

## Stanley Milgram's Obedience Experiments: A Report Card 50 Years Later

Augustine Brannigan

# Springer Science+Business Media New York 2013

Fifty years ago Stanley Milgram published the first report of his studies of obedience to authority. His work (1963) forged the mindset of how social scientists over the next two generations came to explain the participation of hundreds of thousands of Germans in the mass murder of European Jews during the Holocaust. Milgram's model was Adolph Eichmann who was convicted and executed for his role in the deportation of European Jews to death camps created in Poland for their eradication. Eichmann's legal defense, that he was 'just following orders,' suggested that the final solution to the 'Jewish problem' in Europe was engineered by desk murderers remotely positioned in hierarchies of authority across the Nazi bureaucracy. Submission was unquestioned because the decision to eradicate the Jews originated from the sovereign authority. Milgram's murderers were loyal automatons.

Milgram attracted his subjects from the wider community in New Haven and Bridgeport. He recruited an astounding 780 subjects. His work was identified by Roger Brown as 'the most important psychological research' done in his generation. Where Hannah Arendt speculated philosophically that the ranks of Holocaust perpetrators such as Eichmann were unremarkable non-entities, Milgram described in an experimental idiom the ease with which New Haven citizens could be transformed into brutal Nazis without much difficulty. Milgram's work also provoked questions about the ethical treatment of human subjects in a way that helped to shape future policies for the treatment of volunteers in experimental studies. It alerted funding agencies to the necessity of risk assessment of those deliberately misled in studies premised on subject deception. Milgram also championed the proposition that grave questions of human morality could be examined

following the experimental methods that proved to be so effective in the natural sciences. He also contributed to the dogma in social psychology that 'the situation' is one of the most, if not *the* most, important determinant of social behavior. His 1974 book was promoted widely in the popular press, and created a media storm. His scientific portrayal of 'the banality of evil' inspired an artistic outpouring of films, and plays, and remains a point of relevance in studies of holocaust history today. Stanley Milgram died in 1984 at the age of 51.



Stanley Milgram 1933-1984

Enter Gina Perry. Perry is an Australian journalist and writer who took an interest in the Milgram study after learning through personal acquaintances that several persons who

---

A. Brannigan(\*)  
Department of Sociology, University of Calgary, 2500 University  
Dr NW, Calgary, Alberta, Canada T2N 1N4  
e-mail: branniga@ucalgary.ca

participated in a replication of the obedience study at La Trobe University in Melbourne in 1973 and 1974 continued to suffer trauma decades later. That unpublished replication involved some 200 subjects. She turned her attention to the original study, and spent 4 years researching the Milgram archives at Yale University. She listened to 140 audio recordings of the original experiments, and dozens of hours of conversations involving former subjects in the postmortem debriefing with a psychiatrist. She interviewed former subjects and experts familiar with the research, family members of the actors, and read the mountains of documentation and correspondence accumulated during the study. The conclusions she draws from her investigation were disturbing, and will fundamentally challenge the way scholars interpret Milgram and his experiment.

What is known about Milgram and his background? He was interested in the conditions that led to the expression of deep antisemitism in Germany during the Second World War. Some thought that pathological conformity might have had national roots. In his doctoral work, he investigated national differences in defiance of group pressure to conform to judgments that subjects thought were incorrect. In this work, he adopted the protocols of Solomon Asch. Indeed, he found national differences in conformity between Norwegian and French subjects, but nothing that illuminated the German case. As a young professor he sought to raise the ante by creating conditions in which subjects were compelled, not just to *say* something they thought to be untrue, but to act in a gravely inappropriate way. Most students with any postsecondary training in recent years will recall the experiment. The ‘cover story’ was a learning experiment in which potential teachers and learners were recruited from the public by newspaper ads and direct mail solicitation. Persons who presented individually for the study drew names out of a hat for their respective roles. A lab-coated scientist explained the need to determine the effectiveness of punishment on the learning process. The subjects were deceived from the start. The learner (Jim McDonough) and the scientist (John Williams) were amateur actors. Participants were paid \$5 for participation and carfare, a very significant compensation at the time in 1961. They witnessed the physical restraint of the learner in an attached room where he was to be ‘tested.’ The shocking appliance consisted of 30 switches numbered from 15 to 450 V. It buzzed and snapped and exuded technological credibility as the teacher administered the punishments up to a level labeled as ‘severe shock.’ To start things, the teacher read a long list of word pairs, and then, under the supervision of Mr. Williams, the scientist, proceeded to test whether the learner retained knowledge of the information just presented to him. Now the drama began in earnest. The learner apparently had a terrible memory. The shocks escalated in accord with the errors. And he began to protest with increasingly painful moans and screams. This was all piped back to the learner through speakers. The response pattern was all predetermined, and

