Credibility and Incredulity in Milgram's Obedience Experiments: A Reanalysis of an Unpublished Test

Gina Perry[1], Augustine Brannigan[2], Richard A. Wanner[2], and Henderikus Stam[2]

Abstract

This article analyzes variations in subject perceptions of pain in Milgram's obedience experiments and their behavioral consequences. Based on an unpublished study by Milgram's assistant, Taketo Murata, we report the relationship between the subjects' belief that the learner was actually receiving painful electric shocks and their choice of shock level. This archival material indicates that in 18 of 23 variations of the experiment, the mean levels of shock for those who fully believed that they were inflicting pain were lower than for subjects who did not fully believe they were inflicting pain. These data suggest that the perception of pain inflated subject defiance and that subject skepticism inflated their obedience. This analysis revises our perception of the classical interpretation of the experiment and its putative relevance to the explanation of state atrocities, such as the Holocaust. It also raises the issue of dramaturgical credibility in experiments based on deception. The findings are discussed in the context of methodological questions about the reliability of Milgram's questionnaire data and their broader theoretical relevance.

Stanley Milgram's obedience-to-authority experiments are undoubtedly the most famous research in social psychology. Naive volunteers allocated the role of teachers in an experiment purportedly about the effect of punishment on learning were instructed to give a "learner" increasing levels of electric shock each time he failed to recall the correct answer in a memory test. Milgram concluded that the majority of subjects' compliance with the experimenters' commands was evidence of humankind's innate capacity for obedience to authority. Given fresh impetus with the recent availability of Milgram's research materials, the

[1]University of Melbourne, Melbourne, Victoria, Australia
[2]University of Calgary, Calgary, AB, Canada

Corresponding Author:
Dr. Gina Perry, School of Culture and Communication, Faculty of Arts, University of Melbourne, Parkville, Melbourne, Victoria, 3010, Australia
Email: gperry@labyrinth.net.au

research continues to attract scholarly attention and interest with debates about the methodological, ethical, and theoretical claims of the obedience experiments.

Stanley Milgram's papers became available to researchers in 1993 through the Yale University Archives. The extensive guide to the Stanley Milgram Papers describes the contents of the 266 linear feet of files in 424 boxes, which contain Stanley Milgram's professional correspondence and research output from 1950 until his death in 1984 (Yale 2017). A substantial portion of the archives is taken up by material relating to the obedience-to-authority experiments. This portion of the archive contains letters, grant applications, notes, data files, and audio recordings of the experiments themselves. The archives have allowed a growing number of scholars to scrutinize unpublished material and recordings related to the obedience experiments. Scholarly examination of the material has prompted revelations about the extent and nature of debriefing (Nicholson 2011), unstandardized experimental protocols (Gibson 2013a; Russell 2011), unreported data and misrepresentation of results (Modigliani, 1995; Perry 2013b), and incongruities in Milgram's conceptual accounts of the research (Kaposi 2017).

Milgram's design specifically included features that suggested that with the escalation in shocks, the learner began to suffer, that is, emit painful groans, complaints, and a scream. The scientist assured the teacher, however, that "though the shocks may be painful, they are not dangerous" (Milgram 1974:23). These inconsistencies produced a variety of responses. In a comparison of published and unpublished materials, including correspondence between some subjects and Milgram, recordings of experimental sessions, results of subject

surveys, transcripts of psychiatric follow-up interviews, and interviews with former subjects, Perry (2013a) concluded that subjects were not uniform in their belief in the experimental setup, including many who were skeptical of the shocks. This resonated with earlier critics of Milgram who argued that the stress experienced by subjects could be attributed to the ambiguity of the experimental scenario and the confusion of subjects in knowing what to believe (Mixon 1977; Orne and Holland 1968). The matter may have been more indeterminate, however, whereby teachers accepted that the shocks were painful without being harmful, or that the learner reactions were exaggerated (Hollander and Turowetz 2017).

This article builds on the recent analyses by Hollander and Turowetz (2017) of participant explanations of their behavior through an analysis of recorded conversations between subjects and experimental staff after the experiment and on the work of Haslam et al. (2015) in analyzing participants' written responses to a post-experimental survey. These scholars have evidenced a whole spectrum of subject reactions, from skepticism to belief (and everything in between), that raises the issue of the "dramaturgical credibility" of Milgram's design. In this article, we offer a more comprehensive approach by using a combination of questionnaire and interview material that offers broader insights into social psychological experiments based on deception.

A feature of much of the scholarly discourse in the decades since Milgram conducted his experiments has been the profound and provocative nature of his finding of humanity's slavish obedience to authority (Blass 2000; Lunt 2009). While not denying the ethical issues that Milgram's experiment raised regarding the welfare and treatment of subjects,

many current scholarly narratives tend to accept Milgram's assumption that we are all capable of torture and murder at the behest of an authority figure (Russell 2018), a conclusion that is transferred uncritically to the general public and prospective students through the textbook industry (Griggs and Whitehead 2015).

In this article, we explore the degree to which subjects believed the experimental scenario and the influence of such beliefs on the experimental results. Central to the claims Milgram made about uncovering an innate tendency to follow orders is the assumption that his subjects, under the guise of a memory test, and presumably hoodwinked by the experimental deception, continued to follow the commands of an experimenter to administer what were presented as increasingly painful electric shocks to a person who seemed to be suffering.

HOLLANDER AND TUROWETZ'S ANALYSIS OF THE POSTEXPERIMENTAL INTERVIEWS

Milgram's very first probes regarding the effectiveness of his deceptive cover story occurred in the conversations conducted immediately after the termination of the experiments. These were designed to switch the definition of the situation away from the bogus learning task to the subjects' reactions to the administration of shocks. Archival materials indicate that Milgram systematically audiotaped all cases that were tested through some 23 different iterations of the obedience experiment, although tapes for two key experiments (5 and 6) are unaccountably missing. The audiotapes continued after the experiment ended and captured the reflections of the subjects in an informal interview with either the "scientist," John Williams; Milgram himself; or Alan Elms, Milgram's young research

assistant. These interviews were semi-structured and were designed to probe the subjects' reactions to their experience and to try to expose the reasons for their behavior. Hollander and Turowetz (2017) analyzed 117 recordings that were selected from condition 2 (voice feedback), condition 3 (proximity), condition 20 (women as subjects), condition 23 (Bridgeport), and condition 24 (intimate relationships). From the 117 cases, they settled on 91 recordings where the subjects gave at least one discernible account or explanation of their behavior. Hollander and Turowetz's analysis focused on the commonsense reasoning given by the subjects to account for their reactions in the immediate aftermath of the experiments. The 91 recordings included 46 cases where the subjects had been classified as "obedient" (i.e., completed the entire escalation of shocks) and 45 cases defined as "defiant" (i.e., discontinued the regimen at any point).

