Was There a Hawthorne Effect?[1]

Stephen R. G. Jones
McMaster University

The "Hawthorne effect" has been an enduring legacy of the celebrated studies of workplace behavior conducted in the 1920s and 1930s at Western Electric's Hawthorne Plant. This article examines the empirical evidence for the existence of Hawthorne effects using the original data from the Hawthorne Relay Assembly Test Room. Allowing for a variety of other factors, the author assesses whether experimental changes, variously defined, had a common effect that could be regarded as a pure result of the experimentation. The main conclusion is that these data show slender or no evidence of a Hawthorne effect.

I. INTRODUCTION

An enduring legacy of the celebrated studies of workplace behavior conducted at the Hawthorne Plant of the Western Electric Company in the 1920s and 1930s is the so-called Hawthorne effect. Variously defined, the central idea is that behavior during the course of an experiment can be altered by a subject's awareness of participating in the experiment. Though not obviously more than an incidental and intermediate finding for the early researchers, the Hawthorne effect has come to occupy a central role in the methodology of experiments and continues to have widespread influence in social science and management textbooks and research and especially in research in the psychology of education. It is somewhat surprising that, although efforts have been made to identify Hawthorne effects in various areas of field experimental research, there has been no systematic study of the evidence from the Hawthorne Plant itself. My object in this article is to undertake such a study, recognizing the mixed experimental and field nature of the Hawthorne research.

I first review the widespread and largely uncritical acceptance of the

[1] This research was supported in part by a grant from the Labour Studies Programme at McMaster University. I would like to thank participants in the McMaster Labour Lunch for comments on an earlier version of this work. I would also like to thank two anonymous referees for helpful suggestions. Correspondence may be addressed to Stephen R. G. Jones, Department of Economics, McMaster University, Hamilton, Ontario, Canada L8S 4M4.

0002-9602/93/9803-0001$01.50

idea of a Hawthorne effect and then examine the evidence from the Hawthorne studies themselves, with allowance for a wide range of direct experimental and incidental variables, for the role of replacement workers during the five years of study, and for potential interdependence of the workers' output levels. Whether the Hawthorne effect is defined in a narrow or a broad sense, the conclusion is the same: the original Hawthorne studies contain little clear evidence of a Hawthorne effect.

II. INFLUENCE OF THE HAWTHORNE EFFECT

Writing in the widely influential Festinger and Katz (1953) volume, John French noted that a potential merit of field experiments over laboratory experiments is that the former can avoid or minimize artificiality and thereby overcome the problem of generalizing results from the laboratory to real-life situations. French continued:

> That this is not always the case, however, is well illustrated in the famous Hawthorne experiment. From a methodological point of view, the most interesting finding was what we might call the "Hawthorne effect." In order to manipulate more precisely the physical factors affecting production, the experimenters had set up a special experimental room for a small group of girls who were wiring relays. This wiring was separated from the rest of the factory, and the girls working in it received special attention from both outside experimenters and the management of the plant. Careful studies of this wiring group showed *marked increases in production which were related only to the special social position and social treatment they received*. [1953, pp. 100–101; emphasis added]

However, French gave no reference for these careful studies and only cited the original Roethlisberger and Dickson (1939) account—that does not contain any statistical analysis beyond bivariate correlations—so that the exact foundation for his statement is unclear.

Notwithstanding this, there can be little doubt that the Hawthorne effect has entered into the literature as a key fact to be reckoned with in many practical contexts. William Whyte, for example, writes of the Hawthorne studies, "As experiment followed experiment . . . it became abundantly clear that physical changes were not the key. As in the earlier experiment, output did shoot ahead where conditions were changed, but so did output shoot ahead where no changes had been made. . . . The researchers came to the conclusion that output shot up in both groups because in both groups the workers' participation had been solicited and this involvement, clearly, was more important than physical perquisites. The workers were a social system; the system was informal but what it really determined was the worker's attitude toward his job" (1956, p. 34). Similarly, Ruch and Zimbardo write, "*No matter what the research-*

ers did, productivity went up. Even when work conditions were made worse than they were originally, the women worked harder and more efficiently" (1971, p. 372; emphasis added).

More recently, Blalock and Blalock similarly report that, "*Each time a change was made, worker productivity increased,* leaving the impression that each change had a progressive effect. As a final check, the experimenters returned to the original unfavorable conditions of poor lighting, no rest pauses, and no incentive system. Seemingly perversely, productivity continued to rise" (1982, p. 72; emphasis added). Some qualification is provided by Elmes, Kantowitz, and Roediger: "With few exceptions, no matter what changes were made—whether there were many or few rest periods, whether the work day was made shorter or longer, et cetera—the women tended to produce more and more telephone relays. . . . The workers knew that the experimenters expected the changes in working conditions to affect them, so they did" (1985, p. 225). As these sources reveal, the received wisdom is that there were Hawthorne effects at the Hawthorne Plant.[2]

In several areas of active research, the Hawthorne effect has come to be a significant preoccupation for many scholars. In education research in particular, since Desmond Cook's (1962) classic work and the subsequent contribution of Bracht and Glass (1968), there have been many studies that attempt to deal with Hawthorne effects. In their recent survey of 86 such studies, Adair, Sharpe, and Huynh (1989*a*) give a mixed overall assessment of this work: their metanalysis gives no grounds for a Hawthorne versus no-treatment control difference.[3] Yet the view that the original Hawthorne studies provide a firm foundation for the idea of a Hawthorne effect remains firmly entrenched in the literature. As Adair recently wrote, "The investigators began by changing the method of determining wages. During the experiment the investigators also manipu-

