Reflections on "Replicating Milgram" (Burger, 2009)

Arthur G. Miller
Miami University

In "Replicating Milgram: Would People Still Obey Today?" Jerry M. Burger (2009) reported a high base rate of obedience, comparable to that observed by Stanley Milgram (1974). Another condition, involving a defiant confederate, failed to significantly reduce obedience. This commentary discusses the primary contributions of Burger's study in terms of (a) its novel methodological variation on Milgram's original paradigm (the "150-volt solution") and (b) its attention to ethical concerns so as to minimize participant discomfort and ensure institutional review board approval. Burger's technique could unlock research on behavioral aspects of obedience, which has been essentially muted for several decades. However, Burger's intensive efforts to improve the ethics of the study may be exaggerated, are uncertain in their effectiveness, and pose impractical demands. Different procedures used by Milgram and Burger in the modeled refusal condition preclude a clear explanation for the results and challenge Burger's emphasis on the comparability of his and Milgram's experiments. This study documents the complexities of extending research on destructive obedience in the context of contemporary ethical guidelines.

Keywords: Milgram obedience experiments, ethical issues in research with human participants, methodological aspects of obedience research, institutional review boards and the Milgram experiment

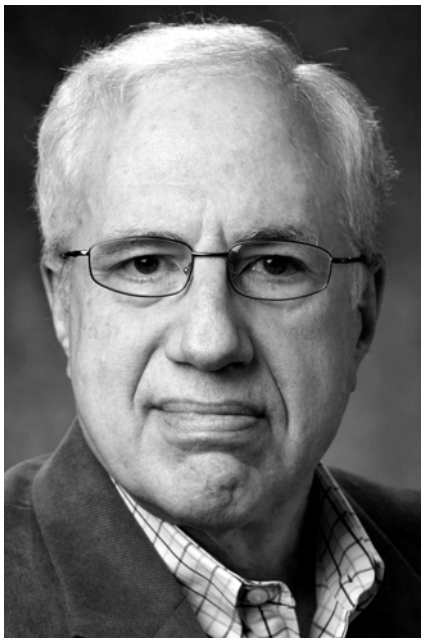
Upon seeing at first the title of Jerry Burger's (2009) article—"Replicating Milgram"—it was as if "Breaking News!" had flashed across my screen, a stunning, even (dare I say) shocking, announcement. Academics are, of course, curious about any new paper in their area of interest, particularly one by an experienced, highly regarded researcher, but this was something quite different and special. If one digs deeply enough in the archives, one will locate several previous replications of Milgram's shock-administration paradigm (for reviews, see Blass, 2000; Miller, 1986, chap. 4) as well as research using different paradigms and measures of obedience (e.g., Brief, Buttram, Elliott, Reizenstein, & McCline, 1995; Meeus & Raaijmakers, 1995). These studies documented rates of obedience to authority (and the influence of situational factors on these rates) comparable to those shown by Milgram (1974). However, most of these experiments occurred decades ago. Research on (behavioral) obedience to authority has been virtually nonexistent in recent years, a peculiar cessation given the unparalleled and nonabating interest in Milgram's research. I had assumed that the time for fresh inquiry and insights on this intriguing work was long past. Burger's "Replicating Milgram" thus provides an invaluable inducement to reexamine the obedience research. What accounts for the long dry spell?

I think that many researchers may have been intimidated in taking on the problem of destructive obedience, primarily because Milgram's work had assumed rather quickly a kind of larger than life essence. After the appearance of his 1974 book, Milgram seemed, in effect, to have had the last word on the subject. Commenting upon the unprecedented impact of Milgram's findings, Ross and Nisbett (1991) eloquently anointed the obedience experiments as "part of our society's shared intellectual legacy" (p. 55). Thus, when contemplating a new obedience investigation, it is not simply a matter of designing a "follow-up" study in the usual sense. Instead, the researcher enters a veritable universe of commentary and controversy. (Witness the articles in this issue.) In this context, generations of researchers, though tempted to follow Milgram's lead, may have been hesitant to invite the unprecedented criticism and hostility that Milgram himself experienced. Conversely, one might think that the glaring celebrity status of the obedience experiments would have, in itself, constituted a powerful inducement for active research in this area, similar to that prompted by Asch's (1956) highly influential conformity paradigm. However, this did not occur. (For a provocative critique of the manner in which Asch's research has been interpreted by social psychologists, see Hodges & Geyer, 2006.)