designed to create an increasingly disagreeable dilemma for the subject: either shock or walk, obey or defy. Each experiment lasted about 50 min, and resulted in levels of agitation among some subjects that were unprecedented in previous psychological research. Milgram’s question was simply how far the teacher would continue to issue painful shocks before defying Mr. Williams’s directives to continue. Milgram’s critical finding was that 65 % of ordinary persons would administer levels of punishment that would appear to be lethal, even where, in one condition, the learner was depicted as a person with an existing heart condition. Despite this apparently profound level of aggression, the conventional wisdom suggested that the agitation associated with the exercise dissipated immediately as the subjects were ‘de-hoaxed.’ All the civility that had been suspended was restored by the debriefing, and having been made whole again, the subject and the scientist departed company on good terms. That was the myth.



Traumatized subjects

Perry discovered a different picture. Herb Winer was ‘boiling with anger’ for days after the experiment (p. 79). At the time, like Milgram, he was an untenured professor at Yale. He confronted Milgram in his office with his concerns about the experiment, particularly about pressure to shock someone with a heart condition. His trauma was so intense that he confided in Perry, nearly 50 years later, that his memory of the event would be ‘among the last things I will ever forget’ (p. 84). After the cover story was explained, Winer became an admirer of Milgram, ‘although he will never forgive him for what he put him through.’ Bill Lee was another subject tracked down by Perry. Bill Menold was unsure of whether the study was a sham or not, but he found it ‘unbelievably stressful...I was a basket case on the way home’ (p. 52). He confided that night in a neighbor who was an electrician to learn more about electrical shocks. Hannah Bergman (a pseudonym) still recalled the experiment vividly after half a century. Her recollections suggested that she ‘was ashamed—and frightened.’ Her son told Perry that ‘it was a traumatic event in her life which opened some unsettling personal issues with no subsequent follow-up’ (p. 112). A New Haven Alderman

complained to Yale authorities about the study: ‘I can’t remember ever being quite so upset’ (p. 132). One subject (#716) checked mortality notices in the *New Haven Register*, for fear of having killed the learner. Another subject (#501) was shaking so much he was not sure he would be able to drive home; according to his wife, on the way home he was shivering in the car and talked incessantly about his intense discomfort until midnight (p. 95). Subject 711 reported that ‘the experiment left such an effect on me that I spent the night in a cold sweat and nightmares because of fears that I might have killed that man in the chair’ (p. 93). None of the previous histories of these experiments even hinted at such reactions, nor was any of this ever reported in the university curriculum. What caused all the trauma?

To say that the de-hoaxing left a lot to be desired would be a gross understatement. In his first publication, Milgram had written that steps were taken ‘to assure that the subject would leave the laboratory in a state of well-being. A friendly reconciliation was arranged between the subject and the victim, and an effort was made to reduce any tensions that arose as a result of the experiment’ (Milgram 1963: 374). Also, ‘at the very least all subjects were told that the victim had not received dangerous electric shocks.’ Perry’s review of the archives indicates that this was simply not the case. In fact, Perry reports that 75 % of the subjects were not immediately debriefed in any serious way until the last 4 out of 23 conditions. Perry reports that subjects in conditions one to eighteen, around 600 people, left the lab believing that they had shocked a man, with all that dramatized agony etched on their conscience (p. 92). This was corroborated by Alan Elms, Milgram’s research assistant in the first 4 conditions. ‘For most people who took part, the immediate debrief did not tell them there were no shocks’ (p. 90). In addition, many of the subjects who met after the completion of the study with the psychiatrist, Dr. Paul Errera, similarly reported they received no debriefing at all (p. 89–107).

At minimum a debriefing would have involved an explanation that the scientist and the learner were actors, the shocking appliance was a fake, all the screams were simulated, and that the teachers were the focus of the study. Perry reports that even where some account was given by Milgram to the subjects, they were told that their behaviours, whether obedient or defiant, were natural and understandable, and that the shocking device had been developed to test small animals, and was harmless to people. So even when it occurred, the debriefing, in Perry’s words, ‘turned out to be another fiction’ (p. 90). In addition, the debriefing was remarkably brief—two minutes—and did not involve any question-answer interaction with the experimenter. Milgram did not want future subjects to be contaminated by accounts from prior subjects about the true nature of the experiment, and so he withheld such information until the experiment was virtually over. A fuller explanation was mailed to subjects a year later, but it does not seem to have consoled any of those interviewed by Perry.