[2] In response to a referee, who fears that past errors may be reinforced through unwitting repetition, I should point out the major inaccuracies in the preceding quotations. The best single reference for careful corrections to these and other accounts is Gillespie's (1991) recent book *Manufacturing Knowledge,* in which a detailed archival study corrects many common misconceptions about events at Hawthorne, in addition to presenting a fascinating argument about the stabilization of conflicting interpretations in the process of "manufacturing knowledge." The women in the relay assembly room were not actually "wiring" relays, as French said (Gillespie 1991, chap. 2). Whyte's claim that physical factors were not the key is in part at odds with the evidence (see Gillespie [1991, pp. 38–48] on the illumination tests and Jones [1990] on physical effects on productivity in the relay room itself). Ruch and Zimbardo's claim that productivity went up no matter what the researchers did is incorrect; also incorrect are Blalock and Blalock's similar statement and the qualified description given by Elmes et al. (Gillespie 1991, table 4, p. 57).

[3] See also Adair, Sharpe, and Huynh (1989*b*) for further discussion of placebo, Hawthorne, and other controls in experimental research.

lated, on different occasions and sometimes concurrently, the length and timing of rest periods, the length of the work week, the length of the work day, and whether or not the company provided lunch and/or beverage. *Productivity seemed to increase regardless of the manipulation introduced* (1984, p. 336; emphasis added).[4]

Finally, I must mention the wider influence of the Hawthorne experiments and the received wisdom of Hawthorne effects. In a leading case of popular business writing from the 1980s, for example, Peters and Waterman 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')" (1982, pp. 5–6).

III. THE HAWTHORNE EXPERIMENTS

A. The Hawthorne Studies and the Hawthorne Effect

The Hawthorne experiments were conducted at the Hawthorne Plant of the Western Electric Company in the late 1920s and early 1930s and involved a variety of different studies of workplace behavior. The illumination experiments, which initially sought to establish a physiological relationship between intensity of illumination and workplace efficiency, predated the main Hawthorne studies themselves and showed that, in some instances, workers could maintain efficiency even under very low intensity of light, a finding that the researchers viewed as quite anomalous.[5] Indeed, Roethlisberger and Dickson (1939, p. 17) even mention a sequence of experiments in which an electrician pretended to alter lighting intensity—simply replacing bulbs by others of equal power—after which the women involved commented explicitly on their preference for the old or new illumination intensity. Overall, they concluded, these experiments "failed to answer the specific question of the relation between illumination and efficiency," but, nonetheless, "they provided great stimulus for more research in the field of human relations" (Roethlisberger and Dickson 1939, p. 18).[6]

[4] As with the earlier textbook quotations, Adair's claim that output increased no matter what the experimental change is incorrect (see, e.g., Gillespie 1991, table 4, p. 57).

[5] Roethlisberger and Dickson (1939, p. 17) cite the case of two "capable and willing operators" who maintained their productive efficiency even when the amount of light was cut to 0.06 of a footcandle, "an amount of light approximately equal to that on an ordinary moonlight night."

[6] Gillespie (1991, pp. 38–48) discusses the industrial illumination tests in some detail and shows that, following Roethlisberger and Dickson's (1939) account, the standard

For quantitative research, a more valuable aspect of the Hawthorne studies was the Relay Assembly Test Room studies, in which five women worked in a technologically independent way producing electrical relays.[7] This study covered 270 weeks, from April 1927 to June 1932, and involved 24 different "experimental periods" of varying length in which working conditions were changed, sometimes by conscious design of the researchers and sometimes out of practical expediency in the face of the declining 1930s economy. Table 1 details the timing of these periods and lists the principal changes made in each.

There is little direct evidence of a Hawthorne effect in the original research. Perhaps the clearest statement of what constitutes the effect was made by Roethlisberger and Dickson in reviewing the changes that occurred in the first seven periods of the Relay Assembly Test Room experiment: "There were those changes introduced by the investigators in the form of experimental conditions; these were well noted and recorded. There was another type of change, however, of which the investigators were not so consciously aware. This was manifested in two ways; first, in a gradual change in social interrelations among the operators themselves . . . ; secondly, in a change in the relation between the operators and their supervisors. . . . *From [the] attempt to set the proper conditions for the experiment, there arose indirectly a change in human relations which came to be of great significance in the next stage of the experiment*" (1939, pp. 58–59; emphasis added). Together with the primarily anecdotal evidence from the illumination studies, these observations, more than anything else in the major account of the Hawthorne

interpretation of the early lighting studies tends to "exaggerate the initial ignorance of the experimenters and misrepresent the process of discovery" (p. 47).

[7] There were actually eight women in the study overall, counting three "replacement" workers, who appear here, as elsewhere, labeled 1A, 2A, and 5A. In fact, workers 1 and 2 were replacements, since 1A and 2A started out in the relay assembly room group but were removed near the start of period 8. It should be pointed out that there is considerable disagreement in the literature over the appropriate interpretation of this replacement of 1A and 2A by 1 and 2. Franke and Kaul (1978), e.g., used it to construct an indicator of "managerial discipline," though such a view was subsequently disputed by responses to their work (Wardwell 1979, pp. 859–60); Schlaifer 1980, pp. 998–99, 1004–5). In the original log for January 25, 1928, the day the replacement occurred, Hibarger, a Hawthorne piece-rate analyst who was the supervisor/experimenter in the Relay Assembly Test Room, stressed that animosity had built up between workers 1A and 2A and some of the other women in the group, though other evidence suggests that these workers' decision to restrict output was a more important factor leading to their replacement (Gillespie 1991, pp. 61–63). The weeks when they "replaced" workers 1 and 2, and the two occasions when worker 5A replaced worker 5, are detailed in table 1 above.

