
Obedience Lite

Alan C. Elms
University of California, Davis

Jerry M. Burger's (2009) partial replication of Stanley Milgram's (1963, 1965, 1974) classic experiments on obedience to authority is considered from the viewpoint of a contributor and witness to the original obedience experiments. Although Burger's replication succeeded in terms of gaining the approval of his local institutional review board, it did so by removing a large portion of the stressful circumstances that made Milgram's findings so psychologically interesting and so broadly applicable to instances of real-world destructive obedience. However, Burger has provided an initial demonstration that his "obedience lite" procedures can be used to extend the study of certain situational and personality variables beyond those examined by Milgram.

Keywords: obedience, authority, Milgram

In September 1960, I entered Yale Graduate School, majoring in personality and social psychology. At the same time, Stanley Milgram, five years older than I and fresh out of Harvard Graduate School, became an assistant professor of psychology at Yale. World War II had ended only 15 years earlier, followed by the Nuremberg War Crimes Trials, where "I was only obeying orders" was not deemed a sufficient excuse for the wartime behavior of Nazi Germans. In 1960, the eight years of the Eisenhower presidency, characterized by widespread social conformity, were coming to a close. Although the United States was not directly involved in any active war, more than 86,000 young American males were drafted into military service during that year alone. Most of them learned to obey their superior officers without protest.

I did not meet Stanley Milgram until the spring semester of 1961, when he was required to give a talk about his research interests to first-year psychology graduate students and I was required to listen. I listened to many psychology faculty members talk about their research interests that semester, but only two really stirred my curiosity: Milgram and another new assistant professor, Lawrence Kohlberg, who was starting his research program on stages of moral development. Both Milgram and Kohlberg were looking for a research assistant. I was looking for a summer job as well as for a temporary faculty sponsor, because my major professor, Irving Janis, was going on sabbatical leave in Amsterdam for a year. Larry Kohlberg was not a dynamic speaker, and his research ideas seemed still somewhat vague, so I decided it would be more fun to work with Milgram, who came across as self-confident and well-organized. I met with Milgram soon after his talk and offered my services for the summer. As far as I know, no

one else applied for the job. He hired me with the understanding that he would also serve as my major professor until Irving Janis returned from Amsterdam. (My Yale classmate Elliot Turiel volunteered to work with Larry Kohlberg and went on to develop his own distinguished research career around the topic of moral development.)

For several weeks early in the summer, Stanley Milgram filled my eight hours a day of working time with mundane tasks. I tabulated and filed responses to the advertisement he had placed in the local newspaper, calling for volunteers to participate in "a study of memory and learning." I devised long lists of word pairs for the research participants to teach or to learn. I typed and made multiple copies of those word-pair lists and of the scripts Milgram wrote for the experimenter and the volunteers to follow. I bought electrode paste at a nearby medical supply house to apply to the learner's wrist in order to get a good electrical connection and to "avoid blisters and burns," as the research participants were told. I helped rehearse the high school teacher who would become the experimenter and the railway accountant who would become the learner in most runs of the experiments. I sorted through our files of respondents to the newspaper ad, assigning equivalent percentages of different ages and occupational categories to each experimental condition. Then I began a daily round of telephone calls to schedule the first batches of participants.

It was only when those first participants arrived at Linsly-Chittenden Hall that Milgram and I, as we watched from behind the big two-way mirrors in the Social Interaction Lab, began to realize that something truly unusual was going on—something quite different from the usual low-key social psychology experiment. Before that summer ended I watched approximately 100 participants, run one at a time, as they moved higher and higher up the sequence of switches on the shock generator. At first Milgram and I were astonished at both the intense emotional involvement displayed by most participants and their high levels of obedience to the experimenter's commands. Soon we tried to outdo each other in predicting which participants would obey fully and which ones would refuse to go further. We began to recognize certain tipping points (without using that term) beyond which a particular participant

My lasting debt to Stanley Milgram should be obvious. I thank Donald B. Ball for helpful comments on a draft of this article.

Correspondence concerning this article should be addressed to Alan C. Elms, Department of Psychology, University of California, One Shields Avenue, Davis, CA 95616. E-mail: acelms@ucdavis.edu



Alan C. Elms

was likely to go all the way. But we never became fully accurate in our predictions even then.