## The Sceptical Subjects

If many subjects were traumatized, there were significant others who had their doubts about the cover story (p. 156). One subject wrote to Professor Milgram the day after his participation. He had inferred that the ‘draw’ for roles was fixed, and that both pieces of paper probably had the word ‘teacher’ written on them. He found the learner unaccountably ‘disinterested’, and was suspicious of all the one-way glass mirrors. He also noticed that the learner was not given his check at the same time as himself. Another noticed that the learner’s check was dog-eared from what appeared to be frequent use. Others engaged in reality testing by asking the learner to tap on the wall if he could hear him. No response. One *lowered* the shock level intentionally, and the learner seemed to express *increased* pain despite this. Others were simply sceptical that Yale would permit anyone to absorb such punishment. Some commented on the fact that no one with a cardiac condition which was under medical surveillance would submit to such intense agitation. Another noted that there was a speaker in the learner’s room, and the sound from the voice did not appear to be coming through the door, as he would have expected. And many suggested that the sounds appeared to be audio recordings. All this was noted in the archives. Under these conditions, the subjects simply played along as required by the experiment, since they assumed that no one would purposely be hurt, and it was all for the good of science.

Milgram was aware of this scepticism, but he dismissed it as a reaction formation. He reasoned that the subjects had acted shamefully, then, in self-defense, they denied anyone was injured, and that they had not done any harm. Perry also turned up a report by a research assistant, Taketo Murata. He examined the maximum shock levels in all 23 conditions, but divided the subjects into those who appeared to be doubters, and those who appeared to be true believers. ‘Taketo found that people most likely to disobey and give lower-voltage shocks were those who said they believed someone really was being hurt’ (p. 164). Perry comments further that ‘only half of the people who undertook the experiment fully believed it was real, and of those, two-thirds disobeyed the experimenter’ (p. 163). There was another area of information leakage that must have piqued the curiosity of some teachers. There were numerous cases where the subjects practically shouted out the correct answer to the learner, but this communication never made a difference in his response. Also, numerous teachers, frustrated by the learner’s poor performance, offered to switch places during the experiment, but again, this offer did not attract any interest or response. This did not always result in outright disbelief, but created some suspicions that things were not exactly as they seemed.

## The Secret Experiment

In his 1974 book, *Obedience to Authority*, Milgram gave the fullest account of the various conditions he investigated. He reported 18 conditions. In the archives, Perry came across 24, although one, the 'Educated Opinion' condition, was not actually an experiment, but an estimation by psychiatrists and university students of the probability that average subjects would be fully obedient given the conditions described to them. Among the unpublished investigations, Perry discovered a remarkable condition that Milgram had kept secret. This was the study of 'intimate relationships.' Twenty pairs of people were recruited on the basis of a pre-existing intimacy. They were family members, fathers and sons, brothers-in-law, and good friends. One was randomly assigned to the teacher role, the other to the learner role. After the learners were strapped into the restraining device, Milgram privately explained the ruse to them, and encouraged them to vocalize along the lines employed by the actor in response to the shocks in previous conditions. The 'intimate relationships' study produced one of the highest levels of defiance of any condition: 85 %. It also produced a great deal of agitation to teachers as the learners begged their friends or family members *by name* to be released. One subject (#2435) went ballistic with the scientist's pressure, and started shouting at him for encouraging him to injure his own son.

Perry speculated that Milgram was ambivalent about this condition for two reasons. On the one hand, 'Milgram might have kept it secret because he realized that what he asked subjects to do in Condition 24 might be difficult to defend' (p. 202). After all, he abused their mutual trust and intimacy to turn the one against the other. On the other hand, the results countered the whole direction of Milgram's argument about the power of bureaucracy. Perry found a note in the archives in which Milgram confessed that 'within the context of this experiment, this is as powerful a demonstration of disobedience that can be found.' When people believed that someone was being hurt, and that it was someone close to them, 'they refused to continue' (p. 202). Given its implication, the finding was never reported.