TABLE 1

MAIN EXPERIMENTAL CHANGES IN HAWTHORNE RELAY ASSEMBLY TEST ROOM

Period	Weeks	Change
1	1–3	None; still in main plant
2	4–7	Move to test room; ↑ in VRest (from 0)
3	8–15	Small group incentive introduced, VRest ↑
4	16–20	Sched stop, Sched time (2 × 5 mins.); VRest ↓
5	21–24	Sched time ↑ (2 × 10 mins.), VRest ↕
6	25–28	Sched stop ↑, Sched time ↑ (6 × 5 mins.), VRest ↓
7	29–39	VRest ↑, Sched stop ↓, Sched time ↓ (15 and 10 mins.)
8	40–46	Raw materials problems, VRest ↓; 1 and 2 replace 1A and 2A
9	47–50	VRest ↓ (mostly), shorter day (1/2 hour less, 4:00 stop)
10	51–62	VRest ↑, return to full working day (4:30 stop)
11	63–71	VRest ↕, five-day week (from five and one-half days)
12	72–83	Days/week ↑, Sched stop and Sched time ↓ (to 0), VRest ↑
13	84–114	Sched stop and Sched time ↑ (15 and 10 mins.), VRest ↓
14	115–23	VRest ↓ (mostly); 5A replaces 5 in week 120
15	124–54	VRest ↓ (slightly)
16	155–58	Raw materials problems end, VRest ↕, seating change
17	159–83	Days/week ↓, VRest ↕; 5 returns to replace 5A
18	184–98	Days/week ↓, VRest ↑ (slightly)
19	199–210	VRest ↕, seating change (back)
20	211–38	VRest ↓
21	239–41	Days/week ↓, VRest ↓
22	242–50	VRest ↓
23	251–53	VRest ↓ (to 0), days/week ↓
24	254–70	Sched stop and Sched time ↓ (to 0), days/week ↓; 5A replaces 5

NOTE.—Changes relative to regular department include a smaller room, more uniform lighting, fans for use in summer, one layout operator for five women rather than for six or seven, a new chute mechanism, fewer relay types to assemble (in general), a new repairs procedure, a test room observer who "took over some of the supervisory functions" (Roethlisberger and Dickson 1939, p. 39), periodic physical examinations, and the freedom to talk more freely while working.

LEGEND.—An upward arrow (↑) = increase; a downward arrow (↓) = decrease; a dual-direction arrow (↕) = increases and decreases within the same time period; VRest = voluntary rest time; Sched stop = scheduled rest stops; Sched time = scheduled rest time; Days/week = working days per week.

experiments, seem consistent with the interpretation of the Hawthorne effect proposed by French.[8]

[8] The only additional evidence regarding Hawthorne effects in the early accounts is given in Whitehead's (1938) meticulous study. In his discussion of Graph H-42 (in vol. 2) that charts weekly group output together with a number of the main experimental variables over the five year study period, Whitehead asks, "Is there any clear indication that this behavior is very markedly influenced by the changes in experimental periods?" (vol. 1, p. 41). His answer is that, except for period 12, when rest pauses were eliminated and group output over the period as a whole fell "about 4% below its trend at that time, . . . changes in output rate do not correspond in time with changes in experimental periods for the most part" (vol. 1, p. 42). He is, however,

B. Interpretation of the Hawthorne Effect for Empirical Investigation

There are three potential views of the usefulness of the original Hawthorne data for studying the Hawthorne effect. The most restrictive of these holds that the general experimental effect is the key. Particular factors relating to the isolation of the group from the rest of the plant, especially the "special attention" accorded to its members—which may be a euphemism for less authoritarian but closer supervision—are critical, in this view, and since there was only one such test room, with no control study, the Hawthorne studies yield essentially one observation. On this reading, the original data can at best be suggestive but can never resolve the issue of Hawthorne effects.

A less restrictive view, and one that is clearly consistent with the standard interpretation placed on the Hawthorne data by many of the authors cited above, is that the Hawthorne effect is related to the explicit changes made at the start of experimental periods, changes that were, of course, known to the workers concerned. In this light, one can look for a common effect on output associated with *any* such experimental period changes, making allowance for other effects from any experimental or environmental variables.

Finally, the broadest interpretation is that the Hawthorne effect might reasonably be related to *any* changes in experimental conditions, not just to the major changes made at the beginning of each experimental period. Thus, one might also expect a change induced by a Hawthorne effect to result from a within-experimental-period change in working conditions. As with the second view, the research strategy that arises from the broad interpretation is to look for some common effect at times when any experimental variable changes, controlling for the experimental variables themselves. This article investigates the consequences of the second and third of these approaches, which are termed the "narrow" and the "broad" definition of an experimental change, respectively.

