

3-6-1980

An Evaluation of Stanley Milgram's Experiments on Obedience to Authority

Steven Alan Samson
Liberty University, ssamson@liberty.edu

Follow this and additional works at: https://digitalcommons.liberty.edu/gov_fac_pubs



Part of the [Other Social and Behavioral Sciences Commons](#), [Political Science Commons](#), and the [Public Affairs, Public Policy and Public Administration Commons](#)

Recommended Citation

Samson, Steven Alan, "An Evaluation of Stanley Milgram's Experiments on Obedience to Authority" (1980). *Faculty Publications and Presentations*. 273.
https://digitalcommons.liberty.edu/gov_fac_pubs/273

This Article is brought to you for free and open access by the Helms School of Government at Scholars Crossing. It has been accepted for inclusion in Faculty Publications and Presentations by an authorized administrator of Scholars Crossing. For more information, please contact scholarlycommunications@liberty.edu.

AN EVALUATION OF STANLEY MILGRAM'S
EXPERIMENTS ON OBEDIENCE TO AUTHORITY

Steven A. Samson

Sociology 560

March 6, 1980

Stanley Milgram's studies of obedience to authority, which began in 1960 and continued for several years, provoked considerable controversy when the results first began to be published. The experiments showed that a high proportion of naive subjects were willing, in what they believed was a learning experiment, to send apparently severe electrical shocks into the bodies of supposed learners at the instruction of a supposed research psychologist. If this description sounds far-fetched, it is only half the story, and the ironies are the most interesting part of it.

Critics challenged the design and the ethics of the experiment. They did not need to be picky. How can you conduct an experiment without a formal control group, without even a pretest? Is it proper to design a realistic but false experiment in order to conduct a somewhat defective but genuine experiment? The current code of ethics within the psychology profession would no longer permit an experimenter to induce subjects to commit acts of evident harm to others. So it is no wonder that the experiments have not been replicated elsewhere, despite their provocative results. These ironies are compounded by the experimenter's dependence on the same kind of expectations about authority he sought to measure in an experimental situation. The entire experiment could be criticized as the moral equivalent of entrapment. These problems will be considered in more detail later.

Milgram wrote two papers on the experiments which will be examined here. The first, "The Behavioral Study of Obedience" (1963), describes a single experiment conducted at Yale

University. The second, "Some Conditions of Obedience and Disobedience to Authority" (1965), described a series of experiments in which equivalent experimental groups, each with fresh subjects, were tested under a variety of conditions. The variation of experimental conditions made comparisons possible between the groups and substituted for the use of a formal control group.

Description

The subjects were adult males, aged 20 to 50 years, residing in the greater New Haven and Bridgeport areas, and engaged in a variety of occupations. Each experimental condition used 40 fresh subjects. Each group was matched, or blocked, according to two criteria: age and occupation. Milgram did not discuss randomization procedures. Subjects were obtained by a newspaper advertisement and direct solicitation by mail in what was apparently a two-step process. This is only one instance of Milgram's lack of precision in his description, however; too much is left to conjecture. For instance, Milgram made only passing reference to his first paper in his second paper, although the criteria for selecting the groups was the same. Perhaps he regarded the first experiment as a pilot study. But the first paper contained a more detailed account of the experimental procedure, though differing in minor details.

In the first study, 20% of the subjects ranged in age from 20-29 years, 40% from 30-39 years, and 40% from 40-50 years. This was held constant for each experimental group. Occupations were represented as follows: 40% were skilled or unskilled

workers, 40% were sales, clerical, or other white collar workers, and 20% were professional people. There was slight variation from the first to the second study in this category. Milgram did not provide any rationale for selecting these ratios or for changing certain details. One gets the feeling at times that Milgram was using a scattershot approach because he had no idea what variables would prove relevant.

