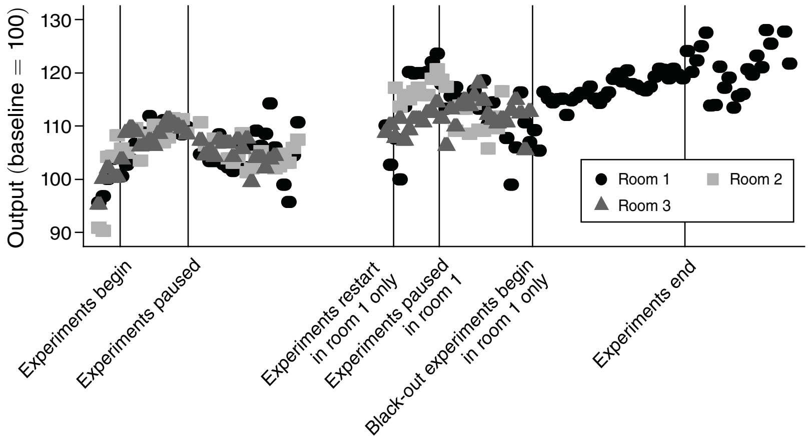

Figure 4. Output During the Illumination Experiments Over Time, by Room

Notes: Values in the figure represent weekly average output per worker, normalized to equal 100 in the baseline period preceding experimentation. All rooms participated in the initial round of experiments; only room 1 was part of later rounds of experimentation. Output data is available only for Room 1 in the latter part of the sample.

In the final two periods shown in Figure 3, output was tracked only for room 1, making it difficult to estimate the experimental impact. Output steadily increased during this last phase of experimentation, rose sharply when the experiments stopped, fell sharply during the following summer, but then rebounded to all-time highs toward the end of the sample period when no experimentation was ongoing.

Table 2 presents regression estimates of the relationship between output and the presence or absence of experimentation. In these specifications, we divide all observations into one of three mutually exclusive categories: the roughly one month baseline period prior to the start of experimentation, times when active experimentation is ongoing,[13] and times when experimentation has been suspended. We include indicator variables for the latter two of these classifications; thus the period before experimentation begins is the omitted category against which the other effects are measured. The same set of controls employed in Table 1 is also included in these regressions. The structure of this table parallels that of Table 1; column 2 adds month-year interactions, and column 3 controls for room-month-year interactions.

Results from the specification in columns 1 and 2 are consistent with a Hawthorne effect. In column 1, when experimentation is ongoing, output is 3–4 percent higher than in the pre-experimentation baseline and the nonexperimental periods, even after controlling for other factors including linear and quadratic time trends and the amount of light. The results change somewhat in column 2, when month-year interactions are included. The implied gap between periods of experimentation and nonexperimentation jumps to nearly 6 percent. Although the experimental period is no longer statistically distinguishable from the pre-period, this comparison is based

[13] This variable is equal to one in the first round of experimentation in all three rooms and in the later rounds of experimenting for room 1 only.

TABLE 2—LONG-TERM RESPONSE OF OUTPUT TO EXPERIMENTAL VARIATION

	(1)	(2)	(3)
Pre-experimentation (baseline)	–	–	–
Experimentation ongoing	3.154	0.787	1.201
	(0.86)	(1.53)	(1.46)
Experimentation paused or completed	−1.187	−4.992	2.743
	(1.60)	(2.13)	(2.41)
Linear time trend	0.023	0.059	0.045
	(0.00)	(0.03)	(0.03)
Quadratic time trend divided by 10,000	−0.019	−0.078	0.040
	(0.03)	(0.30)	(0.29)
R^2	0.70	0.71	0.75
Includes controls?	Yes	Yes	Yes
Includes month-year interactions?	No	Yes	Yes
Includes room-month-year interactions?	No	No	Yes

Notes: The unit of observation is daily output per worker in a room, normalized to equal 100 in the baseline period. Entries in the table are regression coefficients with accompanying standard errors. All specifications include linear and quadratic trends, weather conditions, day and week indicators, day-before-holiday and day-after-holiday indicators, defective inputs, variations in output-checking, and room fixed effects. Specification (2) also controls for room-month-year interactions. The number of observations is equal to 1,502 in all specifications. Full regression results are presented in the online Appendix.

on just a few days of experimental data from the end of December 1924 since only comparisons within month and year identify the parameters.[14]

Column 3, which adds room-month-year interactions, asks a slightly different question of the data. These estimates are identified solely from within-room variation in output in the months when experimentation is started or halted. Consequently, the top two rows of column 3 show whether there is an abrupt jump up or down in output associated with turning the experiment on or off. Unlike the first two columns, the coefficient on experimentation is not statistically different from the coefficient on nonexperimentation, providing little support for the argument that output immediately responds to initiating or ending experimentation.