C. Data Description and Output Levels around Periods of Experimental Change

My investigation begins with simple data description and then moves to more involved statistical models of output determination in the Relay Assembly Test Room. Summary statistics on all the data used in this

---

quick to qualify this finding, noting that "a negative finding is never very convincing; it is always possible that some phenomenon remains undiscovered through a defect in the investigation" (vol. 1, p. 43). Whether my largely negative findings, which corroborate Whitehead's graphical analysis but are at odds with the established reading of Hawthorne effects, are convincing is, of course, for the reader to assess.

study are given in table 2.[9] The output variables for the eight women studied in the Relay Assembly Test Room, including the three replacements, are measured as the mean number of (standardized) relays produced per hour worked, averaged over a weekly measurement interval. Repair time and voluntary rest time are measured in minutes per week as is the general (non-person-specific) scheduled rest time. Small group pay is a dummy variable reflecting the introduction of a group piece rate based on the output of the five workers under study as opposed to pay based on the average output of the plant as a whole: this change was made in week 8 (see table 1 above). Other experimental variables with changes detailed in table 1 and summarized in table 2, were working days per week (hereafter days/week), the use of replacement workers, a change in the seating plan at the workbench, and the number of scheduled rest stops.[10] A change in any of these variables in a week will count as a *broad* experimental change, whereas a *narrow* change is limited to those groups of simultaneous changes listed in table 1. Overall, of the 270 weeks, 24 weeks had a narrow experimental change, while 88 weeks had a broad change.

The first set of results comes from a comparison of individual mean output levels overall—in the week preceding an experimental change, in the week of the change itself, and in subsequent weeks. This is an "unconditional" interpretation of the Hawthorne effect, as suggested by several of the authorities cited above. These figures are given in table 3,[11] and since different workers were present for differing periods and since some periods are dropped to avoid double counting, the sample size upon which the mean is based is given in each case.[12] For each worker, the results display little evidence of a simple Hawthorne effect. Relative to the week preceding the change, mean output falls slightly in the week

[9] The data set employed in this study is based on pioneering work by Franke and Kaul (1978) and Franke (1979, 1980), which I have extended using data from Whitehead (1938) (see Jones 1990). It is available on request to readers who supply a DOS-formatted diskette.

[10] I have given further information about the Relay Assembly Test Room and the variables constructed for this research elsewhere (Jones 1990).

[11] A minor complication in this calculation is that, especially for the broad definition of an experimental change, a given week might be "double counted" as being, say, two weeks after one change *and* one week before the next change, which would tend to make interpretation of the output levels in such a week difficult. I have accordingly omitted all such double counting from the averages recorded in table 3 and from the graphs of output changes in figs. 1 and 2. In doing this, priority is given first to the week preceding the change, then to the week of the change, and so on.

[12] The sample size rises moving from the week preceding a change to the week of a change simply because I count the first week of the relay room study as a "change" and there is no output data for the preceding week.

TABLE 2

SUMMARY STATISTICS FOR ALL VARIABLES

Variable*	N	Mean	SD	Minimum	Maximum
Output 1	218	69.5	4.5	60.1	80.8
Output 2	216	73.0	5.0	59.9	82.4
Output 3	256	63.2	4.6	43.7	72.9
Output 4	256	67.2	7.1	48.1	82.4
Output 5	204	59.8	5.7	47.7	69.8
Output 1A	39	51.6	3.1	44.4	56.4
Output 2A	38	52.0	3.2	45.6	56.8
Output 5A	52	55.7	5.4	43.4	64.2
Days/week	257	5.0	.59	3.0	5.5
Unemployment rate	270	8.8	6.6	3.2	23.6
Repair time 1	231	27.7	15.0	0	59.9
Repair time 2	231	22.5	11.1	0	42.7
Repair time 3	270	19.6	11.8	0	40.7
Repair time 4	270	11.6	7.9	0	30
Repair time 5	214	18.4	12.6	0	69
Repair time 1A	39	20.5	11.6	0	38
Repair time 2A	39	18.9	12.1	0	37.4
Repair time 5A	56	40.0	27.6	0	69
Voluntary rest 1	231	4.5	3.3	0	15
Voluntary rest 2	231	6.7	4.1	0	20.6
Voluntary rest 3	270	6.1	3.3	0	13
Voluntary rest 4	270	7.4	4.1	0	16.5
Voluntary rest 5	214	5.8	2.0	0	8.9
Voluntary rest 1A	39	9.2	4.6	0	13.7
Voluntary rest 2A	39	8.2	5.6	0	17
Voluntary rest 5A	56	3.3	2.1	0	6.4
Raw material problems	270	.15	.36	0	1
Small-group pay	270	.97	.16	0	1
No repair time reports	270	.14	.35	0	1
No voluntary rest reports	270	.09	.28	0	1
No scheduled rest reports	270	.06	.24	0	1
Scheduled rest stops	270	1.7	.91	0	6
Scheduled rest time	270	20.6	9.4	0	30
1A/2A replacement	270	.86	.35	0	1
5A replacement	270	.21	.41	0	1
Seating change	270	.16	.37	0	1
No medical reports	270	.42	.49	0	1
Worker 1 ill	270	.07	.26	0	1
Worker 2 ill	270	.06	.23	0	1
Worker 3 ill	270	.05	.22	0	1
Worker 4 ill	270	.05	.21	0	1
Worker 5 ill	270	.02	.15	0	1
Heat wave	270	.04	.21	0	1
Cold wave	270	.03	.18	0	1
Narrow experimental change	270	.09	.29	0	1
Broad experimental change	270	.33	.47	0	1

* Numbers following variables represent workers.