Another, perhaps more verifiable, rationale for the attenuation of obedience research has been the daunting ethical features of the Milgram paradigm. One frequently encounters the view that it is no longer possible to conduct this type of experiment in the current authoritative, more restrictive climate of university and departmental institutional review boards (IRBs). Recent editions of the American Psychological Association's (2002) ethical guidelines have stipulated, unequivocally, that a participant's wish to withdraw from participation in research must be respected and that this freedom must be clearly communicated prior to the study. In the Milgram paradigm, the experimenter explicitly challenges the participant's right to withdraw. For myself, this challenge has always been the crux of the ethical controversy, more than the stress and emotional displays, or the many deceptions involved. Two other factors may also have been key to ethical developments. One

Correspondence concerning this article should be addressed to Arthur G. Miller, Department of Psychology, Miami University, Oxford, OH 45056. E-mail: millerag@muohio.edu

© 2009 American Psychological Association 0003-066X/09/$12.00
Vol. 64, No. 1, 20–27 DOI: 10.1037/a0014407



Arthur G. Miller

is the infliction of (apparently) increasingly severe pain on an unwilling recipient, a feature seemingly unique to the Milgram studies. The second is the wide distribution of Milgram's film of the obedience research, which has vividly portrayed the emotionally intense impact of his procedure to generations of students.

In the extensive commentary over the past 45 years regarding the ethics of Milgram's paradigm, there is a consensus that Milgram either approached or exceeded the boundaries of ethical propriety in terms of the coercion and stress experienced by many of his participants (Blass, 2004; Miller, 1986, chap. 5). Among the more instructive commentaries on the ethics of the obedience studies, a particularly memorable quotation is that of M. Brewster Smith (1976):

> For myself, I find it quite possible to justify a Milgram study, as a carefully weighed exception, an important study that raises serious ethical questions but can be undertaken with heavy ethical responsibility falling on the investigator's shoulders. I wouldn't do such a study, but I really respect Milgram's right to have done it. (p. 452)

I have no idea of how many investigators have tried to replicate the Milgram paradigm, only to be rejected by their university or departmental IRB. The best guess is that very few if any have done so for many years. So, with Burger's (2009) new study in hand, what do we find that is informative on the problems of studying obedience to authority?

Why Is a Replication of the Milgram Experiment Important?

The most important characteristic of any new study or conceptual analysis is the degree to which it extends the line of inquiry, for example, by testing theoretical ideas, by suggesting new methods, or by posing further questions for research (or stimulating others to do so). Burger (2009) voiced this idea in the last sentence of his article: "I hope future investigators will utilize the 150-volt procedure presented here to address the weighty questions that motivated Stanley Milgram nearly half a century ago" (p. 10). I agree, at least in principle. The major contribution of Burger's study is that it could open a window for researchers to address vital questions that have hovered over the original Milgram experiments. In terms of theoretical understanding, Lee Ross (1988) suggested that we really have no precise explanation for why participants behaved as they did in Milgram's experiments. Plausible ideas here are rampant, but somewhat vague and always numerous. As Burger suggested, obedience could simply be a reasonable reaction to a seemingly legitimate source of influence in a highly ambiguous setting. It could also reflect what Milgram (1974) termed the *agentic shift*, the absolving of a sense of personal responsibility. It could be a fear of defying authority, a hesitancy to "make a scene" and embarrass oneself or ruin the project. Simply by volunteering to be in the study, participants could have been strongly motivated to do whatever the person in charge said. Milgram (1974, Experiment 12) observed, for example, that participants obeyed orders to "stop" shocking the victims as well as orders to shock them. It wasn't simply a matter of taking situational advantage to exercise hostility or aggression. It could reflect the power and value of Science—that is, in the eyes of research participants—to override any other consideration, once participants were psychologically locked into the basic situation. Only further research can provide the convincing means to sharpen theoretical understanding of destructive obedience.

Other questions about Milgram's experiments have lingered for years. Many of these relate to generalizations from the laboratory to real-world scenarios of harm doing and organized evil (Darley, 1992; Miller, 1995, 2004). For example, how would obedience to malevolent authority be influenced if the victim (i.e., learner) was a member of a discriminated or stigmatized group? Would a person who obeyed completely be likely to engage in even further actions against another person, at a later time, in another setting? Do self-perceptions change in predictable ways among those who defy as well as obey authority, and are these perceptions, in turn, predictive of further destructive or prosocial actions? Do individuals rationalize their destructive obedience and see it as desirable and useful behavior? (For an imaginative empirical approach to the escalation issue, but not specifically in the context of obedience, see Martens, Kosloff, Greenberg, Landau, & Schmader, 2007). How do participants in the subordinate role partition responsibility for their actions? What behavioral effects would occur if the "teacher" had an opportunity to reflect upon, and to discuss with others, the nature of the upcoming shock/punishment task instead of being thrown into the escalating situation alone and virtually without warning? What sorts of postexperimental accounts would defiant as well as obedient participants give for their behavior? Rationalizations would be a very predictable

result here (Tavris & Aronson, 2007), but guilt, shame, anger, and other emotions are certainly possible and are of great interest. Would "I was only following orders" be a prominent defense? What about individual differences? Which dispositional constructs are relevant to specific variations of the obedience situation? Kelman and Hamilton (1989) provided an informative discussion of this issue, but the actual linkage of their theoretical constructs to a study of behavioral obedience has not been made. As for gender differences, one would need a complete design covarying the sex of authority, teacher, and learner to clarify this matter. Many of these questions have been raised in numerous discussions of the obedience research over the past decades. None has been fully addressed empirically in the actual context of the Milgram paradigm. Burger's (2009) basic approach to replicating the Milgram paradigm provides, in my view, a method to answer all of these questions and undoubtedly many others.

