

The Deliberative Poll

ried out on "panel attrition," each pursuing the question of who drops out of panel studies in successive waves. I have done these analyses for both two- and three-wave panels in two different countries (the US and France), and with two different types of subject matter—electoral behavior and perceived quality of life. The main finding is monotonous in all these analyses: Variables in the first wave of the study give almost no prediction as to who will and will not give second and third interviews. For the quality-of-life topic, the decision seems essentially orthogonal to any of the variables in the study. For electoral behavior studies this is almost so as well, although there is one small but persistent exception. The repeaters report slightly more political involvement than those who drop out, and a diluted version of the effect—typically not significant—turns up for the few other variables strongly correlated with involvement, such as education. This is probably the most obvious prediction that one might make as to who would choose to repeat a political interview that can be experienced by the politically uninterested as embarrassing, or boring, or both. Again, the key fact is that the discrepancy between repeaters and the parent sample is very small.

Of course, choosing not to continue panel participation after a wave or two is not the same as deciding not to come to Austin, Texas, for a weekend of further discussion of political issues. There are, among other things, major differences in the governing parameters; the time lapse between my panel waves has always been at least a year, and as long as four years; the lapse between first interview and convention attendance is a matter of a few weeks. Also, the rate of attrition is much greater for the convention weekend—as found in the British experiment—than for the later waves of the panel—twice or more as great. But the similar and (to me) counterintuitive general findings of next-to-no-bias in the two types of cases seems to be of the same cloth. I have resisted such findings for many years, and pushed them in more ways than I have related here, such as examination of contrasts in higher-order interactions; but to no avail. Hidden sources of self-selection that might corrupt inference continue to be a worry; but I think the burden of proof in these matters now has shifted to those who believe that large-scale attrition can only mean the advent of damaging self-selection bias for the key variables of the study.

I have been addressing criticisms and possible problems. I should close by noting that the design being used in Jim Fishkin's project has many positive features: Resources will be devoted to one or two longitudinal follow-ups of the convention-goers much later in the campaign, to assess the durability of whatever changes are measurable at the end of that event. (Such assessments can be done, of course, for each of the 30-group replicates as well as for the total sample, if group-dynamic variability is of interest.) Another lively possibility is that one of these later follow-ups might re-interview those who did not go to Austin, as a kind of poor-man's control group, an idea which has appeal mainly if the decision to make the trip is in fact uncorrelated with other variables in the study. Finally, the goal is to make the data surrounding the study available to other researchers just as soon as it is feasible.



Philip E. Converse, long distinguished through his work at the University of Michigan's Institute for Social Research, Chairs the Fishkin project's Technical Review Committee

The Hawthorne Effect is a Common Artifact in Social Research

By John G. Adair

A common concern in social research is whether research participants are behaving normally, as they would if the "microscope" of the social scientist weren't focused on their behavior. Such changes in behavior that contaminate social research are commonly referred to as Hawthorne effects. This methodological artifact derives its name from the research project in which it was first observed that participants' awareness of being in an experiment might influence their behaviors and alter the data for the experiment.

The Birth of an Artifact

The Hawthorne studies, having launched the human relations movement in industry and provided the name for this well-known research artifact, are a social science classic.¹ The research took place at the Western Electric Company's mammoth Hawthorne Works. The company manufactured telephone equipment to supply the Bell System. Although studies designed to find the optimal level of lighting for the

factory were begun in 1924, the Hawthorne Effect probably owes its discovery to a five-year experiment, called the Relay Assembly Test Room Study, that began in 1927.² Five female employees who spent each work day as-

“*The primary contribution of the Hawthorne studies then was to demonstrate how easily, and without the researchers' awareness, an important research project could be contaminated and the data and phenomena distorted simply by social conditions within the experiment.*”

sembling relays were separated from their department and placed into a special test room where all relevant variables could be controlled or evaluated. The study was designed to find the optimal cycle of rest and work periods. However, to make the subjects sensitive to the experimental manipulations, the investigators changed the method of determining wages to base it on the performance of this group of five. During the experiment the investigators also manipulated, 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. Yet as the experiment progressed, productivity seemed to increase regardless of the manipulation introduced. Finally, well into the second year, the investigators decided to discontinue all treatments and return the workers to full work days and weeks without breaks or lunches. Unexpectedly, rather than dropping to pre-experiment levels, productivity was maintained! Obviously, it was concluded, the workers' behavior was influenced by the effects of some other variables—variables that the investigators had unwittingly manipulated.