Milgram describes his general laboratory procedure as follows:

The focus of the study concerns the amount of electric shock a subject is willing to administer to another person when ordered by an experimenter to give the 'victim' increasingly more severe punishment. The act of administering shock is set in the context of a learning experiment, ostensibly designed to study the effect of punishment on memory. Aside from the experimenter, one naive subject and one accomplice perform in each session. On arrival each subject is paid \$4.50. After a general talk by the experimenter, telling how little scientists knew about the effect of punishment on memory, subjects are informed that one member of the pair [i. e., subject and accomplice] will serve as teacher and one as learner. A rigged drawing is held so that the naive subject is always the teacher, and the accomplice becomes the learner. The learner is taken to an adjacent room and strapped into the 'electric chair.'¹

Several observations may be made about the procedure so far. The experimenter was not Milgram himself. The number of sessions numbered in the hundreds, perhaps the thousands, judging from the identification numbers. The experiment was based on such a complex series of deceptions that it is amazing that the experiment remained credible and that the same procedures could be repeated time and again. Milgram claims that the debriefings revealed that the subjects did not suspect the deception, but reliance on a self-report in a conflict situation is inherently risky. The complexities

and built-in ironies of the experiment raised more intriguing possibilities than Milgram could hope to control for.

The naive subject is told that it is his task to teach the learner a list of paired associates, to test him on the list, and to administer punishment whenever the learner errs in the test. Punishment takes the form of electric shock, delivered to the learner by means of a shock generator controlled by the naive subject.²

Milgram built the shock generator and had an engraver do the labeling. Evidently it worked well enough because the experimenter gave the subject a sample shock prior to commencing the procedure. Milgram paid close attention to details that would add to the realism of the situation: a fact which makes his omissions puzzling.

The teacher is instructed to increase the intensity of electric shock one step on the generator on each error. The learner, according to plan, provides many wrong answers, so that before long the naive subject must give him the strongest shock on the generator. Increases in shock level are met by increasingly insistent demands from the learner that the experiment be stopped because of the growing discomfort to him. However, in clear terms the experimenter orders the teacher to continue with the procedure in disregard of the learner's protests. Thus, the naive subject must resolve a conflict between two mutually incompatible demands from the social field.³

Milgram indicated that the responses of the "victim" (the paper contains a great deal of such gallows humor) were standardized on tape. Some of the later experimental conditions, however, brought teacher and learner into increasingly close proximity, and voice cues were provided. The shock generator had 30 clearly marked voltage levels ranging from 15 to 450 volts, each of which was activated by an individual switch. The generator also bore verbal designations ranging from "Slight Shock" to "Danger: Severe Shock." The last three switches did not bear any verbal designation.

This brief description will be elaborated on in the sections that follow. The experiments consisted of a series of variations on this one theme.

Theory

Milgram only briefly considered theoretical issues in his two papers. Some of the flaws in the experimental design may be attributed to the theoretical near-vacuum in which he was operating. Not much was known about the nature of obedience. Instead of providing a definition, Milgram merely observed in the first study that

Obedience is the psychological mechanism that links individual action to political purpose. It is the dispositional cement that binds men to systems of authority. Facts of recent history and observation in daily life suggest that for many persons obedience may be a deeply ingrained behavior tendency, indeed, a prepotent impulse overriding training in ethics, sympathy, and moral conduct.⁴

These "facts of recent history" referred explicitly to the Holocaust. Milgram had a major philosophical problem in mind and this may explain the impression that he attempted to drain the experiment of every possible significant detail. Nobody was more surprised by the results of these experiments than Milgram was with his first pilot studies. Initially, he predicted that subjects would generally balk at a certain point in the experiment. At that time, he felt compelled to introduce many of the features that became part of the experimental procedure because there was little resistance demonstrated by the subjects. They simply followed orders. Even mild pretests by victims proved inadequate. Finally, Milgram settled on a series of increasingly vehement pretests up to the 300 volt level. After that level, all response from the learner-

accomplice ceased. Milgram expressed dismay that most of the subjects continued the "treatment" (26 out of 40 subjects in the first study).

The absence of good theoretical definitions weakened the connection Milgram could make between the experiment and real-life conditions. The results are not generalizable in any precise sense. The whole construct validity problem is left unaddressed. Even worse, the first study did not state any explicit hypotheses.