This suggests that, to an extent, Milgram cherry-picked his results for impact. Perry notes that Milgram worked to produce the astonishingly high compliance rate of 65 %. He assumed that he needed a plurality of his subjects, but not a figure so high that it begged credibility. In pilot studies he tweaked the design repeatedly. At first, there was no verbal feedback from the learner, and every subject, when commanded, went indifferently to the maximum shock. Such a response would suggest that subjects did not actually assume they were doing anything harmful. The verbal feedback from the learner was introduced to create resistance. Milgram also explored a number of Stress Reducing Mechanisms and Binding Factors to optimize compliance. Stress was reduced, for example, by framing the actions as part of a legitimate learning experiment, and by advising the

subjects that there was no permanent damage from the shocks. The binding factors included the gradual 30 step increments from the lowest to the highest shock level on the supposition that once they started, the movement up the shock scale would signal their acceptance of the protocol one step at a time.

Perry also found that there was often a Mexican standoff between the subjects and Mr. Williams as to their point of defiance. This was particularly evident in the all-female design. In their histories of the experiments Blass (2004) and Miller (1986) created the impression that the scientist would use 4 specific prods to encourage the subjects to continue, since that was what Milgram published. 'If the Subject still refused after this last [fourth] prod, the experiment was discontinued' (Blass 2004: 85). The subjects were always free to break off. After listening to the Female Condition (condition 20), Perry concluded: 'this isn't what the tapes showed' (p. 136). Mr. Williams did not adhere strictly to the protocol. This was reflected in post-mortem interviews with Dr. Errera, where three women from the Female Condition suggested that they had been 'railroaded' by Williams (p. 135). He would not relent in his pressure. In one case (#2026), he brought the subject a cup of coffee while she sat idle for 30 min before succumbing to repeated pressure to continue. Another subject was prompted a total of 26 times. This suggests, not only that the results could be cherry-picked between conditions, but also that in any one condition the scientist could elevate the compliance rate by departure from the protocol and the relentless application of pressure. The resulting 65 % compliance in condition 20 was equivalent to the previous highs achieved in two earlier conditions. In the remote feedback design (condition 1), the victim apparently pounded on the wall to signal distress. This reduced compliance from 100 to 65 %. In the cardiac condition, all the elaborate moaning, screaming and demands to be released (condition 5), resulted surprisingly in the same 65 % compliance level. How could such radically dissimilar feedback result in identical levels of compliance? This might be explained in part by the degree to which the scientist adhered strictly to, or departed from, the 4 prod protocol. As Perry's analysis of the Female Condition suggests, the various treatments were simply not standardized. Milgram's conclusion that there were no gender differences in aggression based on a comparison of outcomes in condition 5 and condition 20 does not bear scrutiny.

In his face-to-face dealing with subjects, Milgram assured them that their reactions were normal and understandable. Yet in his book he describes the compliant subjects as acting in 'a shockingly immoral way' (1974: 194). In his notes, he describes them as 'moral imbeciles' capable of staffing 'death camps' (Perry 2012: 260). In the 1974 coverage of his book on the CBS network '60 minutes' program, he portrays the compliant subjects as New Haven Nazis (p. 369), and asserts that one would be able to staff a system of death camps in America with enough people recruited from medium-sized American towns.

## The Obedience Legacy

What are the implications of Perry's critique for the place Milgram is accorded in the canons of historical psychology? After all, we are told that Milgram's results were essentially replicated in 2009. I offer five observations. First, there are many shortcomings in Perry's work. Her evidence is anecdotal in the sense that she was not able to canvas systematically large numbers of former subjects, and after 50 years, this is not surprising. She did have access to all the postmortem interviews of former subjects with psychiatrist Paul Errera, but only 120 of the original 780 subjects were asked to attend these meetings, and only 32 did. Of the 780 individual cases, only 140 audio-recordings were available for Perry's use. Nonetheless, there was convergence among the surviving subjects about the enduring levels of trauma from which they continued to suffer, both in the US and Australia. She also discovered that the majority of subjects were not appropriately debriefed in a timely manner. From the archival materials, there was a mountain of evidence of scepticism among subjects who were not deceived by the cover story. This was a fact Milgram stubbornly refused to acknowledge. Additionally, she discovered a totally new, and previously unreported condition, the 'intimate relationships' study, which dramatically altered the significance of the entire experimental initiative. None of these findings were reported in the previous histories of the experiment by Blass or Miller. The value of her book is to question the scientific and moral significance of the obedience study as it was originally reported.