A Note on Torture

Another value of Burger's (2009) approach is that it reminds us of the importance of torture in the modern world. Milgram's (1963, 1974) original studies were, among other things, an exploration of people's willingness to physically harm others. The actual origin of the "aggression machine" is a matter of mild historical dispute, and many studies have used (deceptively) a shock-delivering apparatus of some kind (e.g., Leonard Berkowitz's, 1993, well-known research on aggression). However, Milgram was perhaps the first researcher to use a procedure in which a person was ordered by an authority to inflict increasingly severe physical pain upon another person. In the 1980s, Meeus and Raaijmakers (1986, 1995) conducted an important conceptual replication of Milgram's paradigm, including several key situational variations that Milgram had originally devised. Their objective was specifically to increase the ecological validity of Milgram's paradigm by using what they termed *mediated violence*—that is, ordering a naive participant to inflict stress-inducing verbal comments upon another person taking a test for a job interview instead of using (apparent) electric shock as the means of harming. These investigators considered using electric shock as "exerting a form of archaic physical violence" (Meeus & Raaijmakers, 1995, p. 159). In other words, it would be more representative of real-world destructive obedience to use a procedure involving psychological distress rather than physical pain. In the current world scene, however, the rise of terrorism and the use of torture have become very prominent, highly controversial, sociopolitical issues (Zimbardo, 2007, chap. 14). Perhaps more than ever, we need to understand the conditions and causes for people's willingness to administer physical pain and inflict punishment (even murder) upon others, particularly under orders from a superior in an organizational hierarchy (Darley, 1992; Moghaddam, 2005).

Major Contributions of This Study

The major lessons of Burger's (2009) study seem to be these: (a) It is possible to conduct a partial replication of the Milgram paradigm under the current guidelines of a university IRB, and (b) given the procedure described in Burger's research, a rate of obedience comparable to that reported by Milgram is observed. Achievement of this outcome required two major changes in the procedure described by Milgram, both designed to minimize the likelihood of unacceptable levels of stress, tension, or harmful aftereffects in participants. The first was implementing a multifaceted screening of participants. The second was reducing the maximum allowable shock level from 450 to 150 volts. Retaining essentially all other aspects of Milgram's methodology, Burger observed an obedience rate (70%) that is comparable to that of Milgram's (82.5%), that is, in terms of those participants willing to exceed the 150-volt level.

I personally find the 150-volt solution to be quite convincing. If one scans Milgram's (1974) data tables for each of the experimental variations (e.g., Tables 2, 3, and 4), the 150-volt level constitutes a unique threshold. More people break off (disobey) immediately after pressing that switch than any other, and if they exceed that level, the odds are high that they continue to the end. Packer (2008) has confirmed this picture with a recent meta-analysis on data from eight of Milgram's (1974) experimental variations. He focused on a particular methodological feature of Milgram's paradigm—that is, that starting *at the 150-volt level*, there is a clearly audible, verbal request (tape recorded) on the part of the learner to be released. Packer hypothesized that the 150-volt level represents a pivotal choice point among the 30 incrementally increasing shock levels, and his results provide strong support for Burger's (2009) position:

> In all studies, disobedience was most likely at 150 v, the point at which the shocked "learner" first requested to be released. Further illustrating the importance of the 150 v point, obedience rates across studies covaried with rates of disobedience at 150 v, but not at any other point; as obedience decreased, disobedience at 150 v increased. In contrast, disobedience was not associated with the learner's escalating expressions of pain. This analysis identifies a critical decision point in the obedience paradigm and suggests that disobedient participants perceived the learner's right to terminate the experiment as overriding the experimenter's orders. . . . Among noncompliant participants, a much higher proportion disobeyed at 150 v (36.88%) than at any other point; the next highest proportion, 10.63%, disobeyed at 315 v. (Packer, 2008, pp. 301–302)

Packer's (2008) analysis is thus consistent with Burger's (2009) contention that the 10th (150-volt) shock level is a kind of alternative "end point" for the study. That is, one knows more about whether or not participants are going to continue obeying orders to press shock levers at the 10th (150-volt) level than at any other single point in the study. Packer's study also provides a valuable reminder of the considerable variability of obedience observed in Milgram's (1963, 1974) original data. Individual differences within many of the experimental variations are one of the most striking features of his results, but they are an exceedingly underreported and ignored feature as well.