TABLE 3

MEAN OUTPUT LEVELS IN WEEKS AROUND EXPERIMENTAL CHANGES

	q1	q2	q3	q4	q5	q1A	q2A	q5A
Overall	69.5	73.0	63.2	67.2	59.8	51.6	52.0	55.7
N	218	216	256	256	204	39	38	52
Narrow definition (Δ):								
Δ(−1)	70.3	72.8	62.1	64.8	58.7	51.4	52.0	57.3
N	16	16	23	23	19	7	7	3
Δ(0)	69.6	71.3	61.3	63.9	58.4	52.3	53.7	56.1
N	17	16	24	24	21	7	6	3
Δ(1)	68.9	71.6	61.5	64.0	58.0	51.2	50.8	56.8
N	17	16	24	24	20	7	7	4
Δ(2)	67.4	71.1	61.0	63.5	57.1	50.9	51.1	58.5
N	14	14	19	19	17	6	6	3
Δ(3)	68.6	71.7	62.1	64.7	59.1	51.9	52.6	53.3
N	12	12	15	15	13	3	3	2
Δ(4)	68.3	71.8	62.7	66.6	59.7	51.1	53.2	56.6
N	12	12	14	14	12	2	2	2
Broad definition (Δ*):								
Δ*(−1)	69.8	73.3	62.4	66.6	59.0	51.0	51.7	57.0
N	63	63	81	80	63	18	18	17
Δ*(0)	69.2	72.4	62.7	65.9	59.1	52.1	52.2	56.3
N	34	33	44	44	35	10	9	9
Δ*(1)	69.5	72.9	63.3	66.3	60.2	51.7	52.0	53.8
N	27	26	33	33	28	6	6	5
Δ*(2)	69.1	72.6	63.3	66.4	59.1	52.9	53.0	51.9
N	19	19	23	24	20	5	5	4
Δ*(3)	70.4	73.9	64.4	69.1	61.6	...	...	55.7
N	14	14	14	14	11			3
Δ*(4)	69.0	72.4	64.3	68.3	61.1	...	...	52.8
N	11	11	11	11	9			2

NOTE:—Here, Δ(*n*) denotes the observation *n* weeks after a narrow experimental change; Δ*(*n*) is used analogously for the broad definition. Also, *N* gives the sample size for each mean. Ellipses indicate that no observations were available for this worker for this week. Double counting of weeks is avoided; see text for explanation.

of the change for each of the core group of workers on the narrow definition, but the pattern is mixed for the changes within the broad category. In both cases, however, the changes are numerically very small. For subsequent weeks, there is similarly no clear pattern, and the movement is once again slight. At conventional significance levels, one certainly could not reject the hypothesis that output had the same mean in each of the weeks surrounding the experimental change.

More generally, the pattern of output levels at *all* periods after an experimental change (and before the next such change) is graphed for the narrow and broad cases in figures 1 and 2, respectively. As can be seen

FIG. 1.—Mean output levels (narrow definition of experimental change)

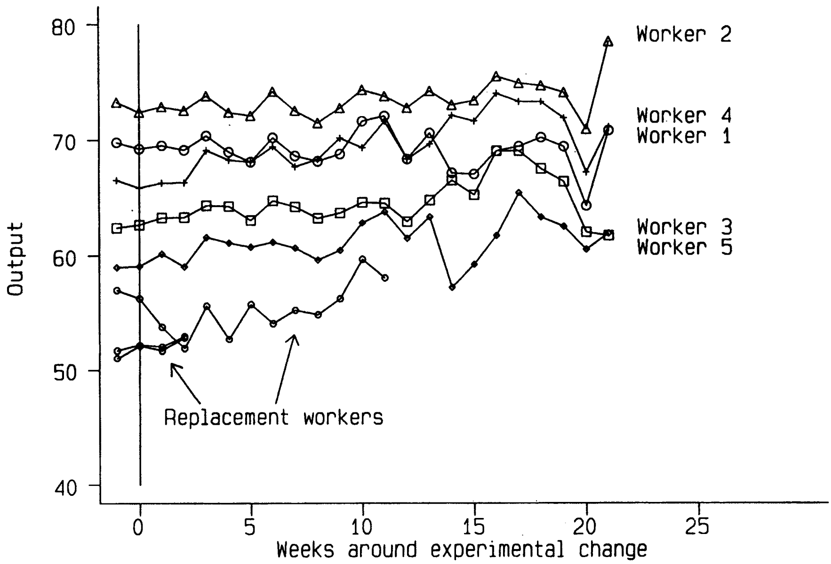

FIG. 2.—Mean output levels (broad definition of experimental change)

from figure 1, although there is considerable diversity in the mean output levels of the workers involved, especially including the three replacement workers who had particularly low average outputs, there is a clear trend only for worker 4, and that trend begins only in the fifth week after a narrow experimental change. For the broad definition in figure 2, worker 4 again has the strongest upward pattern in her mean output levels although, as with figure 1, in no case is there a significant regularity in these mean output data.

D. Models of Worker Behavior and the Hawthorne Effect

One potential problem with the results discussed above is that the experimental changes themselves, which are likely to have direct effects, may obscure any Hawthorne effects that might be present. To control for this, I next present the results of estimating the determinants of individual output levels, controlling for the direct effects of experimental changes.[13]

Two sets of specifications, or models, are employed. The first treats each worker's output as independent of the others, although potentially influenced by a set of common variables, and amounts to the estimation of

$$q_i = X_i b + Z c + u_i, \tag{1}$$

where X_i contains the person-specific variables from table 2, Z represents the variables common to all the workers, and u_i is a person-specific error term. An alternative approach, or model 2, allows for a sluggish response to experimental changes and a potential interdependence in the levels of output of the members of the working group by postulating

$$q_i = \lambda q_i(-1) + X_i b + Z c + \alpha \mathbf{q}_{-\mathbf{i}} + u_i, \tag{2}$$

where $q_i(-1)$ is the own lagged value of output and $\mathbf{q}_{-\mathbf{i}}$ is a vector of contemporaneous output variables for the other members of the working group. This second model follows one I developed earlier (Jones 1990).