At most, a partial theoretical foundation had been laid in earlier studies by Max Weber, Hannah Arendt, Theodor Adorno, Milton Rokeach, and others. Milgram's references to such studies, however, were minimal. Apparently they did not suggest hypotheses which he could test. Milgram's procedure resembled Solomon Asch's experiments in group pressure but Milgram did not acknowledge any debt.

He defined the problem he wished to examine more precisely in the second paper.

In its most general form the problem may be defined thus: if X tells Y to hurt Z, under what conditions will Y carry out the command of X and under what conditions will he refuse?

Milgram thus distinguished between "obedient" and "defiant" subjects. These terms were operationalized according to whether or not the subject carried out instructions. Milgram disclaimed the possibility of generalizing these terms outside the experimental setting. Milgram did not state any hypotheses regarding expected behavior for each variation in the experimental conditions. He did describe obedience as a three-person relationship between an authority, executant, and victim,

which he illustrated with the story of Abraham and Isaac. By doing so, he placed authority and obedience into a context of conflict. A critic might be justified in calling this a philosophical exercise rather than a true experiment. It is easy to conclude that Milgram was testing unformulated assumptions about human nature rather than a formal set of hypotheses. If he had any original expectations, he kept them to himself.

As with several other facets of the experiment, however, Milgram did insert a clever substitute for the missing element. Lacking hypotheses, he substituted predictions by informed observers. In the first study, Milgram selected fourteen seniors in psychology from Yale to predict how many subjects would continue to follow orders even in the face of protests by the learner-accomplices. Their predictions were similar to those obtained from full professors of psychology in some of the later experiments: the respondents greatly underestimated the willingness of subjects to follow orders. This appears to add yet another dimension to the experiment that is not explicitly connected to the purpose, rationale, or design of the project.

Milgram's scattershot approach at least had the useful effect of making discoveries that challenged prevailing assumptions about behavior. The papers are written in a crisp, authoritative manner that makes a positive first impression. It is easy to overlook the flaws because of the weight of analysis, which is quite imaginative. Milgram's ability to draw convincing conclusions from his data testifies both

to his persuasiveness and to his careful elimination of alternative interpretations of the data.

Design

Several problems are apparent, but most are compensated for in some way. The biggest problem is the complexity of the design. It was an ambitious project to begin with and it is matched by a design that is as convoluted and filled with subplots as a mystery novel, and is just as deceptive in appearance. Some of Milgram's improvisations substituted for some of the usual controls.

Milgram made no use of a pretest. Yet it is hard to see how a pretest could be devised that would be salient and materially related to the rest of the experiment without giving it away. Milgram might have varied the reception given the subjects before the experiment or the means by which they were compensated, but the effects of these variations could be expected to be minimal while further complicating the design.

The papers did not identify any control groups, but again it is difficult to tell how useful a purposely identified control group might have been, or even what might have been controlled. There was a built-in control mechanism in that each experimental condition was a variation on the original model, which was described in the first paper. Milgram varied only one specific element in each condition. It is possible to compare the experimental groups with each other because the procedures were standardized and the selection process was replicated from one group to another. This assured a degree of internal validity. Matching, or blocking,

guaranteed replicability.

To some extent, Milgram also controlled for other threats to internal validity. He tested the setting effect and the interviewer effect by deliberately introducing these factors into separate experimental conditions. Regarding the first, Milgram stated that

One must always question the relationship of obedience to a person's sense of the context in which he is operating. [*Italics in the original*].

To explore the problem we moved our apparatus to an office building in industrial Bridgeport and replicated experimental conditions, without any visible tie to the university.⁶

Milgram found that the level of obedience at Bridgeport was not significantly lower than that obtained at Yale. Milgram might have made other variations in the location, atmosphere, dress and manner of the experimenter, but at least he addressed the setting problem. He does not mention financial considerations, which had so restricted his original research, in this study. The irony is that he was so dependent on the symbols of authority--the university, the profession of psychology, his position as a professor, support from the National Science Foundation and the Higgins Fund--in order to study the effect of authority on obedience.