Among those classified as obedient, there were four major accounts identified. In some cases, more than one account was offered for why the subjects continued with the experiment. The most prevalent explanation was that the subjects did not believe the learner was actually being harmed. This finding was astonishing in light of Milgram's earlier assurances of the credibility of the cover story. This was offered in 33 out of the 76 accounts uncovered. "The most common 'obedient' explanation is *L* [learner] *not being harmed*, as 33 (out of 46 total) 'obedient'-outcome participants used it at least once (72%)" (Hollander and Turowetz 2017:660). This was followed by accounts of "following instructions" of the scientist (27 cases), the "importance of the experiment" (11 cases), and the existence of a tacit "contract" after having agreed to participate in the experiment (5 cases).

The first account uncovered among those classified as "defiant" were "unwilling to continue" (e.g., to risk harming L; for religious reasons; against L's will; because L was suffering; to avoid responsibility for harming L) (25 cases). The second account was "faulty design" in the sense that the experiment was thought to be ineffective for studying learning (17 cases). The third explanation (13 cases) was being "unable to continue" (e.g., against L's will; due to nervousness; unable to make L suffer) (Hollander and Turowetz 2017:661–62).

These observations suggest that there was a high degree of heterogeneity in subject reaction to the experimental protocols. Many subjects were taken in by the cover story and were deeply alarmed by the conditions they found themselves in. In the post hoc conversations, they explained resistance primarily out of fear of hurting the learner and skepticism about the utility of punishment in the encouragement of learning. On the other hand, those who were classified as obedient were divided dramatically in the accounts they offered. The most prevalent accounts were those that doubted that any suffering was actually occurring, a significant challenge to the assumption of the effectiveness of Milgram's deception. Other students of the Milgram interviews caution against drawing definitive implications from these tapes because they might be self-exculpatory rhetoric (Gibson et al. 2017). Hollander and Turowetz (2017:658) argue otherwise: "We believe it is possible, and necessary (given the constraints of available data) to make inferences on the basis of post-hoc reports." The claim is based on the immediacy of the interviews within minutes of the experiment, their open-ended nature, and their relative spontaneity. Their conclusions are congruent at least in part with responses to a questionnaire that

Milgram subsequently administered by mail to the subjects.

ORIGINS OF THE SUBJECT QUESTIONNAIRE

Initially Milgram had no plans to follow up with subjects for their feedback on the experiment after the research was over. He planned to supply all subjects with a written debrief to explain the experiment, which had not been provided to the majority of subjects in the laboratory (Perry 2013a). In January 1962, Milgram submitted a second application for continued funding. But responsibility for grant applications in social psychology had changed hands, and Henry Riecken, who had described Milgram's work admiringly as a "bold breakthrough in social psychology" (Milgram 1962a) was replaced by Robert Hall, program director for sociology and social psychology. In April 1962, representatives from the National Science Foundation (NSF), including Hall, visited Yale to observe the experiments in action and to discuss Milgram's fresh application in detail in a three-hour meeting. The NSF's formal refusal of funding for further experiments a month later was unlikely to have been a surprise to Milgram, as the NSF representatives likely expressed their reservations in their meeting, and Hall outlined these reservations explicitly to Milgram over a year later. The reasons NSF terminated funding of the obedience experiments were out of concern for the welfare of subjects and the absence of a basic theoretical framework. He noted that without information about how

> the subjects viewed the situation . . . it would have been impossible to assess whether [the] results were generalizable to real-life situations or were artefacts of a laboratory situation. . . . How do we know if the

situation is really credible to the Ss [subjects]? It seems likely that some of them suspect there was a "catch." This may be operating below the level of awareness of the subject. (Hall 1963)

Milgram replied to acknowledge that he would conduct no further research but asked for additional funds to gather more data on his subjects. On July 12, 1962, he mailed a ten-item questionnaire with the explanatory report to all subjects. His records show he received a total of 730 returns (85 percent) by August 22 (Milgram 1962b). The NSF subsequently gave Milgram one-time funding for the questionnaire analysis and to conduct and transcribe follow-up interviews (Perry 2015).

In his first journal article about his obedience research, Milgram (1963) stressed the dramaturgical credibility of the experiment. He emphasized that "[w]ith few exceptions subjects were convinced of the reality of the experimental situation" (Milgram 1963:375). He vividly described subjects' distress, tension, and emotional strain as evidence for their belief in the experimental scenario. A "mature and initially poised businessman" who was "smiling and confident" at the beginning of the experiment was by the end "reduced to a twitching, stuttering wreck" (Milgram 1963:377). The implication was that the subjects fully believed that what was happening was real, and despite indications that the learner was in increasing pain, 26 out of 40 proceeded to administer the maximum shock. Milgram (1963:376) described the defiant subjects as "frequently in a highly agitated and angered state" but made no reference to their motivation in refusing to continue.

Behind the scenes, as Hollander and Turowetz's (2017) work has suggested, Milgram would have been well aware

from conversations between subjects and the experimenter immediately following the experiment that subjects' perceptions of the experiment showed dramatic variations in subject doubt and belief. After the publication of Diana Baumrind's (1964) critique of his research the following year, Milgram (1964) replied using data from the subject questionnaire items 8, 9, and 10. He argued that the responses suggested that his subjects were overwhelmingly positive about having taken part, supported further experiments in the same vein, and had learned something of personal importance by participating. Milgram also promised that the full results from the questionnaire as well as his interviews with subjects and details of debriefing would be forthcoming. However, he published only one additional item from the questionnaire—item 4, which related to the credibility of the experiment, in Table 7 of his book ten years later (Milgram 1974:172).

When it came to ethical criticisms of his research, Milgram championed the point of view of his subjects, arguing that their views of the situation should be the deciding factor in any assessment of whether or not they had been upset by taking part (Milgram 1977). But while he used responses from the questionnaire data to defend himself against charges of unethical treatment of his subjects, he dismissed data from the same questionnaire when it appeared to challenge his conclusions on methodological grounds (Milgram, in Sabini and Silver 1992). Hoffman, Myerberg, and Morawski (2015:678) note that Milgram dismissed subject skepticism as a "defense function" and a "post facto explanation," that is, an attempt by those who had obeyed to save face by denying any suffering occurred. Fearing contamination of the potential subject pool, Milgram did not conduct a formal debriefing of subjects until after 75 percent of those recruited had

completed the experiment. That meant he had no opportunity to question the majority of the subjects about their belief in the cover story prior to administration of the questionnaire. In a public debate with Martin Orne in 1969, Milgram argued that the subjects' reflections on the experiment contained in the questionnaire were a more valid measure of their reactions than the postexperimental conversations because subjects were so "worked up" immediately after the experiment (Milgram 1969). By the time they received the questionnaire months later, they were "more detached" and presumably better able to reflect on their experience. Describing a typical subject, Milgram said, "He's got time to see his participation in perspective. . . maybe he can speak a little more honestly at that time. As a matter of fact, you get tremendous conclusions from the time factor. . . I think the subjects were basically honest at that point" (Milgram 1969).