In a previous analysis (Miller, 1990), I examined the results of 16 of the experimental variations in Milgram's (1974) program that involved explicit orders to inflict shock on a protesting victim. Out of a total of 540 participants, 324 (60%) *disobeyed* orders at some point in the sequence of commands. In 11 of the 16 studies, more than 50% of the participants defied authority at some point prior to 450 volts. The pervasive, almost reflexive, generalizations made from the Milgram experiments to depictions of the inherent evil in human nature—for example, in explaining the Nazi Holocaust—rarely take into account these individual differences or the powerful effects of subtle changes in the situation as well (Miller, 2004; Miller, Buddie, & Kretschmar, 2002). In this context, Burger's (2009) new procedure could encourage important research on the issues of why individuals differ so markedly in their response to authority as well as the manner in which individual differences interact with situational variation.

Burger's (2009) reasoning, that behavior at the 150-volt level is predictive of what participants would likely do were they to be ordered past that level, is, in my opinion, both imaginative and convincing. Milgram (1974, p. 2) often emphasized the importance of distinguishing between philosophical discussions about obedience and actual behavioral analysis. Armed with Burger's new investigation, we have an opportunity to consider fresh data, to rethink things in the context of both the legacy of Milgram and contemporary perspectives. What does this picture look like?

The Screening of Participants

I have several thoughts on the screening issue. First, there is the matter of one's personal values and beliefs regarding the legitimacy of studying, in a rigorous manner, human behaviors involving a tolerable but clearly unpleasant degree of stress and conflict. This commentary is not the place for an extended discussion on this complex problem. However, three conclusions seem clear to me: (a) Researchers and scholars will always differ in their opinions on this issue; (b) the tide of opinion has shifted in the direction of erring on the side of caution (i.e., it is better to engage in no study than to propose one that appears to entail significant risk regarding participant discomfort); and (c) there is the pragmatic necessity of gaining IRB approval for such a study. I presume that the latter issue was a significant motivation behind Burger's (2009) employing the diverse techniques described in his section on "recruitment and screening" (pp. 5–6; however, Burger did not provide specific information regarding his communications with the IRB, which would have been useful to readers). Thus, if one wishes to retain significant elements of Milgram's procedure and still pass the threshold of an ethical research procedure in the light (or darkness) of current IRB guidelines, one needs to "keep what you can and change what you must."

Burger's (2009) screening procedure has both methodological and ethical dimensions. One requirement was to find participants unfamiliar with the Milgram studies in order to avoid suspicion of the deceptions involved, the possible meaning of their behavior, and so forth. Another goal was to exclude participants who would be likely to react in an unacceptably emotional way to the experiment. This objective amounted to an extensive clinical assessment involving a series of stress-relevant questions as well as personal interviews with each prospective participant. Several points can be made here. First, in principle, Burger's approach seems quite sensible. Obtaining participants who are naive to the Milgram studies is certainly necessary from a methodological perspective. Attempting to reduce emotionally devastating effects or symptoms in participants both during and after their participation is also reasonable in light of the ethical controversy and the IRB guidelines currently operative.

However, despite the sensible and reasonable arguments Burger (2009) made for his methodology, his approach does not strike me as practical or perhaps even doable. For example, how viable would it be to conduct an extended research program using this paradigm in a typical undergraduate population or university setting? The requirements for immediate debriefings as well as large numbers of naive participants (even if from nonuniversity sources) for subsequent studies would pose significant challenges. The relatively extreme measures taken to screen participants for their ability (or lack thereof) to withstand the stress of the procedure also present a number of difficulties or ambiguities. Some of these problems are related to what Burger described as "the 150-volt solution" (p. 2). If, in stopping participants at the 150-volt level, one would avoid "exposing them to the intense stress Milgram's participants often experienced in the subsequent parts of the procedure" (Burger, 2009, p. 2), why would it then also be necessary to engage in the extended screening itself? Using the 450-volt procedure, Milgram (1974) did not report any harmful psychic aftereffects in his participants, nor, to my knowledge, has anyone else.

Thus, one gets the impression of ethical overkill here. By reducing the effective endpoint of the shock generator as well as screening out a large proportion (well over 50%) of the initial sample of volunteers, there is a convincing sense that the prospects of undue discomfort in the experiment have been significantly minimized—but at an enormous cost in terms of efficiency (time, expense, personnel) and precision. Without additional evidence that would be very difficult to obtain, there is no absolute certainty that any of the measures used here were in fact necessary or effective. From an ethical point of view, the 150-volt solution might, in itself, have been more than satisfactory, that is, without the additional screening. Conversely, the screening might have made the 150-volt endpoint unnecessary. (I would personally opt for the 150-volt endpoint and far less screening.)