Set in an era when the intrusive time-and-motion study dominated industrial research, the Hawthorne investigators had feared that employees taken from their regular work to be placed in a test room would be resistant to their experiment. To overcome this anticipated negative reaction, supervision was removed, special privileges were allowed, and considerable attention paid to designated workers. All these changes were intended to provide a controlled experiment. The researchers were forced to conclude, however, that it was these manipulations that had caused the subjects to improve their overall productivity. Thus the Hawthorne Effect was born.

Controlling for a Hawthorne Effect

In spite of its applicability to research in general, over the years the artifact has become of particular con-

cern to educational researchers, program evaluators, and others who conduct experiments in naturalistic settings. In educational environments, for example, studies introducing computers, special experimental curricula, or novel teaching methods are thought to be especially vulnerable to artifactual effects. At one time, it was thought that such biases may extend to "fully half the classroom study conclusions," so that many research methods texts lead students of education to expect that "many...experiments report changes and improvements that are due to the Hawthorne Effect."³

These textbooks, however, provide little guidance concerning the factors the contemporary researcher should guard against. Nevertheless, educational researchers regularly use Hawthorne control groups which exemplify the three components of the Hawthorne Effect. The participants receive special attention, are given a novel (yet meaningless) task, and are told that they are the subjects of an experiment but not given the full experimental treatment. It is these three conditions that specify the variables defining the Hawthorne artifact.⁴ Such control groups give the investigator knowledge about whether Hawthorne effects might operate within their study, but no control over its effects. Much like placebos given in medical experiments, these pseudo-experimental manipulations are created to check on the presence of the Hawthorne Effect. If the researcher finds no evidence of this effect, he or she can claim that the study is free of artifact. But to make that claim the researcher must provide reasonable assurances that Hawthorne effects were adequately tested.

Measuring a Hawthorne Effect

Unfortunately, there is no scientific knowledge of the proper placebo for a Hawthorne control. Various manipulations are introduced based solely on the researchers' judgment of what is appropriate. My colleagues and I collected all educational studies employing Hawthorne (and no-treatment) controls and analyzed them to determine the size of the effect associated with the three

components of the artifact.⁵ The results were disappointing—no control procedure was more effective than any other and the overall impact of Hawthorne effects on these education studies was essentially zero. No measurable Hawthorne effect was generated by these manipulations. The Hawthorne researchers and subsequent social methodologists have not defined the artifact precisely enough. Moreover they have not provided adequate procedures to enable us to measure or control its results. Nonetheless, the Hawthorne artifact continues to hang like a cloud over all social research.

The primary contribution of the Hawthorne studies then was to demonstrate how easily, and without the researchers' awareness, an important research project could be contaminated and the data and phenomena distorted simply by social conditions within the experiment. Some forty years after this insight, researchers working within the context of the psychological laboratory have empirically documented researchers' and participants' artifactual contribution to the research process.⁶ Unlike the Hawthorne studies, this work views participants in social research as cooperative, actively engaged in solving the problem of how to behave in a task that may be only partially explained by a researcher who may have some hidden objectives. This work has found that subjects, sensitive to cues in the research context, formulate their own hypotheses for the study, and this, rather than the experimenter's hypothesis, can compellingly guide their behavior.⁷

Researchers have also been shown to unwittingly convey their outcome expectations to participants.⁸ For example, although identical instructions were read to subjects by all experimenters, significant differences were obtained in the ratings of neutral photographs, based solely on the experimenter's expectation for positive or negative ratings. Further research has shown that the researchers' tone of voice subtly and inadvertently conveyed their expectations for subjects' performance on this simple task. Within more complex stud-

The Deliberative Poll

ies very little researcher-participant bias is required to produce significant amounts of artifact. These examples from the laboratory apply equally to field experimentation and expand our understanding of artifact in social research.