In total, the results of six items from the ten-item questionnaire have been published. Milgram (1964) published three reports on the theme of subject well-being and one on subject skepticism (Milgram 1974). In support of Milgram's claim that no harm was done, Blass (2004) published additional data from two more questionnaire items that asked subjects how they felt during the experiment and whether they were bothered by it afterward. The partial and piecemeal publication of these data and the fact that five of the six questionnaire items have been on the theme of subject well-being has tended to keep the focus on ethical as opposed to methodological issues in the obedience experiments.

But methodological misgivings did not abate. Echoing the NSF, Orne and Holland (1968:285) questioned whether the subjects were "taken in" by the experiment and concluded that, without examining this in detail, Milgram and his subjects may well have entered "a pact of ignorance," in which the subject maintains a mask of naivete to safeguard the integrity of the experiment. In fact, Hall (1963) suggested that Milgram "should get post-experimental interviews to find out how Ss perceived various features of the experimental condition." Orne and Holland argued that in social psychological experiments using deception, it was not clear who was deceiving whom. They questioned the plausibility of the experiment and gave credit to subjects as "sentient" beings (1968:285) actively attuned to cues about how they were expected to behave and alert to incongruities in the experimental setup. This was an opportunity for Milgram to reply with a presentation of his questionnaire data on what subjects said they believed during the experiment. He chose not to do so. Milgram received subject feedback, including phone calls and letters, from many former participants that detailed a wide variety of triggers for suspicion. The dog-eared check given to the learner, the cries emanating from what sounded like a speaker, the impassivity of the experimenter in the face of the learner's supposed pain, the experimenters' refusal of teachers' offers to change places with the learner, suspicions that the laboratory mirror was two-way, and the unlikelihood of Yale conducting an experiment involving torture were variously identified by subjects as incongruent elements in the experimental setup (Perry, 2013a).

MURATA'S ANALYSIS

With the questionnaire data complete, in late summer of 1962, Milgram instructed his research assistant, Taketo Murata, to take a closer look at the questionnaire data on the credibility of the deception and its effect on his subjects. This was the only time in the project when Milgram linked the questionnaire data on an

individual level with the subjects' behavior in the experiment. Milgram had obtained five-level scale responses from his questionnaire concerning the degree of conviction that subjects had during the experiment about how real the shocks were, varying from fully believing to fully not believing. He asked Murata to compare the degrees of obedience between those subjects who said they were doubtful that the shocks were painful and those who were certain they were. The report was titled "Reported Beliefs in Shocks and Level of Obedience" (Murata 1962). Murata (1962:1) wrote, "The following is a condition-by-condition analysis to determine whether shock level reached was affected by the extent to which the subject believed that the learner was actually receiving shock." Murata listed conditions 1 through 24, although there is no shock associated with condition 21. This condition contained the estimates given by various groups (psychiatrists, college students, and middle-class adults) of how persons would be expected to act in such an experiment. This condition is omitted in our analysis. The report's main results were a comparison of the mean shock levels administered by subjects who had "fully believed" (FB) in the reality of shocks versus those classified as having "not fully believed" (NFB). The mean shock levels varied from 1 to 30, representing the gradations in shock from 15 to 450 volts in 15-volt increments. Table 1 displays the mean shock level for each group (FB vs. NFB) as well as the numbers in each experiment. Murata found that in 18 of 23 experiments, those subjects who fully believed the learner was getting painful shocks gave *lower* levels of shock than subjects who doubted the shocks were real. In two conditions, the shock levels were equal; and in three conditions, they were higher. Murata conducted a nonparametric sign test to determine whether subtracting the NFB means from the FB

means resulted in an equal distribution of positive and negative effects (as assumed by the null hypothesis). On the evidence (18 negatives, 3 positives), he rejected the null hypothesis (Murata 1962).

AN EXTENDED ANALYSIS

The sign test may provide a quick determination of significant differences in binomial distributions, but it discards the data where the means are equivalent and fails to incorporate other relevant data, such as the differences in the experimental designs. It is possible, however, to provide a robust estimate of any significant difference between means by incorporating the number of observations on which the means in each of the 23 experiments are based. In Table 2, we estimate two ordinary least squares (OLS) regression models in which mean number of shocks constitutes the dependent variable, weighting the results for the number of cases in each experimental condition.[1] In addition to a bivariate model with level of belief as a predictor (Model 1), we added a set of effect-coded variables for the experimental conditions using the "intimate relationship" condition as the reference category (Model 2). In effect, coding the coefficients is interpreted as deviating from the unweighted grand mean of the dependent variable. The effect of the left-out category equals the negative of the sum of all other coefficients.[2] We did not include

---

[1]We checked the distribution of the residuals from these models for normality by graphing them against a generated normal distribution and with a quantile-quantile plot of studentized residuals against the cumulative normal. Both plots indicate that the assumption of normality is met by these data.

[2]For a detailed explanation of effect coding, see Alkharusi (2012). Unlike indicator (dummy) coding, this method involves generating separate variables for each value of a categorical variable by assigning +1, 0, or −1 to each value.

Table 1. Mean Shock Levels by Level of Belief and Experimental Condition

Condition	Experimental condition designation	FB mean	FB n	NFB mean	NFB n	Sign
1	Remote feedback	26.92	11	29.29	17	Negative
2	Voice feedback	26.85	13	24.91	23	Positive
3	Proximity	19.38	26	22.60	10	Negative
4	Touch proximity	17.17	24	18.58	12	Negative
5	Coronary tape: Williams & McDonough	23.30	23	25.73	15	Negative
6	Coronary tape: Elges & Tracy	21.52	23	23.15	13	Negative
7	Groups for disobedience	14.93	15	17.11	19	Negative
8	Qualifying volunteer	19.94	16	24.22	18	Negative
9	Groups for obedience	26.50	16	24.56	18	Positive
10	Benign experimenter	10.00	6	10.00	9	Zero
11	Group choice	15.09	22	19.82	11	Negative
12	Authority as Victim	10.00	13	10.00	3	Zero
13	Non Trigger	27.55	22	30.00	13	Negative
14	Carte blanche	4.19	21	7.19	16	Negative
15	Double Authority (common man shocked)	10.00	8	10.30	8	Negative
16	Double authority (experimenter shocked)	18.00	5	24.44	9	Negative
17	Common Man Authority	25.83	6	23.10	10	Positive
18	Experimenter Departs	20.23	22	25.93	14	Negative
19	No Experimenter	18.24	21	20.57	14	Negative
20	Women	24.76	21	26.06	16	Negative
22	Common-man authority 2	14.88	8	17.89	9	Negative
23	Bridgeport replication	19.50	14	25.56	9	Negative
24	Intimate relationship	14.55	11	17.00	3	Negative
	Average means and total frequencies ($N = 656$)	19.05	367	21.73	289	

Note: FB refers to subjects who ''fully believed'' in the reality of the shocks; NFB refers to those who had ''not fully believed'' in it. Sign refers to the difference between the FB mean and the NFB mean.

indicators of level of significance in this table, since we were able to reject the null hypothesis at $p < .001$ for all coefficients, with the one exception noted in further discussion.