Although there is certainly an impression of ethical safety in Burger's (2009) procedure, it is largely based on intuition, self-report, and uncertain clinical acumen. On the plus side, one could say that Burger simply did the best he could do to guarantee (to the IRB as well as others) a significantly more ethical obedience study. Of concern here is the image of the research operations, what one might

term the *impression management* of the experiment. What impressions are conveyed in describing the procedure to, say, an IRB or funding agency? Is this project perceived as safe for every participant? These are not inconsequential matters for the experienced researcher. On the minus side, Burger's procedures may in fact have been either unnecessarily or insufficiently restrictive. What actually were his participants' emotional reactions during or after the study? Unlike Milgram, who paid extraordinary attention to these matters, Burger did not provide any information. One presumes that all went well here, but one cannot be certain. There is also a sense in which the *lack* of emotional distress in participants is problematic.

Avoiding Stressed Participants—At What Cost?

In considering the meticulous attention to the emotional welfare of participants in this study, the reader is, at first, likely to be impressed. After all, this is presumably the only approach that would be sanctioned by an ethical board of review. There is also the impression that Burger (2009) has been ingenious in conceiving a methodology that accomplishes the seemingly impossible, that is, a meaningful inquiry into destructive obedience to authority that is, at the same time, ethical. But has this really been achieved? Indeed, can it be achieved? If so, are there serious costs? As always, it depends on the details and how one construes them.

It is commonly assumed that the majority of Milgram's participants were under various degrees of duress, many at genuinely extreme levels. Milgram created a scenario involving moral conflict. Participants were presented with competing sources of influence—an authority figure's insistent orders to continue harming a victim pleading to be released. In his original publication, Milgram (1963) noted two major findings: One was the unexpectedly high rate of obedience (26 of 40 proceeding to 450 volts).

> The second unanticipated effect was the extraordinary tension generated by the procedures. One might suppose that a subject would simply break off or continue as his conscience dictated. Yet, this is very far from what happened. There were striking reactions of tension and emotional strain. (Milgram, 1963, p. 377)

These emotional reactions were not viewed by Milgram (1974) as merely an unanticipated side effect of the procedure (although they were unexpected at the very initial stages of the obedience studies) but were regarded as a major substantive finding. Powerfully, they revealed that participants would, in certain situations, exhibit very high rates of obedience while simultaneously experiencing extreme discomfort and personal misgivings regarding their own behavior.

According to this line of reasoning, to the degree that Burger's (2009) new study failed to induce these emotional reactions, one can question its parallel to Milgram's original investigation. Given the 150-volt ceiling, Burger has shown comparability in terms of behavioral obedience (with respect, specifically, to Milgram's [1974] "new baseline study," i.e., Experiment 5). (In Experiment 1, described in his first publication [Milgram, 1963], no participant broke off prior to the 300-volt level.) But Burger did not observe the second of Milgram's major findings, that is, the powerful indicators of emotional conflict and the crucial fact that the presence of these strongly negative emotional reactions did not predict disobedience. As Milgram (1974) noted, "Many people were unable to realize their values in action and found themselves continuing in the experiment even though they disagreed with what they were doing" (p. 6).

It seems likely that *if* Burger's (2009) participants had, in fact, been taken past the 150-volt level, the indications of stress and emotion reported by Milgram would have appeared. But one cannot be certain. Here, Burger's use of intense screening comes into play. Because of that screening, participants who would have shown more emotion or stress could have been precluded from being in the study. So, the "half-full, half-empty glass" metaphor seems relevant. Yes, we have a procedure that seems to fulfill some aspects of "replicating Milgram" but at a rather serious cost. Because of the need to modify the paradigm for IRB approval (among other reasons), it is "good" that intense (if any?) displays of stress and emotion were not observed in this study. The project presumably could not have been conducted if these reactions were elicited. But because those features are so fundamental to Milgram's findings, and to the inferences that have been drawn from them (e.g., Miller, 2004; Waller, 2002), it is difficult to argue too strongly for comparability, that is, for replication.

The Value of Noncomparability

There could be a distinct value in studying behavioral obedience even if the powerful emotional component is not present. It might be the case that the earlier stages of the shock series—Burger's (2009) focus—are more homogeneous or "pure" in terms of reflecting a less conflicted decision on the part of participants to obey the experimenter. At the lower shock levels, the motivations of the participants may consist essentially of "volitional obedience" rather than the recalcitrant or dissonance-laden obedience that seems so characteristic of responses at the higher shock levels. The decisions of participants at the earlier shock stages may reflect different processes than decisions at higher levels of shock. Hypothesized changes in self-perception, or self-serving rationalizations accompanying the escalation of shock levels, may not (yet) have taken place among the majority of participants. As a result, this situation may provide an opportunity to test the influence of other causal factors as well as to examine participants' verbalized accounts of their behavior without the confounding presence of varying degrees of high tension, physical symptoms of stress, and so forth.