[13] It should be noted that, while simple theoretical regression techniques were known by the time of the Hawthorne studies (Stigler 1986), instrumental variable methods (as are used for some of the models below) came to fruition after the Second World War. Also, it should be stressed that such techniques as were known at the time of the Hawthorne studies did not have wide currency among empirical researchers. Computing limitations were also undoubtedly severe, so that it is reasonable to conclude that the present statistical methods were not readily available at the time of the Hawthorne studies themselves. Nonetheless, there is clear value for present-day researchers to use such multivariate techniques in a reanalysis of the original data, essentially since such techniques permit the researcher simultaneously to control for a large number of factors that might otherwise be confounded with the Hawthorne effect.

In model 1, a single equation estimate is appropriate, while for model 2 the endogeneity of other workers' concurrent output levels means that an instrumental variable estimate is preferred: the instruments I use are lagged output levels and person-specific variables, such as voluntary rest or repair time. Finally, in view of the role of replacement workers in the Relay Assembly Test Room, I employed a variant of model 1 in which the left-hand-side variable was the output of worker 1A *or* worker 1 (and similarly for workers 2A and 2, 5A and 5), with all of the explanatory variables being interacted with the dummy variable representing the particular replacement in question. I refer to this as the "interacted model."

To assess the presence of a Hawthorne effect in the context of these models, equations (1) and (2) were estimated for each worker and included a dummy variable with the value of 1 in each week of experimental change (and 0 when no change occurred). Sample sizes varied for model 1 and for the interacted model depending on the length of time the worker spent in the group, although model 2 was always estimated on the 159 weeks when all five core group members were present and had been present in the preceding week. The estimated coefficients on these dummy variables are given in table 4.[14] For model 1, there is little evidence of a common (Hawthorne) effect in either the narrow or broad case; the pattern of signs is mixed with the only significant coefficients suggesting a negative Hawthorne effect, controlling for the other variables. For the interacted model, which only matters for the three workers who were replaced at some point in the five-year study, the results are equally checkered. Similarly, when allowance is made for partial adjustment and potential interdependence of the worker's output levels, the dummy variables representing an experimental change are uniformly insignificant. As with the unconditional results of table 3, there seems to be essentially no evidence of a Hawthorne effect.

E. Patterns of Residuals

The final diagnostic check I employ is to examine the pattern of residuals from equations (1) and (2) when estimated excluding the experimental change dummy variables. The presence of a marked pattern of such residuals in periods following a change could suggest some type of Hawthorne effect, perhaps of a form too subtle to be adequately captured by a single contemporaneous dummy variable. Accordingly, three of the specifications in table 4 (I am excluding the interacted specification) were reestimated and the residuals plotted. Figures 3 and 4 illustrate these results in which, for each number of weeks after an experimental change,

[14] Full results for these estimated equations are available on request from the author.

TABLE 4

ESTIMATES OF EXPERIMENTAL CHANGE EFFECTS

	*q*1	*q*2	*q*3	*q*4	*q*5
Narrow Definition:					
OLS specification 1	.93	−.74	−.09	−.72	.46
	(.73)	(.83)	(.49)	(.77)	(.58)
OLS interacted model...	1.00+	−.30	. . .	. . .	.37
	(.60)	(.73)			(.61)
OLS specification 2	.52	−.95+	.18	.56	.21
	(.59)	(.53)	(.51)	(.77)	(.56)
IV specification 2	.51	−.89	.29	.51	.59
	(.63)	(.56)	(.58)	(.81)	(.63)
Broad definition:					
OLS specification 1	.14	−.94+	−.09	−.32	−.91*
	(.47)	(.52)	(.34)	(.53)	(.40)
OLS interacted model...	.28	−.65	. . .	. . .	−.67
	(.42)	(.47)			(.42)
OLS specification 2	.43	−.49	.09	.19	−.50
	(.39)	(.35)	(.34)	(.51)	(.37)
IV specification 2	.37	−.40	.06	.14	−.36
	(.41)	(.37)	(.36)	(.52)	(.39)

NOTE.—Each entry gives the estimated coefficient on a dummy variable that takes the value "1" in each week of experimental change (0 otherwise) entered into eqq. (1) or (2). Control variables for specification 1 for worker i (i = 1, . . . , 5) were repair time of i, voluntary rest time of i, a dummy variable indicating if worker i was ill (but worked in the week in question), days per week, a dummy for raw material problems, scheduled rest time, scheduled rest stops, a dummy for the seating change, dummy variables for a heat wave or a cold wave (as reported for the week by the *Chicago Tribune*), and dummy variables for the absence of data on scheduled rest stops and for the absence of medical reports. In addition, replacement worker dummies were included for 5/5A (for i = 1, . . . , 4) and for 1/1A and 2/2A (for i = 3,4,5). Sample sizes for the OLS specification 1 were, respectively, 218, 216, 256, 256, and 204. For the interacted model, these controls were all interacted with the replacement dummy variable, as well as entering the equation themselves. Sample sizes were 257, 254, and 225, respectively. For specification 2, additional regressors were the lagged dependent variable and the contemporaneous output levels of the other four workers. In this case, all estimates are for the consistent core sample of 159 weeks as detailed in the text. Numbers in parentheses are SEs.