Milgram tested the interviewer effect, in a sense, when he varied the proximity of the experimenter to the subject. He found that the physical presence of an authority figure was an important force contributing to the subject's obedience or defiance. Likewise with the presence of the victim.

As the victim is brought closer, the subject finds it harder to administer shocks to him. When the victim's position is held constant relative to the subject, and the authority is made more remote, the subject finds it

easier to break off the experiment. This effect is substantial in both cases, but manipulation of the experimenter's position yielded the more powerful results.⁷

From this, Milgram concluded that "obedience to destructive commands is highly dependent on the proximal relations between authority and subject."⁸

As far as confounding variables are concerned, then, the design of the experiment was uneven. Maturation was not a question, since it was a one-shot ordeal for the subjects. The papers did not indicate whether there were any significant differences between age groups or occupation groups. This was an unfortunate omission. The absence of a formal control group was a drawback. Random assignment was not used. Prospective subjects were matched according to specific attributes. The fact that this was an experiment was clear to the participants, even though the nature of the experiment was not divulged, so that the setting effect could not be eliminated. But variations in the setting made little difference. Pilot studies compensated for the lack of a pretest to some extent. Blinding was not used, as far as the experimenter and his accomplice were concerned, but neither individual used in the experiments was a professional psychologist, or professional actor. Their actions were carefully standardized during the pilot studies and varied selectively. Blinding, instead, operated at other levels: Milgram evidently was not physically present and the purpose of the experiment was not divulged. Indeed, Milgram deliberately created a setting effect, turning it into an element of the experiment in the guise of a learning experiment.

Measurement

By standardizing the elements of the procedure, Milgram made the experimental groups, for all purposes, interchangeable. The shock generator was scaled in such a way that the subject could refuse to depress the next higher switch at any point in the experiment. A sequence of four standardized "pretests" could be used by the experimenter to bring a balky subject in line before terminating the experiment. They ranged from a simple request to an outright demand that the subject continue.

One variable was the point at which pretests from the victim began. In the first study, the victim was enclosed in a separate room and pounded on the wall at shock levels 300 and 315, but was afterward silent. Level 300 was in the "Intense Shock" range while level 315 was the first of four switches in the "Extreme Intensity Shock" range. Five subjects stopped immediately after the pounding began at level 300. Four more stopped after level 315. Two stopped at 330, following the start of the 'silence' interval; three others eventually stopped, one each at the next three levels. Twenty-six subjects continued to the end.

The interval of pretests from the victim was changed for the second study. Four separate experimental conditions tested the effect of the proximity of the victim. The first condition (Remote Feedback) fit the description of the study above. The second condition (Voice Feedback) was identical to the first except for the substitution of vocal pretests for pounding. The Proximity condition had the victim placed in the same room as the

subject, at a distance of $1\frac{1}{2}$ feet. The fourth condition (Touch-Proximity) required the subject to force the victim's hand on a shockplate when he began pretesting at the 150-volt level. The voltage-level apparently was not varied systematically in this series. No change in the voltage-level was mentioned until the fourth condition. Milgram was evidently interested more in getting subjects to rebel than in carefully planning his strategy. The results were striking:

Expressed in terms of the proportion of obedient to defiant subjects, the findings are that 34 percent of the subjects defied the experimenter in the Remote condition, 37.5 percent in Voice Feedback, 60 percent in Proximity, and 70 percent in Touch-Proximity. [N=40]

The low "n" increases the size of any error, but significant differences may be noted between the first two conditions and the last two.

The operational definitions for the experimental conditions appear appropriate. Milgram plotted the "Mean Maximum Shock" and found that increasing proximity was correlated with decreasing intensity: the mean declined from 405 volts (switch 27) for the Remote condition to 270 volts (switch 18) for the Touch-Proximity condition. Validity and reliability problems center on the absence of consistency in the interval of pretests as well as on the lack of any appropriate means of grading the proximity changes. But the results bear out a common sense judgment that the greatest change should take place between the Voice Feedback and the Proximity conditions.