The Failure of the Disobedient Model Condition

Milgram viewed the susceptibility of obedience to situational influences as one of the most significant and unheralded findings of his research program:

A second, and no less important aspect of the research concerns the systematic changes introduced in each of the 18 experimental conditions. The degree of obedience varied sharply depending on the exact manner in which the variables of the experiment are arranged in an experimental condition . . . yet, in the popular press, these variations are virtually ignored, or assumed to be of only minor importance. (Milgram, 1979, pp. 7–8)

The "modeled refusal" condition was the most logical variation to include in Burger's (2009) study. Numerous studies on conformity and obedience have demonstrated the powerful effects of witnessing defiance to social pressure (Asch, 1956; Meeus & Raaijmakers, 1995; Milgram, 1974).

The failure to find significant effects of the refusing model on participants' behavior in Burger's (2009) study is puzzling. The most obvious possibility is that for various reasons, the manipulation was not sufficiently salient or forceful. Burger concurred: "If participants saw 10 out of 10 other individuals refuse to continue the procedures, the descriptive norm would be clear" (p. 10). (Fewer than 10 out of 10 would, I think, be sufficient.) The presence of unpredicted but considerable obedience in this condition does at least attest to the reliability of the paradigm itself. As Burger indicated, "I interpret this high rate of obedience in the modeled refusal condition as a demonstration of the power of the situational forces leading participants to go along with the experimenter's instructions" (p. 10). Of course, had this condition influenced behavior in the predicted direction, the same explanation or (at least) description—that is, the power of situational forces—would have been equally applicable.

The "Torn" Participant

There are other possibilities to consider. One is the likelihood that Burger's (2009) paradigm failed to generate moral conflict, at least in sufficient strength to motivate participants to "find a way out." Notice that in addition to predicting significantly less obedience in the modeled-refusal condition, Burger implied that strong emotions would be experienced. In describing the rationale for predicting the effects of witnessing a defiant model, Burger (2009) stated,

Because participants are *torn* between doing what the experimenter tells them and not wanting to hurt the learner, I reasoned that they might eagerly rely on this limited norm information to conclude that refusing the experimenter's instructions is appropriate. (p. 4, italics added)

This idea of "the torn participant" is also the rationale that Milgram (1974) used to predict (and ultimately observe) relatively low rates of obedience in a variety of experimental variations. That is, because participants are fundamentally (and increasingly, over trials) opposed to the orders they are receiving, they are primed to use any plausible rationale or available channel (Ross & Nisbett, 1991) to defy the experimenter's commands. Numerous variations in fact resulted in far less obedience, for example, the "touch proximity" or "experimenter absent" variations (Milgram, 1974).

Burger's (2009) use of the term *torn* is thus ambiguous. Taken literally, *torn* suggests intense conflict and could imply that Burger in fact expected a considerable amount of tension and stress, contrary to the stated aim of his modified paradigm. However, perhaps a person could be torn and still not enter that zone of ethically unacceptable agony that appeared to characterize many of Milgram's respondents.

Are Meaningful Comparisons Across Paradigms Even Possible?

Other factors add to the ambiguity regarding the modeled refusal condition. In comparison to Milgram's (1974) "Two Peers Rebel" variation (pp. 116–121), Burger (2009) made at least three intentional and possibly consequential changes: (a) He introduced the defiant model before the participant had made any responses whatsoever; (b) he used only one instead of two defiant models; and (c) he used a procedure in which the modeled refusal was "less dramatic" (p. 4) than that used by Milgram. From this perspective, there is really no way to compare Burger's procedure to that of Milgram because (a) Burger duplicated some aspects of Milgram's procedure to the letter but (b) also made vital changes in other aspects at the same time. Burger noted, for example, that Milgram introduced the first of two defiant peers at the 150-volt level, that is, at a point in the shock series in which 33% of the levers had already been pressed. Burger wanted to make it easier for participants to disobey earlier; hence he presented the defiant model prior to any behavior on the part of the naive participant. Yet even with this rather substantial procedural change, there were no significant effects. One major lesson of Burger's study, therefore, is that there are serious interpretive difficulties that are inherent in a procedure that retains certain features of a classic experimental paradigm but changes other features in an effort to bring the research under the threshold for ethical scrutiny. It is hard to reconcile these considerations with Burger's (2009) conclusion that "in short, I am as confident as a psychology researcher can ever be that my findings can be legitimately compared with Milgram's" (p. 10).

The Four Prods

In Milgram's script for the experimenter's role, the 4 increasingly strident and demanding prods have always struck me as one of the most important features of the obedience paradigm—as the most explicit operationalization of authority. The fourth prod reads as follows: "You have no other choice, you *must* go on" (Milgram, 1974, p. 21, italics in the original). Taken literally and out of context, these words are clearly preposterous. They violate any reasonable sense of ethics. In context, of course, they appear to have had considerable persuasive force.