Those first experiments took place 46 years ago. Most social-psychological experiments from that era have long since been forgotten or invalidated or absorbed into the field's amorphous body of data and theory. But the Milgram studies continue to be a focus of interest and debate, not only among social psychologists but in other scholarly fields and in popular culture, and not only in the United States but around the world. The Milgram studies were extraordinary for their time, but in a sense they have been frozen in time, as various sets of highly restrictive rules and regulations subsequently established by the federal government, the American Psychological Association, and campus institutional review boards (IRBs) have discouraged or prevented other researchers from replicating and extending Milgram's work. Jerry Burger's (2009, this issue) attempts to replicate certain portions of the Milgram experiments are the first published efforts at close replication in the United States in several decades.

Questions about Burger's (2009) "Replication"

When I first heard of Burger's (2009) work, several questions sprang to mind: How did he get his research approved by his local IRB? To what degree did he indeed "replicate Milgram," as his article's title claims? And how far were Burger's participants willing to go up the ascending "shock" levels they thought they were administering to the apparent victim—or as Burger's subtitle asks, "Would people still obey today?"

How to get a research project past the local campus IRB may appear to nonresearchers as one of the less important questions to ask. It may even seem an inappro-

priate question. But for anyone who regards the Milgram studies as among the most valuable experiments ever done in social psychology—and I am one of the many who continue to regard them so—it is a highly salient question. Any experimental design so productive of intriguing data and so widely publicized would ordinarily have stimulated hundreds, if not thousands, of similar experiments by now: experiments attempting to expand upon the initial findings, to identify variables that might narrow or modify the import of the initial findings, and to generate new data that might challenge those early findings. Within a fairly narrow time span after Milgram published his basic body of research (Milgram, 1963, 1965), it was still possible to conduct such research in the United States or in Europe, relatively free from contention with IRBs. Indeed, in 1962 I applied for funds to do a substantial replication of the basic Milgram design in Germany, with Milgram's enthusiastic encouragement and incorporating his procedures for protecting participants' psychological well-being. When my grant application was rejected by the Social Science Research Council (probably in part because neither Milgram's research design nor his data had yet appeared in print), I continued to discuss with him the possibility of doing such a replication for my doctoral research project. He discouraged me from doing so, reasoning that when I entered the job market, reference letters from an unknown assistant professor named Milgram would carry far less weight than letters from my distinguished faculty advisor, Irving L. Janis. Neither Milgram nor I nor Janis foresaw in the early 1960s that after I earned my doctorate with a dissertation on role playing and attitude change, my opportunities for further exploration of obedience to authority as an important social-psychological variable would be permanently blocked by new ethical rules.

My subsequent research career necessarily went in other directions, as did Milgram's and that of most other social psychologists. There was a flurry of interest in pencil-and-paper simulations of the Milgram studies in the early 1970s (e.g., by Don Mixon, 1972). Those simulations yielded results that were on the whole similar to Milgram's, but they did not tell us anything new or interesting about obedience in the real world. Milgram's research was so startling, so gripping, in large part because its rather ordinary adult participants found themselves in an intensely involving emotional experience that required them to make important moral choices. Indeed, it was the achievement of such emotional intensity in a laboratory setting that stimulated the development of ethics rules so sweeping that they would have prevented Milgram himself from conducting his obedience research if they had been in place when he started.

So how did Jerry Burger (2009) get his research approved by the IRB at a respectable American university? He did so by going back to the research design drawing board and coming up with an experiment that might best be described as "obedience lite." In so doing, Burger neatly answered my first question. But in looking beyond that answer, I find that he has not done so well with my second

question: To what degree was he able to replicate the Milgram experiments?

Criticism of the ethics of the Milgram studies has focused on two broad issues: the amount of deception used by the experimenters and the emotional stress generated in participants. Burger (2009) appears to have agreed with Milgram's (1964) contention and mine (Elms, 1982) that some degree of deception is necessary and acceptable in studies of obedience, as long as it is revealed to participants soon afterward. Burger's approach to deception differed from Milgram's mainly in informing participants *immediately* after the conclusion of the "shock" sequence that "the shock generator was not real and that the confederate was not receiving electric shocks" and that the experiment was studying obedience to authority rather than memory and learning (Burger, 2009, p. 7). In the original Milgram studies, most participants were at that same point told only that the shock generator's labels of shock intensity were appropriate for small animals but not for humans and that the confederate had been receiving considerably milder shocks than his behavior indicated. Because Milgram anticipated a much more extensive series of experiments than Burger did, he maintained some degree of deception of most participants until he was ready to send them a full report on the intent of the research, the experimental procedures, and his overall research findings. Nonetheless, his debriefing procedures at the end of each experimental hour were designed (as were Burger's) to relieve participants of what most of them experienced as the crucial deception-induced belief: that they had been administering severe harm to another human being via the shock generator.