In the bivariate model, the effect of belief is sizeable, –2.66. The negative sign means that those who fully believed that the shocks were real on average delivered 2.66 fewer shocks. Model 2 adds the experimental conditions as predictors. Every experimental condition, with the sole exception of the "no experimenter" condition, had a significant positive or negative effect on number of shocks compared to the unweighted grand mean. Although the odds ratio associated with belief declines to 2.158, it is still large and significant beyond .001. These results also suggest that the levels of obedience/defiance, controlling for belief in harm, were highly variable across the various experimental groups and responsive to the specific experimental designs. The "carte blanche" condition permitted the teachers to choose their own levels of shock. This reduced shocks by nearly 14 points compared to the mean over all conditions, by far the largest effect. When the "benign

Table 2. Frequency-Weighted Ordinary Least Squares Regressions of Mean Shock Levels on Level of Belief and Experimental Condition

Predictor variable	Model 1	Model 2
Level of belief		
Did not fully believe (reference category)		
Fully believed	–2.664	–2.158
Experimental condition		
Carte blanche		–13.985
Benign experimenter		–9.834
Double authority (common man shocked)		–9.553
Authority as victim		–8.943
Intimate relationship		–3.925
Groups for disobedience		–3.597
Common-man authority 2		–3.208
Group choice		–2.592
Touch proximity		–1.618
No experimenter		–0.230
Proximity		1.136
Double authority (experimenter shocked)		2.214
Bridgeport replication		2.488
Qualifying volunteer		2.524
Coronary tape: Elges & Tracy		2.790
Experimenter departs		3.069
Common-man authority		4.236
Coronary tape: Williams & McDonough		4.868
Voice feedback		5.693
Groups for obedience		5.791
Women		5.850
Remote feedback		7.706
Nontrigger		9.120
Constant	21.723	28.403
Observations	656	656
R^2	0.047	0.968

Note: Regressions weighted by number of cases in each condition. Experimental conditions use effect coding with conditions ranked on size of effect. All coefficients significant beyond $p < .001$ except ''no experimenter present.''

experimenter'' encouraged the teachers to desist after complaints from the learner, this reduced the teachers' shocks by 9.8 points. When the experimenter demonstrated the learner's task (''authority as victim'') and a confederate teacher insisted on his punishment, the maximum shock level given by the naive teacher was reduced by 8.9 shocks. Our findings acknowledge that, in addition to variability in belief, levels of shock are dramatically responsive to variations in the conditions of interaction, which systematically varied the balance of power across experiments.

It is noteworthy that the level of belief (Model 1) accounts for just 4.7 percent of the variance explained in mean shock, even though the effect is robust and statistically significant. With the introduction of the set of indicator variables for the experimental condition (Model 2), the percentage of variance explained rises to nearly 97 percent, suggesting that most of the variation observed in the mean number of shocks is due to how

the experiment was structured. We would note, however, that such a large value for the *R*-squared is in part due to the fact that the data have no individual-level variance, that is, all the variables are summary measures.

Milgram never published Murata's results directly, but he did employ them indirectly. Milgram (1974:172) summarized the patterns of defiance versus obedience contingent on the five levels of belief in harm that were collected in the questionnaire. Murata dichotomized belief into "fully believing" versus "not fully believing." "Not fully believing," however, was actually composed of four levels of disbelief: (1) those who had *some* doubts but thought the learner was *probably* getting shocked, (2) subjects who were *not sure*, (3) subjects who thought the learner was *probably not* getting shock, and (4) subjects *certain* of no shock. Milgram's table contained 658 cases, virtually identical to the 656 employed by Murata.

The relationship between the five levels of belief and the two levels of obedience based on Milgram's original publication is reported in Table 3. We differ from Milgram's presentation of the data (see Appendix) by analyzing defiance versus obedience for the different levels of belief (using estimates weighted by number of observations). Surprisingly, Milgram percentaged the table in the wrong direction in his 1974 book. This is a critical point. The Murata analysis was premised on treating the level of obedience as the dependent variable and the level of belief as the independent variable, that is, whether the shock level was affected by the level of belief. Milgram's (1974) presentation of the data conceals this. Milgram concluded famously that "three-quarters of the subjects (the first two categories) by their own testimony acted under the belief that they were administering powerful shocks" (1974:172). This

Table 3. Cross-Tabulation of Obedience/ Defiance by Level of Belief

Level of belief	Obedient	Defiant	Total
1 (Belief high)	139	230	369
	37.7%	62.3%	100%
2	75	83	158
	47.5%	52.5%	100%
3	18	22	40
	45%	55%	100%
4	47	28	75
	62.7%	37.3%	100%
5 (Belief low)	11	5	16
	68.8%	31.2%	100%
Total	290	368	658
	44.1%	55.9%	100%

Note: Pearson chi-square = 21.364, $p < .001$, with 4 degrees of freedom. Cells show number of cases and row percentages.

approach leads to some mistaken inferences. In point of fact, referring to the original table, only the first category of subjects (47.9 percent) expressed *no* doubt, that is, less than half, not three-quarters. In fact, fully 20 percent of the obedient subjects were skeptical about the reality of the shocks. Additionally, Milgram's approach suppresses the *relative* impact of the different levels of belief on obedience and defiance. Milgram's reference to "three-quarters of the subjects" appears to have been referring specifically to the *obedient* group summarized vertically in the original table (47.9 + 25.9 = 73.8 percent). But what he fails to mention is that in the same table where belief in the shocks was greatest (62.5 + 22.6 = 85.1 percent), this was the category in which subjects were *most defiant*. It is hard to imagine the impact of Milgram saying "85 percent by their own testimony refused to administer powerful shocks when acting under the belief that they were painful." Conventionally, tables are percentaged in the direction of the independent variable, as presented in Table 3. This approach captures the fact that as subjects became

Table 4. Frequency-Weighted Logistic Regressions of Defiance/Obedience by Dichotomized Level of Belief and Ordinal Level of Belief

Variable	Belief dichotomized	Degree of belief
Belief dichotomized		
Belief high (belief low = reference category)	2.571***	
Degree of belief (ordinal)		
Belief highest = 1		3.640*
2		2.435
3		2.689
4		1.311
Belief lowest = 5 (reference category)		
Observations	618	658
Pseudo R^2	0.020	0.024
Likelihood-ratio χ^2	16.78***	21.38***

Note: Dichotomized level of belief does not include level 3. Coefficients are odds ratios.
*$p < .05$. **$p < .01$. ***$p < .000$.

less convinced of pain, their defiance decreased and their obedience increased. Also, Milgram's original table obscured the observation that across all conditions, the majority—368 out of the 658 subjects (56 percent)—were defiant, not obedient.