C. *Excess Sensitivity to Experimental Manipulation in Light as Evidence of a Hawthorne Effect*

Empirical results presented thus far do not exploit the information we have on how total light is composed. As noted earlier, one manner in which a Hawthorne effect might manifest itself is with output being more responsive to changes in artificial light than to fluctuations in natural light. The logic for this argument is that workers know that the experimenter does not control fluctuations in natural light which are driven by weather and seasonal changes. Only changes in artificial light are manipulated, and

[14] Identification in column 2 comes from two sources of variation: months in which experimentation is turned on or off, and cross-room variation in period 2 when room 1 was subject to experimentation, but not the other two rooms. The results are similar when the sample is limited to only the latter source of identification by using only data from the second period.

TABLE 3—RESPONSE OF OUTPUT TO VARIATIONS IN THE LEVEL OF ARTIFICIAL VERSUS NATURAL LIGHT

	(1)	(2)	(3)
Artificial light	0.056	0.033	0.0329
	(0.02)	(0.02)	(0.02)
Natural light	−0.045	−0.018	−0.021
	(0.03)	(0.03)	(0.03)
Linear time trend	0.061	0.075	0.075
	(0.01)	(0.03)	(0.03)
Quadratic time trend divided by 10,000	−0.558	−0.283	−0.282
	(0.12)	(0.38)	(0.37)
R^2	0.75	0.76	0.77
Includes controls?	Yes	Yes	Yes
Includes month-year interactions?	No	Yes	Yes
Includes room-month-year interactions?	No	No	Yes

Notes: The unit of observation is daily output per worker in a room, normalized to equal 100 in the baseline period. Entries in the table are regression coefficients with accompanying standard errors. All specifications include linear and quadratic trends, weather conditions, day and week indicators, day-before-holiday and day-after-holiday indicators, defective inputs, variations in output-checking, and room fixed effects. Specification (2) also controls for room-month-year interactions. The number of observations is equal to 1,502 in all specifications. Full regression results are presented in the online Appendix.

thus, if workers are particularly influenced by experimental variation, they should be more responsive to artificial light. In addition, changes in artificial light were brought to the attention of the workers on the morning of the new treatment, whereas weather driven changes in natural light were not reported to the workers. As such, one might even expect to see productivity rise in response to experimentally induced *reductions* in man-made light if the Hawthorne effect is stronger than any direct impact of reduced light on the production process. To the best of our knowledge, this implication regarding Hawthorne effects has not been previously recognized or empirically tested.

Table 3 presents results for this new test. The dataset utilized in Table 3 includes fewer observations than in the two previous tables because we observe the breakdown between artificial and natural light in only a subset of the data.[15] The key right-hand-side variables in Table 3 are the measured quantities of natural and artificial light in a given room on a particular day. In all other regards, Table 3 parallels the earlier tables. In both columns 1 and 2, artificial light enters positively (in one case statistically significant, but not in the other), whereas natural light is negative and insignificant in both columns. The actual magnitude of the impact of fluctuations in artificial light is not particularly large, however. A one-standard deviation change in artificial light increases output by less than 1 percent. While clearly circumstantial, this pattern is directionally consistent with a Hawthorne effect because the increase in output in response to artificial light is greater than for natural light.[16]

[15] When we do not separate natural and artificial light, but rather combine the two to obtain total light, each foot-candle of additional light is associated with a statistically significant, but relatively small, increase of 0.0865 units of output. Thus, a one standard deviation increase in light raises output by 1–2 percent. Yet, adding month by year interactions yields a coefficient that is statistically insignificant and trivial in magnitude.

[16] An additional caveat regarding this estimation approach is that natural light is likely to be measured with greater error than artificial light since the light readings were gathered at just a few specific times during the day.

IV. Conclusions

The Hawthorne plant studies, and the concept of a Hawthorne effect that emerged from this seminal research, stand among the most influential social science research of the twentieth century. The purported influence of observation on measured treatment effects in these experiments has led to a proliferation of research and methodologies to control for the confounding influence that scrutiny can have (e.g., Desmond L. Cook 1962; Glenn W. Harrison and List 2004). Outside of the academy, results from the Hawthorne studies bolstered the human relations movement, considerably influenced employee/employer relations, and remain an important influence on the optimal incentive schemes employed in the workplace.

This study returns to the very evidence that induced this contemporary wave of thought by examining new data that was presumed lost. Ironically, there is little evidence of the type of Hawthorne effect widely attributed to these data when one subjects them to careful analysis. We do see evidence that workers, over a longer time horizon, appear to respond positively to experimentation. We also propose and test a new manifestation of the Hawthorne effect: whether workers differentially respond to natural and man-made light. We find some weak evidence that workers respond more acutely to the experimental manipulations than to naturally occurring fluctuations in light.

The illumination studies have been hailed as being among the most important social science experiments of all time, but an honest appraisal of this experiment reveals that the experimental design was not strong, the manner in which the studies were carried out was lacking, and the results were mixed at best. Perhaps fittingly, a meta-analysis of the research testing the enormous body of research into Hawthorne effects triggered by this initial study yields equally mixed results (Adair, Sharpe, and Hyunh 1989).