Burger's (2009) main divergences from the Milgram procedures involved efforts to reduce the emotional stress experienced by participants not only through their more complete debriefing at the end of the experimental hour but also by an overall gentling of their research participation. He accomplished this in several ways. First, participants were subjected in advance to several levels of screening, designed in large part to reject individuals most likely to experience the research procedures as stressful. Second, participants who continued on into the experimental hour were given a considerably milder sample shock than in the Milgram studies. (The administration of 15 volts rather than 45 may sound like a small difference, but even with the electrical contact increased by electrode paste it amounts to a slight tickle rather than a really unpleasant jolt.) Third, and most divergent from Milgram, Burger consistently stopped the experimental procedure when participants reached the 150-volt level, the point where the "learner" first clearly voiced an objection to being shocked.

Burger (2009) justified stopping all participants at this point (if they had not stopped already) by observing that in the Milgram studies, participants who pressed the 150-volt switch almost always continued to the 450-volt upper end of the shock sequence. Thus, stopping participants at that point in the replication would presumably result in the loss of little behavioral data while sparing participants the stress of continuing to shock the learner while he or she was screaming in pain or had apparently collapsed into silence.

This procedural change on Burger's part is likely to have been the decisive factor in getting his research proposal approved by the local IRB. Unfortunately, it also purges the most distinctive features from Milgram's basic research design and greatly diminishes the replication's generalizability to any real-world issues.

Various other researchers have tried mild or "lite" obedience tasks, such as directing participants to eat a stack of bitter-tasting crackers until the experimenter tells them to stop (e.g., Kudirka, 1965). So we already know that many participants (at least in the usual college student samples) will be quite obedient in carrying out such commands. Milgram's research was distinctive in that participants were ordered to engage in apparently *destructive* acts, acts that appeared to injure another human being, potentially causing permanent physical damage and perhaps even death. Milgram's obedient participants continued to repeat such apparently destructive acts in spite of the learner's increasingly audible distress and often in spite of their own expressed reservations about obeying the experimenter. As they did so, Milgram observed nonverbal as well as verbal expressions of participants' emotional stress, ranging from sweating and hair pulling to tears, moans, and laughing jags. Milgram was not attempting to study such stress indicators as a primary variable. Instead, he noted them as signs that participants were continuing to obey the experimenter well beyond their own usual comfort levels and in spite of whatever beliefs and values they might hold about harming an innocent individual. He was also able to observe the verbal rationalizations many participants expressed after their destructively obedient behavior ended. Such behavioral data, most of which emerged beyond the 150-volt tipping point, helped Milgram make the case that the destructive obedience shown by his experimental participants resembled real-world phenomena such as the obedience of German soldiers during the Second World War (Milgram, 1963, p. 371) and American soldiers during the My Lai Massacre in Vietnam (Milgram, 1974, pp. 183–186). If Milgram had stopped all participants at some level short of the final 450-volt switch on his shock generator, he might still have been able to make that case effectively. But if he had stopped everyone at the 150-volt level as in Burger's (2009) partial replication, the possibility of such generalizations to the real world would have been much less apparent and much more arguable.¹

Then there is my third question, as phrased in Burger's (2009) subtitle: "Would people still obey today?" Over

¹ In a meta-analysis of data across eight experimental conditions of the Milgram studies, Packer (2008) identified 150 volts (the first shock level at which the learner demanded to be released from the experiment) as the point where disobedience most often occurred. Packer's finding might be seen as supporting Burger's decision to stop all participants at the 150-volt level. But Packer also found that nearly two-thirds of disobedient participants in those conditions (63.12%) disobeyed at a shock level other than 150 volts—suggesting, in my interpretation of Packer's results, that Burger lost potentially interesting data on many disobedient as well as obedient participants by stopping everyone at 150 volts.