The cross-tabulation in Table 3, in which the chi-square statistic testing the association is significant, shows in percentage terms that defiance generally declines significantly from the highest level of belief in pain to the lowest and, conversely, that obedience increases with skepticism of pain. It is difficult to identify with precision an effect size on obedience/defiance across levels of belief in such a cross-tabulation. In order to determine the effect size, we estimated frequency-weighted logistic regressions to identify changes in defiance across levels of belief. The models shown in Table 4 report the effects of level of belief and experimental condition. In Model 1, we dichotomized the belief variable used by Milgram to contrast the top two versus the bottom two measures of belief, while omitting the middle category, to provide an overall estimate of effect of belief on defiance. This resulted in an odds ratio of 2.57, suggesting that those who had

a high level of belief that the shocks were real were 2.57 times more likely to be defiant than those who had a low level of belief. No matter whether one focuses on obedience or defiance, the effects of levels of belief are highly significant, both statistically and substantively.

Model 2 shows that compared to those in the baseline category (belief level 5), the level of defiance increases by 3.7 times when the subjects fully believed the learner was being shocked compared to being certain that the learner was not being shocked. Even though the effects of the intermediate degrees of belief are not statistically significant, this provides even more compelling evidence that belief in the reality of the shocks strongly affected subjects' obedience or defiance. More generally, this means that variations in dramaturgical credibility resulted in dramatic variations in the levels of obedience and defiance.

The effect of level of belief on the number of shocks administered, shown in Table 2, and the strong effect of level of belief on defiance, shown in Table 4, are complementary in their implications. They both indicate that those who were stronger in their belief that the shocks

were real both administered fewer shocks and were more likely to be defiant. The effect of belief on obedience/defiance, however, is much stronger than the effect on number of shocks. Of course, we cannot make a direct comparison, since one effect is captured by an odds ratio and the other by an OLS regression coefficient. Nevertheless, the effect on number of shocks, –2.664, while statistically significant, is relatively small compared to the overall mean number of shocks across all designs, 19.05. The odds ratio associated with the dichotomous version of belief, however, indicates that those with a high level of belief were over two-and-a-half times more likely to express defiance, a very large, significant effect. An even larger effect is observed when we compare those who fully believed that the shocks were real to those who were certain that the learners were not experiencing shocks.

DISCUSSION

This analysis, like other studies based on the Milgram archives, raises critical questions about the significance and meaning of the obedience studies. Like Hollander and Turowetz's (2017) analysis of postexperimental admissions, it suggests that both the perception of pain and skepticism about inflicting pain appear to have been significantly associated with the behavior of the subjects. Miller (2014:216) says that "understanding precisely why different specific individuals in the same experimental condition decide to obey or withdraw remains a major unanswered question in the obedience story." The results reported here, which control for different individuals in the same conditions, suggest that one significant explanation for their behavior was whether they thought the experiment was actually painful to anyone. If we examine the research in its entirety

across the 23 experiments, the majority of subjects perceived that the learner was suffering and defied the experimenter. Those who were less successfully convinced by the cover story were more obedient.

There are several caveats that should temper the interpretation of results. First, while 85 percent of the questionnaires on which Murata's analysis was based were returned, it is impossible to determine whether there was a systematic bias in the return rate and how this might have influenced the results. This is unlikely, however, given the high return rate.

Second, there are differences in the conclusions reached by Murata and Milgram due in part to the different approaches to their data analyses. Murata employed only two levels of belief (full belief versus everything else), while Milgram used all five levels of belief, from full belief to full skepticism. Similarly, Murata used the actual mean score levels of shocks, while Milgram used complete obedience versus any level of defiance. As a result, the mean difference reported by Murata is modest but significant (2.66), and the explained variance is modest (4.53 percent). The effect size in Milgram, however, with the larger variance in the measure of belief, yields a highly substantive measure of the impact of belief on defiance/obedience reflected in the changing odds ratios (from the baseline to the highest level of belief, an increase to a factor of over 3.6).

There is a third issue, which concerns the very idea of belief, what it represents, and when and if it can be measured reliably. This is a more intractable problem. Hollander and Turowetz (2017) take the view that the immediate open-ended experimental *conversations* probably yield relatively reliable and valid measures of the subjects' perceptions of what had transpired in the previous hour,

when their memories were vivid. Their conclusions, however, based on a painstaking methodology, rely on a sample of about 10 percent of all the subjects. In contrast, in a public debate with Orne, Milgram (1969) argues that the *questionnaire* results, which are based on a plurality of all subjects, are more valid because the passage of time had allowed the nerves to cool and the experience to be put in some perspective. He claims at other times, however, that denial of pain is not an automatic token of belief, since it may reflect a defense mechanism to manage shame. In a short preface to Murata's (1962) paper, he says that "probably some of it is a real relationship and some of it is defensive," although he never explains how one would establish the difference. This disparity highlights the difficulty of drawing definitive conclusions from either type of evidence.

Gibson (2013b) and Gibson et al. (2017) raise the situated nature of accounts containing the expressions of belief and what these utterances achieve as speech acts in specific settings. If we adopt their perspective, it follows that the social scientist cannot arbitrarily privilege one context of utterance and expression over others in the search for evidence of "real belief." The positions expressed in the structured questionnaires may or may not have correlated with the postexperimental utterances offered in explanations for the subjects' choices. This was not measured. One line of inquiry that might cast light on this would be to probe expressions of belief over time (after the experiment, in a later interview, and in a subsequent questionnaire) to establish consistency, but such repeated intrusions would themselves be subject to question for inadvertently coaching, if not implanting, recollections of experience.

In addition, it remains to be determined whether differences in debriefing influenced the subsequent perception of the experiment generally and the responses to questions about perceptions of harm in particular (see Turowetz and Hollander 2018). Among the 25 percent of cases that experienced the full post hoc oral debriefing were the "women" condition (which had one of the highest levels of obedience) and the "intimate relationship" condition (which had one of the lowest). Any differences in belief in pain that might be attached to the debriefing process would be significantly confounded by differences in the actual experimental conditions as identified in Table 2.