Burger (2009) indicated (p. 7) that he followed Milgram's four sequential prods to the letter. This faithfulness to Milgram's procedure is surprising because the prods—the third and fourth in particular—could well be regarded as the single most glaringly unethical feature of Milgram's

paradigm. Nevertheless, Burger also indicated, on p. 9: "The participants were told explicitly and repeatedly that they could leave the study at any time and still keep their $50." There is an apparent inconsistency between the fourth prod and these repeated reassurances to the participants of their right to withdraw from the experiment. Which of these seemingly contradictory pieces of information were participants expected to heed? How did the university IRB appraise this particular issue?

Conclusions

It is relatively easy to sit back in one's armchair and, basking in the soothing illumination of hindsight, reflect, at times critically, on someone else's research efforts. It is a far more challenging task to do the research itself. I applaud Jerry Burger for his considerable efforts and for giving psychologists the opportunity to reconsider the Milgram research and the possibilities for its continuation. He has produced a bold and imaginative technique to rekindle research on a famous problem entailing extraordinary methodological and ethical difficulties. Gratitude here must, of course, be tempered with judgmental acumen. Evidence suggests that when one is evaluating a study on a particularly important or consequential topic, there are judgmental biases that may operate to blunt or soften criticism (e.g., Wilson, DePaulo, Mook, & Klaren, 1993). As a commentator, I have tried not to overlook some of the limitations of this study as well as its strengths. Both of these characteristics can be instructive in terms of future developments.

As noted, I find Burger's (2009) 150-volt solution to be a promising approach. Keeping the original, 30-lever, 450-volt shock generator in full display for participants while using effectively only 33% of the total range of those levers is an ingenious methodology. This kind of creative inquiry is in the best tradition of science. A host of empirical questions could be profitably studied using this technique. The ethical ramifications are still, of course, important and complex. The ethical safeguards described by Burger seem unusually taxing to me. I have noted the possibility that the intense screening of participants would not be required if the 150-volt ceiling on punishment effectively prevented unacceptably high levels of stress. In this context, it is conceivable that Burger's study could now be cited—for example, for IRB review committees—as empirical evidence for the ethical propriety of his modified procedure, and thus future investigators might not be obliged to replicate his intensive screening.

Assuming that Milgram's original procedure is now clearly out of bounds, something "has to give" for essential aspects of this paradigm to remain viable. Whether Burger's (2009) particular strategy is ultimately the wisest or most productive course remains to be seen. I would add, however, that the idea of a direct comparison between Milgram's original findings and those of Burger (or future studies) is ill advised and unnecessary. It was tempting and perhaps logical for Burger to utilize this "side by side" data comparison in his initial study. In a sense, of course, Milgram's findings would always be available for comparisons, whether Burger emphasized them or not. But as I have suggested, there are simply too many differences between this study and the earlier obedience research to permit conceptually precise and useful comparisons.

Because of the seemingly boundless interest in Milgram's particular approach to investigating destructive obedience, it is understandable that there would also be strong interest in the viability of a similar, but less ethically problematic, methodological strategy, that is, conveying the impression that one is inflicting physical pain on another person under experimenter instructions. However, it should be emphasized that there are, in principle, a variety of different methodological approaches that one could take to investigate destructive obedience (Brief, Dietz, Cohen, Pugh, & Vaslow, 2000; Darley, 1995, 1999; Meeus & Raaijmakers, 1995). Destructive obedience takes many forms. For example, submitting to the dictates of authority and inhibiting the expression of one's personal disagreement could occur in the context of group decision making—for example, in an organizational policy meeting devoted to planning genocide or other, more mundane but still immoral decisions (e.g., Darley, 2001). Here, the response of *silence* on the part of subordinates could, in principle, be an extremely consequential form of destructive obedience to authority yet one very different in form and context from the pressing of shock levers in Milgram's approach or Burger's (2009) modified paradigm.

Stanley Milgram once told me that a major source of his regret was that the obedience research had stimulated so much ethical and methodological controversy rather than substantive research on obedience itself. He would thus, I think, be very enthusiastic regarding Burger's (2009) approach but probably dismayed at what he would regard as overly restrictive ethical guidelines. In terms of these guidelines and IRB approval, things are very different for social psychological research today than in decades past. Sacrificing the welfare of research participants in the pursuit of vital knowledge on unpleasant behaviors is, of course, controversial (at best) and possibly very dangerous. For good reasons, researchers cannot be put in charge of evaluating the ethics of their own procedures. On the other hand, knowledge about the destructive capacities of human beings always remains sorely needed. In the context of these values and concerns, Burger's new study represents, I think, an impressive effort to fill a very serious void. He has provided a pointed reminder that even (indeed especially) for the Milgram study (and other studies as well), psychologists must not simply settle for "Do No Research" as another ethical guideline.

REFERENCES

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57,* 1060–1073.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs, 70*(9, Whole No. 416).

Berkowitz, L. (1993). *Aggression: Its causes, consequences, and control.* New York: McGraw-Hill.

Blass, T. (2000). The Milgram paradigm after 35 years: Some things we

now know about obedience to authority. In T. Blass (Ed.), *Obedience to authority: Current perspectives on the Milgram paradigm* (pp. 35–59). Mahwah, NJ: Erlbaum.