the past quarter century, ever since Stanley Milgram's tragically early death, I have often been asked to give talks or interviews in his place about the obedience studies. On almost every such occasion, the interviewer or a member of the audience has raised just that question. In response, I have usually expressed optimism that a current measure of obedience to destructive authority would find substantially *less* obedience than Milgram did. While noting a lack of fresh data to support that optimism, I have hopefully suggested that by now a large portion of our populace should have learned important lessons about unthinking obedience from the teachings of Martin Luther King Jr. and other social activists, from deceptive government efforts to obtain widespread obedience to destructive edicts—and also from the publicizing of Milgram's results through the mass media and in college and high school textbooks. From a first look at Burger's figure of 70% full obedience (vs. Milgram's 82.5% at the same point in his most comparable condition), my optimism appears to have received some support. But the difference between those two percentages does not come close to statistical significance. So much for my hopeful expectations about substantially lower obedience!

However, several factors complicate or weaken any direct comparison between Burger's (2009) percentages of obedient participants and Milgram's. First, Burger screened out anyone who reported having heard about the Milgram studies. I understand his reasons for doing so, but those potential participants who remembered having heard something about the Milgram studies should also, in my admittedly optimistic view of the power of education, have been among the most *disobedient* participants if they were permitted to participate. The practice of excluding them is likely all by itself to have raised Burger's average level of obedience well above what it would have been had no screening for sensitivity to questions of obedience taken place. Second, several other potential participants were rejected "to exclude any individual who might have a negative reaction to the experience" (Burger, 2009, p. 2). Burger did not indicate how such judgments were made by the screeners, but surely some volunteers who gave signs that the experience might upset them would have added to the percentage of disobedient participants if they had been included in the experiment. And third, Burger's attempt to make experimental participation less distressing by giving participants only "a very mild 15-volt shock" (p. 2) rather than the more intense sample shock in the Milgram experiments may have led them to assume that the shock generator was not really that shocking. For all these reasons, the participants who survived Burger's double screening and who then faced the redefined shock board may have been considerably more obedient on average than an un-screened population presented with an exact replication of the Milgram procedures would be today.

Making Further Use of Burger's (2009) Design

Despite my substantial reservations about Burger's (2009) initial findings, I am pleased that he has been able to revive

the possibility of conducting research on obedience to authority somewhat along the lines that Milgram initiated. Though direct comparisons of absolute levels of obedience cannot be made between the 150-volt maximum of Burger's research design and Milgram's 450-volt maximum, Burger's "obedience lite" procedures can be used to explore further some of the situational variables studied by Milgram as well as to look at additional variables. Burger already began such exploration by examining the effects of adding a second "teacher" who modeled disobedient behavior for the real participant. Milgram looked at a considerable variety of other group-influence patterns relevant to obedience as well as introducing substantial variations in the characteristics of the authority figure and the learner/victim. Burger and other social psychologists should be able to come up with many additional situational variables that have remained untouched during the Dark Age of obedience research proscription.

Burger's (2009) procedure can also be used to study a category of phenomena that remained largely unexplored by Milgram: personality variables. As social psychologists, both Milgram and Burger have strongly emphasized situational variables that influence obedience. Milgram did not make much effort to collect data on the personalities of his obedience participants, but he did recognize that situational factors interact with individual differences in producing obedience: "Situations producing the greatest obedience could do so by triggering the most powerful, yet perhaps the most idiosyncratic, of motives in each subject confronted by the setting. Or they may simply recruit a greater number and variety of motives in their service" (Milgram, 1965, p.72). In recognition of such individual variations in response to situational factors, Milgram collected demographic data on all his participants. Further, he was willing to facilitate my exploration of personality and life-history variables that might distinguish fully obedient participants from those who defied at least one of the experimenter's commands (Elms, 1972, pp. 128–136; Elms & Milgram, 1966).

To study such variables, I invited 40 of Milgram's participants from the original "proximity series" to return to the Yale campus several months after their experimental participation. I interviewed and gave various personality tests to each participant in an individual session lasting about two hours. To maximize potential contrasts in personality patterns, I recruited 20 participants who had been fully obedient even with maximal cues for disobedience and 20 who had disobeyed even with maximal cues promoting obedience. To my disappointment (though not, I think, to Milgram's), these two behaviorally very different sets of participants did not display any especially dramatic personality differences on the measures I used. But they did show *some* differences that appeared to be related to their levels of obedience.