$^{+}P \geq .10$.

$^{*}P \geq .05$.

each person's residual is averaged.[15] In addition, for each such week, the average is computed across individuals (weighted by the number of such weeks that the individual was in the sample). The circles in these figures are the individual mean residuals while the triangles, joined by a line, are the (weighted) mean of these individual mean residuals.

[15] These calculations avoid all double counting of periods in the same way as for table 3 and figs. 1 and 2. See n. 11 above. In addition, since the models are only estimated for workers 1–5, and not for the replacements, the associated residuals are graphed only for the five core members of the group.

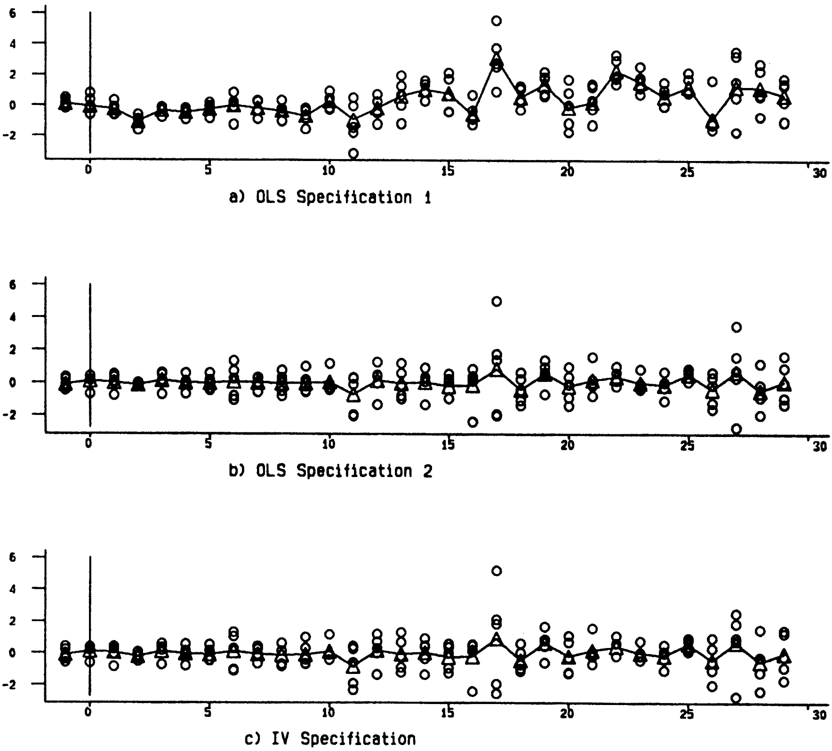

FIG. 3.—Residuals in weeks around experimental change (narrow definition of experimental change).

Figure 3 shows the results for the three specifications viewed through the narrow definition of an experimental change. Under OLS specification 1, the pattern seems to be of slightly negative residuals shortly after a change; these residuals are offset by some larger positive effects beyond 15 weeks after the experimental change. Of course, these latter effects are based on very small samples (since most experimental changes were followed by another change within 15 weeks) and are not significant.[16] In addition, it seems hard to construct a convincing scenario in which a Hawthorne effect would take over three months to develop. For the two other specifications, the pattern of residuals is still flatter, providing no evidence of any systematic error that could suggest a Hawthorne effect.

[16] For the narrow definition, there are four weeks when workers 1–4 record output levels 15 weeks after an experimental change; for worker 5, there are only three such weeks.

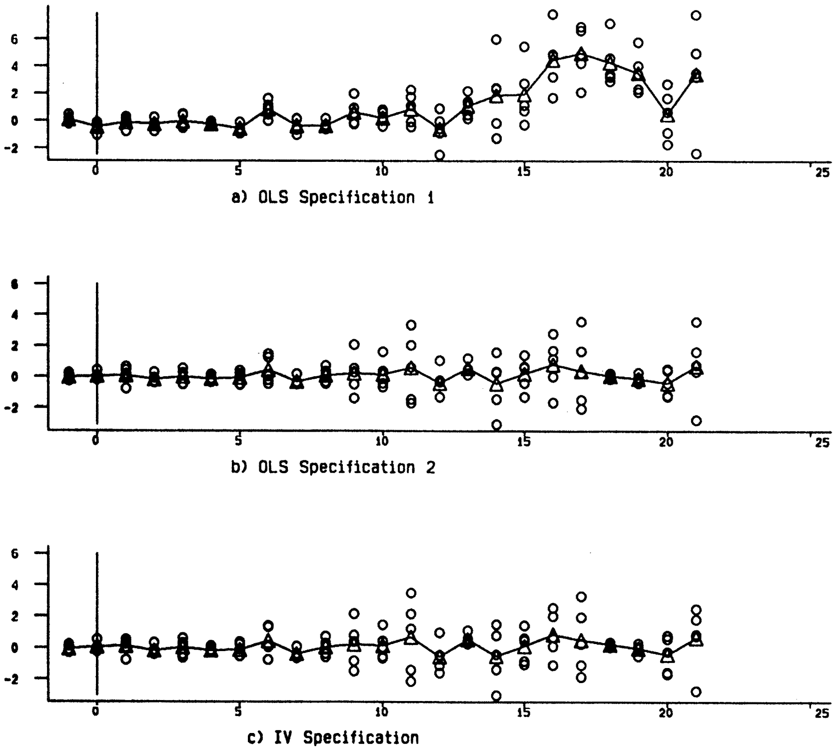

FIG. 4.—Residuals in weeks around experimental change (broad definition of experimental change).