Perhaps the most important lesson to be learned from the original Hawthorne experiments is the power of a good story. The mythology surrounding the Hawthorne experiments arose largely absent careful data analysis, and has persisted for decades even in the face of strong evidence against it generated by Franke and Kaul (1978) and Jones (1992). While our research is probably no more likely than the previous papers to put an end to such myths, at a minimum it raises the costs of propagating these stories among those who are concerned with scientific accuracy.

REFERENCES

Adair, John G., Donald Sharpe, and Cam-Loi Huynh. 1989. "Hawthorne Control Procedures in Educational Experiments: A Reconsideration of Their Use and Effectiveness." *Review of Educational Research*, 59(2): 215–28.

Blalock, Ann B., and Hubert M. Blalock. 1982. *Introduction to Social Research.* Englewood Cliffs, NJ: Prentice-Hall.

Bloombaum, Milton. 1983. "The Hawthorne Experiments: A Critique and Reanalysis of the First Statistical Interpretation by Franke and Kaul." *Sociological Perspectives*, 26(1): 71–88.

This could lead to attenuation bias for natural light. Furthermore, we are making a structural assumption on the marginal value of artificial and natural light to the production process.

Cook, Desmond L. 1962. "The Hawthorne Effect in Educational Research." *Phi Delta Kappan*, 44(3): 116–22.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics*, Vol. 4, ed. T. Paul Schultz and John A. Strauss, 3895–3962. Amsterdam: Elsevier.

Franke, Richard Herbert, and James D. Kaul. 1978. "The Hawthorne Experiments: First Statistical Interpretation." *American Sociological Review*, 43(5): 623–43.

Freedman, Jonathan, David O. Sears, and J. Merrill Carlsmith. 1981. *Social Psychology*. 4th ed. Englewood Cliffs, NJ: Prentice-Hall.

Gale, E. A. M. 2004. "The Hawthorne Studies—A Fable for Our Times?" *Quarterly Journal of Medicine*, 97(7): 439–49.

Gillespie, Richard. 1991. *Manufacturing Knowledge: A History of the Hawthorne Experiments.* New York: Cambridge University Press.

Gneezy, Uri, and John A. List. 2006. "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments." *Econometrica*, 74(5): 1365–84.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature*, 42(4): 1009–55.

Hart, C. W. M. 1943. "The Hawthorne Experiments." *Canadian Journal of Economics and Political Science*, 9(2): 150–63.

Heckman, James J. 1991. "Randomization and Social Policy Evaluation." National Bureau of Economic Research Technical Working Paper 107.

Jones, Stephen R. G. 1992. "Was There a Hawthorne Effect?" *American Journal of Sociology*, 98(3): 451–68.

Kompier, Michiel A. J. 2006. "The 'Hawthorne Effect' Is A Myth, But What Keeps the Story Going?" *Scandinavian Journal of Work, Environment, & Health*, 32(5): 402–12.

Levitt, Steven D., and John A. List. 2007. "What do Laboratory Experiments Measuring Social Preferences Reveal About the Real World," *Journal of Economic Perspectives,* 21(2): 153–174.

List, John A. 2006. "Field Experiments: A Bridge between Lab and Naturally Occurring Data." *Advances in Economic Analysis & Policy*, 6(2): Article 8.

Malani, Anup. 2006. "Identifying Placebo Effects with Data from Clinical Trials." *Journal of Political Economy*, 114(2): 236–56.

Mallach, Stanley, and Steven Smith, ed. 1977. "Records of Industrial Relations Experiment Carried out by Western Electric Company at the Hawthorne Works, Hawthorne, Illinois." Milwaukee, WI: University of Wisconsin Library Microfilms.

Mayo, Elton. 1933. *The Human Problems of an Industrial Civilization.* New York: Macmillan Company.

Peters, Thomas J., and Robert H. Waterman, Jr. 1982. *In Search of Excellence: Lessons from America's Best-Run Companies.* New York: Harper & Row.

Rice, Berkeley. 1982. "The Hawthorne Defect: Persistence of a Flawed Theory." *Psychology Today*, 16(2): 71–74.

Roethlisberger, F. J., and William J. Dickson. 1939. *Management and the Worker.* Cambridge, MA: Harvard University Press.

Snow, C. E. 1927. "Research on Industrial Illumination." *Tech Engineering News*, 8: 257–82.

Wrege, Charles. 1986. *Facts and Fallacies of Hawthorne: A Historical Study of the Origins, Procedures, and Results of the Hawthorne Illumination Tests and Their Influence upon the Hawthorne Studies*. New York: Garland.

This article has been cited by:

1. Michael Kosfeld, , Susanne Neckermann. 2011. Getting More Work for Nothing? Symbolic Awards and Worker Performance. *American Economic Journal: Microeconomics* **3**:3, 86-99. [Abstract] [View PDF article] [PDF with links]
2. Oriana Bandiera, , Iwan Barankay, , Imran Rasul. 2011. Field Experiments with Firms. *Journal of Economic Perspectives* **25**:3, 63-82. [Abstract] [View PDF article] [PDF with links]