Obedient and defiant participants in my sample did not differ significantly on any of the standard scales of the Minnesota Multiphasic Personality Inventory (MMPI) or on a nonstandard MMPI scale intended to measure social dominance (perhaps similar to Burger's, 2009, measure of

“desire for control”). They did differ significantly on a nonstandard MMPI scale developed as a measure of social responsibility (Gough, McClosky, & Meehl, 1952), with the defiant participants scoring higher on that scale. However, the empirical definition of the scale does not clearly indicate why a high score might predispose someone to defy an authority figure’s commands. A clearer pattern was found for measures of authoritarian tendencies. Fully obedient participants appeared rather consistently more authoritarian than disobedient participants on the original California F scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) as well as on several biographical indices and on patterns of attitudes toward the experimenter, the learner, and Yale University. I did not directly measure any personality variable resembling Burger’s measure of “empathic concern,” and he did not measure any variable (other than behavioral obedience itself) that resembled my measures of authoritarianism. Future research employing his partial-Milgram-replication procedures can explore the role of all these as well as other personality variables.

One final note: In his general comments on Milgram’s obedience research, Burger (2009) observed, “In truth, Milgram’s work is more properly described as a series of demonstrations rather than as an experiment” (p. 1). Similarly, in a book review dealing with the Milgram studies Burger (2002) stated,

In fact, the obedience “studies” were not really studies (or experiments) in the sense that participants were randomly assigned to experimental and control conditions. Rather, Milgram provided a series of demonstrations of an amazing phenomenon, altering the procedures slightly each time and comparing the findings to those from earlier demonstrations. (p. 665)

Here Burger appears to have misunderstood either the nature of Milgram’s research or the defining characteristics of experimental social psychology. Milgram’s (1963) first publication on obedience may have appeared to be no more than a demonstration, since it described only a single experimental condition in detail. Milgram wanted to get that paper into print as quickly as possible in order to establish his priority in developing the shock generator as a means of measuring behavior. However, Milgram (1965, 1974) subsequently described several clusters of experimental conditions, such as the proximity series, in which

participants were indeed randomly assigned to various conditions (I assigned them!) and data from different conditions were compared with a baseline control condition. Certain other now-famous psychological studies, such as Zimbardo’s (2007) so-called Stanford Prison Experiment, are indeed more accurately described as demonstrations than as experiments. (Zimbardo referred to his own research as an “extended case study” on p. xii.) But Milgram’s studies of obedience were, by any proper definition of the term, true experiments—as was, on a much more limited scale, the replication conducted by Jerry Burger (2009).

REFERENCES

- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper.
- Burger, J. M. (2002). Four decades and counting [Review of the book *Obedience to authority: Current perspectives on the Milgram paradigm*]. *Contemporary Psychology: APA Review of Books*, *47*, 665–667.
- Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist*, *64*, 1–11.
- Elms, A. C. (1972). *Social psychology and social relevance*. Boston: Little, Brown.
- Elms, A. C. (1982). Keeping deception honest. In T. L. Beauchamp, R. R. Faden, R. J. Wallace Jr., & L. Walters (Eds.), *Ethical issues in social science research* (pp. 232–245). Baltimore, MD: Johns Hopkins University Press.
- Elms, A. C., & Milgram, S. (1966). Personality characteristics associated with obedience and defiance toward authoritative command. *Journal of Experimental Research in Personality*, *1*, 282–289.
- Gough, H. G., McClosky, H., & Meehl, P. E. (1952). A personality scale for social responsibility. *Journal of Abnormal and Social Psychology*, *47*, 73–80.
- Kudirka, N. K. (1965). *Defiance of authority under peer influence*. Unpublished doctoral dissertation, Yale University, New Haven, CT.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371–378.
- Milgram, S. (1964). Issues in the study of obedience: A reply to Baumrind. *American Psychologist*, *19*, 848–852.
- Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*, *18*, 57–76.
- Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.
- Mixon, D. (1972). Instead of deception. *Journal for the Theory of Social Behavior*, *2*, 145–177.
- Packer, D. J. (2008). Identifying systematic disobedience in Milgram’s obedience experiments: A meta-analytic review. *Perspectives on Psychological Science*, *3*, 301–304.
- Zimbardo, P. (2007). *The Lucifer effect: Understanding how good people turn evil*. New York: Random House.