Blass, T. (2004). *The man who shocked the world: The life and legacy of Stanley Milgram.* New York: Basic Books.

Brief, A. P., Buttram, R. T., Elliott, J. D., Reizenstein, R. M., & McCline, R. L. (1995). Releasing the beast: A study of compliance with orders to use race as a selection criterion. *Journal of Social Issues, 51,* 177–194.

Brief, A. P., Dietz, J., Cohen, R. R., Pugh, S. D., & Vaslow, J. B. (2000). Just doing business: Modern racism and obedience to authority as explanations for employment discrimination. *Organizational Behavior and Human Decision Processes, 81,* 72–97.

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist, 64,* 1–11.

Darley, J. M. (1992). Social organization for the production of evil. *Psychological Inquiry, 3,* 199–218.

Darley, J. M. (1995). Constructive and destructive obedience: A taxonomy of principal–agent relationships. *Journal of Social Issues, 51,* 125–154.

Darley, J. M. (1999). Epilogue: Methods for the study of evil-doing actions. *Personality and Social Psychology Review, 3,* 269–275.

Darley, J. M. (2001). The dynamics of authority influences in organizations and the unintended action consequences. In J. M. Darley, D. M. Messick, & T. R. Tyler (Eds.), *Social influences on ethical behavior in organizations* (pp. 37–52). Mahwah, NJ: Erlbaum.

Hodges, B. H., & Geyer, A. L. (2006). A nonconformist account of the Asch experiments: Values, pragmatics, and moral dilemmas. *Personality and Social Psychology Review, 10,* 2–19.

Kelman, H. C., & Hamilton, V. L. (1989). *Crimes of obedience: Toward a social psychology of authority and responsibility.* New Haven, CT: Yale University Press.

Martens, A., Kosloff, S., Greenberg, J., Landau, M. J., & Schmader, T. (2007). Killing begets killing: Evidence from a bug-killing paradigm that initial killing fuels subsequent killing. *Personality and Social Psychology Bulletin, 33,* 1251–1264.

Meeus, W. H. J., & Raaijmakers, Q. A. W. (1986). Administrative obedience: Carrying out orders to use psychological-administrative violence. *European Journal of Social Psychology, 16,* 311–324.

Meeus, W. H. J., & Raaijmakers, Q. A. W. (1995). Obedience in modern society: The Utrecht studies. *Journal of Social Issues, 51,* 155–175.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67,* 371–378.

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York: Harper & Row.

Milgram, S. (1979). Preface. *Soumission à l'autorité: Un point de vue experimental* [Obedience to authority: An experimental view]. Paris: Calmann-Lévy.

Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science.* New York: Praeger.

Miller, A. G. (1990). A perspective on Kelman and Hamilton's *Crimes of obedience. Political Psychology, 11,* 189–201.

Miller, A. G. (1995). Constructions of the obedience experiments: A focus upon domains of relevance. *Journal of Social Issues, 51,* 33–54.

Miller, A. G. (2004). What can the Milgram obedience experiments tell us about the Holocaust? Generalizing from the social psychology laboratory. In A. G. Miller (Ed.), *The social psychology of good and evil* (pp. 193–239). New York: Guilford Press.

Miller, A. G., Buddie, A. M., & Kretschmar, J. (2002). Explaining the Holocaust: Does social psychology exonerate the perpetrators? In L. S. Newman & R. Erber (Eds.), *Understanding genocide: The social psychology of the Holocaust* (pp. 301–324). New York: Oxford University Press.

Moghaddam, F. M. (2005). The staircase to terrorism: A psychological exploration. *American Psychologist, 60,* 161–169.

Packer, D. J. (2008). Identifying systematic disobedience in Milgram's obedience experiments. *Perspectives on Psychological Science, 3,* 301–304.

Ross, L. D. (1988). Situationist perspectives on the obedience experiments [Review of the book *The obedience experiments: A case study of controversy in social science*]. *Contemporary Psychology, 33,* 101–104.

Ross, L. D., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* New York: McGraw-Hill.

Smith, M. B. (1976). Some perspectives on ethical/political issues in social science research. *Personality and Social Psychology Bulletin, 2,* 445–453.

Tavris, C., & Aronson, E. (2007). *Mistakes were made (but not by me): Why we justify foolish beliefs, bad decisions, and hurtful acts.* New York: Harcourt.

Waller, J. (2002). *Becoming evil: How ordinary people commit genocide and mass killing.* New York: Oxford University Press.

Wilson, T. D., DePaulo, B. M., Mook, D. G., & Klaren, K. J. (1993). Scientists' evaluation of research: The biasing effects of the importance of the topic. *Psychological Science, 4,* 322–325.

Zimbardo, P. (2007). *The Lucifer effect: Understanding how good people turn to evil.* New York: Random House.