Second, her analysis of Milgram's formal advocacy of the primacy of 'the situation' with the harsh moral judgments Milgram offered in print and in privacy regarding the obedient subjects illuminates the enormous disconnect between Milgram's scientific posturing and his conclusions about the defects of those who obeyed. Perry: 'He associated obedient behaviour with lower intelligence, less education, and the working classes' (p. 298). Defiant subjects were smart, educated and middle class. The obedient were impervious to the suffering they caused, were remorseless and 'unthinkingly obedient' (p. 243). Yet the behavior exhibited in the lab was frequently marked by deep anguish and empathy, even by persons compelled to obey, not by 'the situation,' but by the relentless badgering by Williams, their own self-doubts, and a sense that all was done for defensible, scientific reasons. Just as Kant's scientist must presuppose space and time before the laws of physics and chemistry can be deduced, the psychologist must presuppose that all human experience occurs in situations. But situations do not explain differences in behaviour.

Third, the recent historiography of the Holocaust in the work of Browning and Goldhagen emphasized the agency of the ordinary perpetrators. These were persons who acted, not out of fear of reprisal from superiors, not out of duress or fear of disobedience, nor were they blindly obedient. They acted out of a sense of duty and responsibility. Cesarani's and Lipstadt's accounts of Eichmann depict a man who was the Final Solution's greatest advocate. These accounts stand in dramatic contrast to the banality of evil that Milgram's work perpetuated, the image of automatons who had no moral agency once they put on a uniform. The experimental account he attempted to construct actually misrepresented the original phenomenon, just as the accounts of his subjects suppressed their agency, and their struggles with him in his own lab.

Fourth, anytime someone raises questions about the validity or reliability of the Milgram experiment, he or she is told that it has already been replicated all over the world, so that criticizing Milgram is essentially a waste of time. Milgram reported that his work had been replicated in Australia, Germany, Italy, and South Africa, suggesting the findings were universal. However, according to Perry, 'the Australian study found significantly lower levels of obedience than Milgram's', as did the Italian and Germany studies; the South African study was a student report based on 16 subjects (p. 307). In 2009 Jerry Burger reported in *American Psychologist* that he had partially replicated Milgram's condition 5 (the cardiac condition). His original study was financed by the ABC network as a reality TV show, and was designed to produce comparable results. It was not designed to examine any of the methodological questions raised by Perry. In a second co-authored publication, Burger suggested that the central phenomenon he studied was something other than obedience, since the prods that appeared to look most like direct commands were ones that were singularly ineffective in producing compliance. In addition, his subjects were told that the learner, like themselves, could disengage for any reason at any point in the process. What could one infer from all that hollering that failed to result in the learner ever quitting?

Last point. Social psychology is an awkward discipline since it occupies a cognitive space already filled with common sense, the judgment of ages, and insight derived from lived experience. It has undertaken to replace this sort of human knowledge with something based on rationality derived from adherence to the experimental method. In Rodrigues and Levine's *100 Years of Experimental Social Psychology* (1999), the contributors came to two conclusions. First, they repeatedly noted that the field had not resulted in a significant body of cumulative, non-trivial knowledge. Second, whatever it was that formed the core of the discipline did not share substantial consensus among its contributors (Brannigan 2002). Perry's disturbing investigation of the Milgram archives will not change any of that.

## References

- Blass, T. 2004. *The Man who Shocked the World: the life and legacy of Stanley Milgram*. New York: Basic Books.
- Burger, J., Girgis, Z. M., & Manning, C. C. 2011. In their own words: Explaining obedience to authority through an examination of participants' comments. *Social Psychological and Personality Science*, 2, 460–466.
- Brannigan, A. 2002. The Experimental Turn in Social Psychology. *Society: Social Science and Modern Society*, 39(5), 74–79.
- Perry, Gina. 2012. *Behind the Shock Machine: The untold story of the notorious Milgram psychology experiments*, Melbourne and London: Scribe Books. New York: The New Press, 2013.
- Milgram, S. 1963. Behavioral Study of Obedience. *Journal of Abnormal Psychology*, 67, 371–78.
- Milgram, S. 1974. *Obedience to Authority*. New York: Harper & Row.
- Miller, A. G. 1986. *The Obedience Experiments: A case study of controversy in social science*. New York: Praeger.
- Augustine Brannigan is Professor Emeritus of Sociology at the University of Calgary where he taught from 1979 to 2012. His most recent book is *Beyond the Banality of Evil: Criminology and Genocide*(Oxford University Press 2013). This book is based in part on his field work in Rwanda and Tanzania. He also authored *The Rise and Fall of Social Psychology* (Aldine Transaction 2004), and *The Social Basis of Scientific Discoveries* (Cambridge University Press 1981) which was re-released in 2010. He has published widely in criminology, criminal justice and social psychology.