The residuals in figure 4, which covers the three specifications for the broad definition of an experimental change, are very similar when averaged by week after the occurrence of an experimental change. Although the mean error is positive in OLS specification 1 from 15–20 weeks after a change, this is based on very small samples and is insignificant.[17] For the other two specifications, the point estimates seem to fluctuate very closely around zero, with no clear departure even many weeks after an experimental change when sample sizes become quite small.

IV. CONCLUSION

In this article, I have tested the evidence for a Hawthorne effect by examining the quantitative data on individual output levels collected

[17] In fact, for the broad definition, there is only one week when workers 1–5 record output levels 15 weeks after an experimental change.

over five years during the original Hawthorne studies. Contrary to the conventional wisdom in much research and teaching, I found essentially no evidence of Hawthorne effects, either unconditionally or with allowance for direct effects of the experimental variables themselves. My result appears to be robust across a wide variety of specifications, alternative samples, and two definitions of experimental change.

The one remaining interpretation of the Hawthorne effect that could survive my investigation is the first one I mentioned above, namely that the whole 270-week period of study was but one experiment, and that all of the various changes introduced at the start of the study and maintained throughout were one experimental change. Since we have no data on a control group, this interpretation means that there is, in essence, only one data point. Clearly, the original Hawthorne data are not adequate to the task of assessing this interpretation of what constitutes a Hawthorne effect. However, it is important to note that established interpretations of the Hawthorne studies (e.g., French 1953; Whyte 1956; Blalock and Blalock 1982; Adair 1984) never define the Hawthorne effect in this manner. In this context, I must conclude that there is slender or no evidence of a Hawthorne effect in the Hawthorne Relay Assembly Test Room. Finally, in light of these results, I must also conclude that the Hawthorne effect is largely a construction of subsequent interpreters of the Hawthorne experiments. A fruitful line of sociological enquiry, already initiated in part by Gillespie (1991), would explore the social and historical context of the reception of the Hawthorne experiments and the process whereby the Hawthorne effect has become enshrined as received wisdom in the social sciences.

REFERENCES

Adair, J. G. 1984. "The Hawthorne Effect: A Reconsideration of the Methodological Artifact." *Journal of Applied Psychology* 69:334–45.

Adair, J. G., D. Sharpe, and C. L. Huynh. 1989*a*. "Hawthorne Control Procedures in Educational Experiments: A Reconsideration of Their Use and Effectiveness." *Review of Educational Research* 59:215–28.

———. 1989*b*. "Placebo, Hawthorne, and Other Artifact Controls: Researchers' Opinions and Practices." *Journal of Experimental Education* 57:341–55.

Blalock, A. B., and H. M. Blalock, Jr. 1982. *Introduction to Social Research*. Englewood Cliffs, N.J.: Prentice-Hall.

Bracht, G. H., and G. V. Glass. 1968. "The External Validity of Experiments." *American Educational Research Journal* 5:437–74.

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

Elmes, D. G., B. H. Kantowitz, and H. L. Roediger III. 1985. *Research Methods in Psychology*. New York: West Publishing.

Festinger, L., and D. Katz, eds. 1953. *Research Methods in the Behavioral Sciences*. New York: Holt, Rinehart & Winston.

Franke, R. H. 1979. "The Hawthorne Experiments: Re-view." *American Sociological Review* 44:861–67.
———. 1980. "Worker Productivity at Hawthorne (Reply to Schlaifer)." *American Sociological Review* 45:1006–27.
Franke, R. H., and J. D. Kaul. 1978. "The Hawthorne Experiments: First Statistical Interpretation." *American Sociological Review* 43:623–43.
French, J. R. P. 1953. "Experiments in Field Settings." Pp. 98–135 in *Research Methods in the Behavioral Sciences,* edited by L. Festinger and D. Katz. New York: Holt, Rinehart & Winston.
Gillespie, R. 1991. *Manufacturing Knowledge: A History of the Hawthorne Experiments.* New York: Cambridge University Press.
Jones, S. R. G. 1990. "Worker Interdependence and Output: The Hawthorne Studies Reevaluated." *American Sociological Review* 55:176–90.
Peters, T. J., and R. H. Waterman. 1982. *In Search of Excellence: Lessons from America's Best-run Companies.* New York: Harper & Row.
Roethlisberger, F. J., and W. J. Dickson. *Management and the Worker.* Cambridge, Mass.: Harvard University Press.
Ruch, F. L., and P. G. Zimbardo. 1971. *Psychology and Life,* 8th ed. Glenview, Ill.: Scott Foresman.
Schlaifer, R. 1980. "The Relay Assembly Test Room: An Alternative Statistical Interpretation." *American Sociological Review* 45:995–1005.
Stigler, S. M. 1986. *The History of Statistics: The Measurement of Uncertainty before 1900.* Cambridge, Mass.: Harvard University Press.
Wardwell, W. I. 1979. "Critique of a Recent Professional 'Put-Down' of the Hawthorne Research." *American Sociological Review* 44:858–61.
Whitehead, T. N. 1938. *The Industrial Worker: A Statistical Study of Human Relations in a Group of Manual Workers,* 2 vols. Cambridge, Mass.: Harvard University Press.
Whyte, W. H., Jr. 1956. *The Organization Man.* New York: Simon & Schuster.