A separate issue is the potential inconsistency between the post hoc interviews and the later questionnaire data. This remains a thorny problem not only for us but for anyone who wishes to advance an explanation for differences in subject behavior based on inferences about their mental states, including Milgram. It was Milgram whose research raised the issue of obedience and its responsiveness to perceptions of harm. If his premise was correct, then our work has raised an ironic linkage between belief and obedience that was unexplored by Milgram and the obedience literature. One of the potentially redeeming elements of our revisiting the Murata paper, and accepting, even provisionally, the validity of the questionnaire responses, is that it re-centers the issue of responsibility and agency that Milgram's work tended to overlook. Milgram's work appears to have been in denial of two salient facts. On the one side, many subjects were not effectively hoaxed, as noted by Orne and Holland (1968) and Baumrind (1964). On the other side, when the studies are taken as an integrated paradigm of research, the most prevalent event that Milgram recorded, contrary to what he reported, was defiance. Even if the quantum of influence presented here is not enormous, a significant contributor to subject

defiance appears to be the reported perception of pain that restrained aggression, a factor that the obedience literature has failed to fully appreciate.

THEORETICAL RELEVANCE

Our analysis offers little support for the theory that Milgram's subjects, despite the usual dictates of conscience, became inattentive functionaries who inflicted painful shocks against an innocent person indifferent to the resulting consequences (see Fenigstein 2015). This perspective is associated with Milgram's initial speculation about the agentic state and has become largely discredited (Brannigan and Perry 2016). The dominant alternative explanation is the engaged followership perspective advanced by Haslam and Reicher (2012) in a remarkable series of publications that cover the work of Zimbardo as well as that of Milgram. These scholars suggest that obedient subjects identify with the experimenter rather than the learner, and follow the experimenter's directions, believing themselves to be "contributing to a moral, worthy, and progressive cause" (Haslam et al. 2015:60). Teachers administer shocks in the service of this larger cause. Even though they may suspect that what they are doing is wrong, they "believe it to be right" (Haslam et al. 2015:78). Our analysis of the Murata (1962) evidence is largely inconsistent with this conclusion, given that the majority of subjects defied the "scientific" pressure to conform and did so in response to the perceived suffering of the learner. For those who continued, in what was at times a highly ambiguous and stressful situation, at least some suspected that there was deception at work and proceeded on the basis that the learner was not being harmed. This is consistent with the findings of Hollander and Turowetz (2017) that (a) disbelief in

harm was the predominant response in obedient subjects and (b) fear of hurting the learner was a most salient account for defiance.

On the basis of their analysis of responses to Milgram's collection of subjects' comments obtained from the questionnaire, Haslam et al. (2015) conclude that most subjects were "happy to have been of service" and "the majority professed that they were happy about their participation" (2015:79). This is an important foundation for the engaged followership perspective. How does one resolve this discrepancy between Murata's quantitative analysis of data from the same questionnaire? Haslam et al.'s analysis drew comments from one subset (section 13) of 1,057 transcribed comments that Milgram had classified into 17 thematic groups. The 140 comments Haslam et al. employed had been labeled by Milgram as "thoughts about the value of having participated in the research." It is unclear, however, how many individuals made comments that appeared in section 13, let alone whether they represented the majority of subjects. In addition, Milgram notably included section 6, which dealt with "feelings and suspicions during study." This included 131 entries. This may have included feelings of suspicion that later eventuated in doubt. Again, we have the problem that these are counts of comments, not individuals, and that individuals may have made comments that ended up in various thematic sections. In our view, relying exclusively on section 13 to make the case for engaged followership seems to dismiss prematurely the potential relevance of the majority of comments. It is also regrettable that the comments were not linked systematically to the subjects' record of obedience, as was the case in Murata's (1962) employment of questionnaire data.

Our findings suggest a degree of subject agency that challenges the power

hierarchy implied by the binary of leader–follower and the emphasis on the attachment to the scientist upon which the engaged followership model is based. It opens up a multiplicity of potential motivations for subject behavior heterogeneity in both belief and the situational structure. This was the conclusion drawn by Hollander and Turowetz (2018:307) when they advocated "a variety of reasons" for persons acting in obedience-constraining situations, including a commitment to scientific objectives, that is, engaged followership. Turowetz and Hollander similarly argued that "no single social psychological process uniquely suffices to explain [subject] actions but rather that compliance resulted from multiple processes involving a complex interplay of situational forces and individual dispositions" (2018:89). A problem with the engaged followership perspective is that acknowledgement of its broad reach, when widened beyond "the epistemic capital of science" (Haslam and Reicher 2018:294) to include the attachments related to religion, professional obligations, concern with living up to a contract, and fear of legal liability, tends to lose its specific traction as a factor encouraging obedience. Haslam et al. (2015:67) mention the case of the lawyer who broke off because he thought he would be held to a higher legal standard. Another individual claimed that his military training made him reluctant to break off since that was not what professional soldiers are in the habit of doing. Other comments concerned the contractual payment and the implied obligation to complete the assignment. In 1961, the average salary in Connecticut was about $2.40 per hour, so $4.50 for participation and transportation was probably a very attractive source of income for many workers (U.S. Bureau of the Census 1976:380). By implication, the subjects were pulled by a number of strands of prior institutional engagements over and above those arising from the prestige of science.

The recent exchange between Haslam and Reicher (2018) and Turowetz and Hollander (2018) was preoccupied with obedience. One contribution in this article is to shift emphasis to the 56 percent of subjects who resisted the experimenter and exercised a degree of self-control and independence by breaking off. The majority of subjects disobeyed the experimenter, and such resistance was correlated with a belief that the learner seemed to be in pain, among other things. This suggests that Milgram's framing of the experiments as an explanation for atrocities of the Holocaust has constrained discussion and understanding of the subject behavior in terms of aggression. Our position points in contrast to notions of empathy and altruism as dominant subject reactions to the experiment.

PERENNIAL QUESTIONS ABOUT EXPERIMENTATION, MILGRAM, AND THE ARCHIVES

In the late 1960s and early 1970s, the obedience study was ineluctably drawn into the so-called crisis in social psychology (Stam, Radke, and Lubek 2000). Milgram's work was challenged in the first instance for its questionable ethics but later attracted doubt as social psychology questioned the epistemological status of the experiments. As the crisis dissipated and social psychologists moved on from deceptive experimentation, the obedience experiments continued to live in the textbooks of social psychology and in popular representations, and the mythology around the experiments only grew. In the early 2000s, as partial replications were undertaken and the archives became available to scrutiny by historians, Milgram's research once again became the focus of critique (Burger, Girgis, and Manning 2011). Milgram's was among a vast swath

of classical psychological studies whose findings were debatable (Danziger 2000; Haslam and Reicher 2012, 2017), as were Zimbardo's (Le Texier 2018). The recent "replicability crisis" has renewed interest in Milgram's actual research achievements. There is one difference, however, between this current crisis and that of the late 1960s: the archives have sparked a renaissance for understanding Milgram by introducing new evidence that assists in interpreting obedience, defiance, and related experimental behavior. They raise questions with both what the obedience research has actually accomplished and its theoretical relevance. In that respect, the current series of critiques may be serious enough to warrant the reconsideration of the studies in toto and to call into question the utility of deception, except where it is "absolutely necessary," as recommended by Cook and Yamagishi (2008).