Fishkin Meets Hawthorne

The desire in any social experimentation such as the Fishkin experiment is for the research participants to behave veridically, independently, and normally. The researcher has designed the project so that these conditions, and hence the goals of his research, will be realized. Yet the three components of the Hawthorne Effect may be inescapably present. Participants in the Fishkin experiment, like the five specially-selected relay assembly test-room workers, have been selected and separated from their peers to participate in an experiment—a highly publicized, novel social experiment. Like the Hawthorne participants, they are receiving abun-

dant special attention—a free trip, national media attention, and an enormous boost to self-esteem. Not only the social scientist's "microscope," but America's television sets will be focused on their behavior. They know this is a novel experiment, and that its success depends on their behavior. They will be highly sensitive to cues to guide their responses. Will the evidence they are to judge be truly balanced, or will subtle expectancies be transmitted? In short, will they respond normally "on stage," independent of any biases, or be susceptible to the pressures known to produce social artifact? These are questions one must consider in evaluating the Fishkin experiment.

Endnotes

¹ E. Mayo, *The Social Problems of an Industrial Civilization* (Cambridge, MA: Harvard University Press, 1945).

² F.J. Roethlisberger and W.J. Dickson, *Management and the Worker*, (Cambridge, MA: Harvard University Press, 1939).

³ See D. Cook, "The Hawthorne Effect in Educational Research," *Phi Delta Kappan*,

Vol. 44, 1962, pp. 116-122; and W.R. Borg and M.D. Gall, *Educational Research: An Introduction* (3rd ed.) (New York: Longman, 1979), respectively.

⁴ J.G. Adair, D. Sharpe, and C.L. Huynh, "Hawthorne Control Procedures in Educational Experiments: A Reconsideration of Their Use and Effectiveness," *Review of Educational Research*, Vol. 59, 1989, pp. 215-228.

⁵ *Ibid.*

⁶ J.G. Adair, *The Human Subject: The Social Psychology of the Psychological Experiment*, (Boston: Little Brown, 1973).

⁷ M.T. Orne, "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications," *American Psychologist*, Vol. 17, 1962, pp. 776-783.

⁸ R. Rosenthal, *Experimenter Effects in Behavioural Research*, (New York: Appleton-Century-Crofts, 1966).



John G. Adair is professor,
Department of Psychology
University of Manitoba
Winnipeg, Canada

Groups are Unpredictably Transformed by Their Internal Dynamics

By R. Scott Tindale

Small groups are used in our society in many contexts. Juries, school boards, planning commissions, cabinets, advisory councils, etc., all play central roles in the institutions for which they are formed. The prevalence of small groups in the US is partly a function of political ideology: As compared to lone individuals, they provide a better representation of constituency interests and can encompass a wider diversity of opinion. However, they are also seen as effective for solving problems, making important decisions, and the like. Although there is an abundance of evidence supporting the effectiveness of groups, there is also a growing body of research showing that using small groups can sometimes lead to unexpected, and occasionally problematic, outcomes.¹ As James Fishkin embarks on his delibera-

“
Recent research has shown that faulty information or biased information processing strategies can produce polarization toward incorrect or biased positions. Thus, group discussion does not always lead group members to hold more “reasoned and informed” positions on an issue.
”

tive poll, recognition of the potential for such unexpected outcomes in relation to the planned small group discussions leads to questions about what can be generalized from the results.

As an initial caveat, I should point out that Fishkin's use of small groups is somewhat outside the range of the types of groups that have received the majority of research attention. Although information exchange and social influence are integral parts of small group dynamics in most contexts, rarely are such things the sole purpose of the group task. For example, jury members exchange ideas about evidence and attempt to influence each other, but such processes are invoked for the purpose of reaching a unanimous verdict. More recently, focus groups have been used to generate information, but rarely are the consumers of such information the group members themselves. Probably the clos-