The operational definitions of obedience and defiance are simple and straightforward. Milgram did not attempt to account for balkiness in his remarks. Defiance was the complete refusal

to administer any more shocks.

Milgram's operational definitions of obedience and defiance, the experimental conditions, and the scale of the shock generator were quite imaginative. He recorded the comment of one subject, who was in electronics, on the impressive appearance of the instrument. The only change that might have been helpful would be to vary the proximity of the learner-accomplice and the experimenter more systematically. But Milgram was more interested in the effect of qualitative changes than simple quantitative changes. This lack of comparability was not a problem that could be altogether avoided given his purposes.

Analysis

Milgram's papers did not refer to tests of statistical significance. This was another serious omission. Since the experiments were one-shot arrangements, Milgram did not have to account for attrition. Milgram also did not make a series of follow-up studies on the subjects. Instead, he used a post-treatment debriefing that was designed to reconcile the subject with the learner-accomplice and the experimenter. Milgram did not comment on the difficulties that may have occurred at this point. Milgram obtained self-reports on "tension and nervousness" from 137 subjects in the Proximity experiments but did not note the number of subjects who declined comment. Another graph plotted predicted and obtained behavior in voice feedback.

Milgram was concerned with possible aftereffects of the treatment and took steps designed to reassure the subject. But the ethical problems are still evident. Speaking of the

debriefing, Milgram said that

It consisted of an extended discussion with the experimenter and, of equal importance, a friendly reconciliation with the victim. It was made clear that the victim did not receive painful electric shocks. After the completion of the experimental series, subjects were sent a detailed report of the results and full purposes of the experimental program. A formal assessment of this procedure points to its overall effectiveness. Of the subjects, 83.7 percent indicated that they were glad to have taken part in the study; 15.1 percent reported neutral feelings; and 1.3 percent stated that they were sorry to have participated.... Four-fifths of the subjects felt that more experiments of this sort should be carried out, and 74 percent indicated that they had learned something of personal importance as a result of being in the study. Furthermore, a university psychiatrist, experienced in outpatient treatment, interviewed a sample of experimental subjects with the aim of uncovering possible injurious effects resulting from participation. No such effects were in evidence.

Despite these reassurances, the ethical dilemmas involved represent the most serious problem with the experiment. Many of the safeguards he discussed have all the appearance of being afterthoughts. Milgram's account of the experiments conveys a sense of continual improvisation with no clear goal in mind. The results of the experiments are fascinating, and the use of ethical dilemmas in experimental procedures has continued, notably in the studies of Lawrence Kohlberg on the moral and cognitive development of children. Philosophically, these experiments may be placed in the context of the call by Joseph Fletcher and others for a "situation ethics" as an alternative to "legalism" on the one hand and "antinomianism" on the other. The use of moral dilemmas in such a manner represents an extreme, perhaps impractical, case. The results were chilling, but the question arises: how can they be generalized? Milgram never dealt with the external validity problem at the most fundamental level. The discussion section

of the first paper is meticulous, if speculative. But Milgram did not make the transition from the unique situation to ordinary life. This is also a defect of the philosophical program, situation ethics, he appears to share. The unsettling effect of the papers is full of literary suspense. But as a piece of social science research, Milgram's paper lacks some of the expected rigor in composition. It is thus possible to evaluate it from other than a strictly scientific (or philosophical) point of view.

NOTES

- 1 Stanley Milgram, "Some Conditions of Obedience and Disobedience to Authority," in Stanley Milgram, The Individual in a Social World: Essays and Experiments (New York: Addison-Wesley Publishing Company, 1977), p. 103.
- 2 Milgram, ibid., pp. 103-104.
- 3 Milgram, ibid., p. 104.
- 4 Stanley Milgram, "Behavioral Study of Obedience," in Elliot Arensen, ed., Readings about the Social Animal (San Francisco: W. H. Freeman and Company, 1973), p. 20.
- 5 Milgram, op.cit., p. 102.
- 6 Milgram, ibid., p. 115.
- 7 Milgram, ibid., p. 110.
- 8 Milgram, ibid., p. 110.
- 9 Milgram, ibid., p. 121.