CONCLUSION

Based on the work of Taketo Murata (1962), this article has tackled the high levels of heterogeneity in subject perceptions of pain in the Milgram obedience experiments. This analysis does not assume to propose the last word on Milgram's ingenious experiments, however. What we offer is a multivariate interpretation based on an acknowledgment of the simultaneous operation of a variety of different and often diverging and converging behaviors. It has also been our purpose to explore the narrative about "Milgramesque behaviors" to capture the high levels of defiance and their foundations in the commonsense perception of harm as well as the justifications for resistance identified in qualitative studies—religious reasons, unwillingness to hurt the learner, rejection of the premise that punishment expedites learning, doubt that Yale would permit torture, and so on. This is part of a shift away from the historical assumption that subject behavior was primarily about "obedience" and introduces neglected notions of empathy and altruism as influences on subject behavior. It does not dismiss the engaged followership explanation for obedience but proposes that this explanation is by no means the only one. We have stressed that defiance was associated on the one side with the perception of pain and obedience was associated with subject skepticism on the other. The key finding of this study, that obedience to authority is not as unreasoning and automatic as Milgram would have had us believe, ought to encourage significant revisions in fair-minded textbooks and other historical accounts of the development of social psychology.

APPENDIX

Milgram's Original Table

Table 7. Responses to Question on Belief

During the Experiment	Defiant	Obedient	All subjects
(1) I fully believed the learner was getting painful shocks.	62.5% (230)	47.9% (139)	56.1% (369)
(2) Although I had some doubts, I believed the learner was *probably* getting the shocks.	22.6% (83)	25.9% (75)	24.0 (158)
(3) I just wasn't sure whether the learner was getting the shocks or not.	6.0% (22)	6.2% (18)	8.1% (40)
(4) Although I had some doubts, I thought the learner was probably not getting the shocks.	7.6% (28)	16.2% (47)	11.4% (75)
(5) I was certain the learner was not getting the shocks.	1.4% (5)	3.8% (11)	2.4% (16)

Source: Milgram (1974:172).

ACKNOWLEDGMENTS

ORCID ID

Gina Perry https://orcid.org/0000-0002-3007-0996

REFERENCES

Alkharusi, Hussain. 2012. "Categorical Variables in Regression Analysis: A Comparison of Dummy and Effect Coding." *International Journal of Education* 4(2):202–10.

Baumrind, Diana. 1964. "Some Thoughts on Ethics of Research: After Reading Milgram's 'Behavioral Study of Obedience.'" *American Psychologist* 19:421–23. http://dx.doi.org/10.1037/h0040128.

Blass, Thomas. 2000. *Obedience to Authority: Current Perspectives on the Milgram Paradigm*. Mahwah, NJ: Lawrence Erlbaum.

Blass, Thomas. 2004. *The Man Who Shocked the World: The Life and Legacy of Stanley Milgram*. New York: Basic Books.

Brannigan, Augustine, and Gina Perry. 2016. "Milgram, Genocide and Bureaucracy: A Post-Weberian Perspective." *State Crime Journal* 5(2):287–305. http://www.jstor.org/stable/10.13169/statecrime.5.2.0287.

Burger, Jerry M., Zachary M. Girgis, and Caroline C. Manning. 2011. "In Their Own Words: Explaining Obedience to Authority through an Examination of Participants' Comments." *Social Psychological & Personality Science* 2(5):460–66. doi:10.1177/1948550610397632.

Cook, Karen S., and Yamagishi T. 2008. "A Defense of Deception on Scientific Grounds." *Social Psychology Quarterly* 71(3):215–21.

Danziger, Kurt. 2000. "Making Social Psychology Experimental: A Conceptual History, 1920– 1970." *Journal of the History of the Behavioral Sciences* 36:329–47.

Fenigstein, Allan. 2015. "Milgram's Shock Experiments and the Nazi Perpetrators: A Contrarian Perspective on the Role of Obedience Pressures during the Holocaust." *Theory and Psychology* 25(5):581–98.

Gibson, Stephen. 2013a. "'The Last Possible Resort': A Forgotten Prod and the in Situ Standardization of Stanley Milgram's Voice-Feedback Condition." *History of Psychology* 16:177–94. https://doi.org/10.1037a0032430.

Gibson, Stephen. 2013b. "Milgram's Obedience Experiments: A Rhetorical Analysis." *British Journal of Social Psychology* 52:290–309. https://doi.org/10.1111/j.2044-8309.2011.02070.x.

Gibson, Stephen, Grace Blenkinsopp, Libby Johnstone, and Aimee Marshall. 2017. "Just Following Orders? The Rhetorical Invocation of 'Obedience' in Stanley Milgram's Post-experiment Interviews." *European Journal of Social Psychology*. https://doi.org10.1002/ejsp.2351.

Griggs, Richard A., and George I. Whitehead. 2015. "Coverage of Recent Criticisms of Milgram's Obedience Experiments in Introductory Social Psychology Textbooks." *Theory and Psychology* 25(5):564–80. doi:10.1177/095935431560123.

Hall, Robert. 1963. Letter to Stanley Milgram from Robert Hall dated November 13, National Science Foundation. Stanley Milgram Papers, Yale University Library, Box 43, Folder 128.

Haslam, S. Alexander, and Stephen D. Reicher. 2012. "Contesting the 'Nature' of Conformity: What Milgram and Zimbardo Studies Really Show." *PLOS Biology* 10(11):e1001426.

Haslam, S. Alexander, and Stephen D. Reicher. 2017. "50 Years of 'Obedience to Authority': From Blind Conformity to Engaged Followership." *Annual Review of Law and Social Science* 13:59–78.

Haslam, S. Alexander, and Stephen D. Reicher. 2018. "A Truth That Does Not Speak Its Name: How Hollander and Turowetz's Findings Confirm the Engaged Followership Analysis of Harm-Doing in the Milgram Paradigm." *British Journal of Social Psychology* 57:292–300. https://doi.org/10.1111/bjso.12247.

Haslam, S. Alexander, Stephen D. Reicher, Kathryn Millard, and Rachel McDonald. 2015. "'Happy to Have Been of Service': The Yale Archive as a Window into the Engaged Followership of Participants in Milgram's 'Obedience' Experiments." *British Journal of Social Psychology* 54:55–83. https://doi.org/10.1111/bjso.12074.

Hoffman, Ethan, N. Reed Myerberg, and Jan G. Morawski. 2015. "Acting Otherwise:

Resistance, Agency and Subjectivities in Milgram's Studies of Obedience." *Theory & Psychology* 25(5):670–89. https://doi.org/10.1177/0959354315608705 17.

Hollander, Matthew H., and Jason Turowetz. 2017. "Normalizing Trust: Participants' Immediately Posthoc Explanations of Behaviour in Milgram's 'Obedience' Experiments." *British Journal of Social Psychology* 56:655–74. https://doi.org/10.1111/bjso.12206.

Hollander, Matthew H., and Jason Turowetz. 2018. "Multiple Compliant Processes: A Reply to Haslam and Reicher on the Engaged Followership Explanation of 'Obedience' in Milgram's Experiments." *British Journal of Social Psychology* 57:301–309. doi:10.1111/bjso.12252.

Kaposi, David. 2017. "The Resistance Experiments: Morality, Authority and Obedience in Stanley Milgram's Account." *Journal for the Theory of Social Behavior* 47(4):382–401. https://doi.org/10.1111/jtsb.12137.

Le Texier, Thibault. 2018. *Histoire d'un Mensonge: Enquête sur l'expérience de Stanford*. Paris: E?ditions La Decouverte.

Lunt, Peter. 2009. *Stanley Milgram: Understanding Obedience and Its Implications*. London: Palgrave MacMillan.

Milgram, Stanley. 1962a. Letter to Jerome Bruner from Stanley Milgram dated February 6. Stanley Milgram Papers, Yale University Library, Box 1a, Folder 3.

Milgram, Stanley. 1962b. Obedience log book. Stanley Milgram Papers, Yale University Library, Box 46, Folder 163.

Milgram, Stanley. 1963. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology* 67(4):371–78. https://doi.org/10.1037/h0040525.

Milgram, Stanley. 1964. "Issues in the Study of Obedience: A Reply to Baumrind." *American Psychologist* 19(11):848–52. http://dx.doi.org/10.1037/h0044954.

Milgram, Stanley. 1969. Debate between Stanley Milgram and Martin Orne, University of Pennsylvania, February 19. Stanley Milgram Papers, Yale University Library, Box 22.

Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View*. New York: Harper and Row.

Milgram, Stanley. 1977. "Subject Reaction: The Neglected Factor in the Ethics of Experimentation." *Hastings Center Report* 7(5):19–23.

Miller, Arthur. G. 2014. "Why Are the Milgram Experiments So Extraordinarily Famous—and Controversial?" Pp.185–223 in *The Social Psychology of Good and Evil*, edited by A. G. Miller. New York: Guilford Press.

Mixon, Don. 1977. "On the Difference between Active and Nonactive Role-Playing Methods." *American Psychologist* 32(8):676–77.

Modigliani, Andre R. F. (1995). "The Role of Interaction Sequences and the Timing of Resistance in Shaping Obedience and Defiance to Authority." *Journal of Social Issues* 51(3):107–23. https://doi.org/10.1111/j.1540-4560.1995.tb01337.x/.

Murata, Taketo. 1962. "Reported Belief in Shocks and Level of Obedience." Research report. Stanley Milgram Papers, Yale University Library, Box 45, Folder 148.

Nicholson, Ian. 2011. "'Torture at Yale': Experimental Subjects, Laboratory Torment and the 'Rehabilitation' of Milgram's 'Obedience to Authority.'" *Theory & Psychology* 21(6):737–61. https://doi.org/10.1177/0959354311420199.

Orne, Martin T., and Charles H. Holland. 1968. "On the Ecological Validity of Laboratory Deceptions." *International Journal of Psychiatry* 6(4):282–93.

Perry, Gina. 2013a. *Behind the Shock Machine. The Untold Story of the Notorious Milgram Psychology Experiments*. New York: New Press.

Perry, Gina. 2013b. "Deception and Illusion in Milgram's Accounts of the Obedience Experiments." *Theoretical & Applied Ethics* 2(2):79–92. Project MUSE database.

Perry, Gina. 2015. "Seeing Is Believing: The Role of the Film *Obedience* in Shaping Perceptions of Milgram's Obedience to Authority Experiments." *Theory & Psychology* 25(5):622–38. htpps://doi.org/10.1177/0959354315604235.

Russell, Nestar J. C. 2011. "Milgram's Obedience to Authority Experiments: Origins and Early Evolution." *British Journal of Social Psychology* 50(1):140–62.

Russell, Nestar J. C. 2018. *Understanding Willing Participants: Milgram's Obedience Experiments and the Holocaust*, 2 vols. London: Palgrave Macmillan.

Sabini, John, and Maury Silver, eds. 1992. *The Individual in a Social World: Essays and Experiments*. New York: McGraw Hill.

Stam, Henderikus. J., Lorraine L. Radtke, and Ian Lubek. 2000. "Strains in Experimental Social Psychology: A Textual Analysis of the Development of Experimentation in Social Psychology." *Journal of the*

History of the Behavioral Sciences 36(4):365–82.

Turowetz, Jason, and Matthew H. Hollander. 2018. "From 'Ridiculous' to 'Glad to Have Helped': Debriefing News Delivery and Improved Reactions to Science in Milgram's 'Obedience' Experiments.'" *Social Psychology Quarterly* 81(1):71–93. https://doi.org/10.1177/0190272518759968.

U.S. Bureau of the Census. (1976). *Statistical Abstract of the United States 1976*. Washington, DC: Government Printing Office. https://babel.hathitrust.org/cgi/pt?id=mdp.39015021301612; view=1up;seq=3.

Yale. 2017. Guide to the Stanley Milgram Papers. MS 1406, July 1993, revised April 2017. Yale University Library Manuscripts and Archives. http://hdl.handle.net/10079/fa/mssa.ms.1406.

BIOS

Gina Perry *is an Australian writer and author of Behind the Shock Machine: The Untold Story of the Notorious Milgram Psychology Experiments* (2012) and *The Lost Boys: Inside Muzafer Sherif's Robber Cave Experiment* (2018). Both works draw on extensive archival research and interviews with experimental participants. She completed her PhD at the University of Melbourne, where she is an associate in the School of Culture and Communication.

Augustine Brannigan is professor emeritus of sociology at the University of Calgary in Canada. His most recent major work was *Beyond the Banality of Evil: Genocide and Criminology* (2013), based in large part on his fieldwork on the Rwandan genocide. Earlier notable books include *The Rise and Fall of Social Psychology: The Use and Misuse of the Experimental Method* (2004) and *The Social Basis of Scientific Discoveries* (1981).

Richard A. Wanner is professor emeritus in the sociology at the University of Calgary and academic director of the Prairie Regional Research Data Centre. His recent research has focused on public perceptions of the Olympics, the effects of immigration on the Canadian economy, internal migration, and trends in social stratification. His work has appeared in many scholarly journals, including the *Journal of Urban Affairs, International Migration Review, Canadian Journal of Sociology*, and *Sociology.*

Henderikus Stam is a professor of psychology at the University of Calgary, and an adjunct professor in the Department of History. His recent work has focused on contemporary theoretical problems in psychology and the historical foundations of twentieth-century psychology. In 2015 he was honored with the American Psychological Foundation's Joseph Gittler Award for his contributions to the philosophy of psychology.