Thinking Critically About Psychology's Classic Studies

Revisiting Studies by Milgram, Harlow, Mischel, Sherif and Others and What They Mean Today

BY CAROL TAVRIS

LAST YEAR MARKED THE 50TH anniversary OF Stanley Milgram's experiments on obedience to authority, an event that inspired a conference, many reflective papers, and a popular book of vitriolic criticism. The occasion got me thinking again about the eternal dilemma for us psychology instructors and textbook writers: How much time should we devote to teaching the classics v. making way for new (and often yet-unreplicated) research, and how should we teach them? In every generation, certain studies get planted in our books and lectures, and they tend to become rooted there. Over time it gets harder to decide how much to prune—let alone decide if it's time to uproot them. We stop looking at the original studies closely, let alone critically; they just sit there in our courses like grand historical monuments.

However, it's good to reexamine them for two reasons: One is for our own sake, to refresh our memories and rethink their contributions; the other is for our students' sake. Students today are as eager to reject unflattering or counterintuitive portrayals of humanity as students were decades ago. Teaching the classics therefore means finding new ways of persuading students that these findings *do* apply to them, despite the errors or limitations of the original studies.

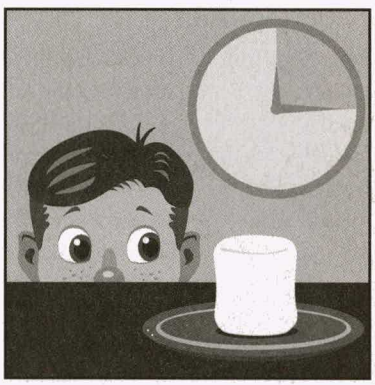
The relationship between cultural events and research is a two-way street: an event may stimulate research, and research may influence the larger culture. Many instructors still tell the story of Kitty Genovese, murdered in 1964, to show how it launched a long and productive line of experimental studies on bystander apathy, deindividuation, diffusion of responsibility, and intervention. Thanks to two recent books and a critical reassessment that appeared in the journal *American Psychologist* in 2007, we now know that almost all of the details that were reported at the time were wrong. Turning Genovese's death into a story of urban alienation was largely the work of A.M. Rosenthal at the *New York Times*, who wanted the image of the "38 witnesses doing nothing" to be in the story's headline, where it quickly went the media equivalent of "viral" in 1964. In fact, most of the neighbors who heard her screams could not see her—let alone watch her murder from their windows—and thought it was just another drunken domestic fight. As it turns out, only three neighbors understood the attack for what it was and failed to respond. Three is too many; but a newspaper story about three craven witnesses would not have been as shocking.

Should teachers eliminate the Kitty Genovese story out of embarrassment that we got it wrong or weren't skeptical enough? Not necessarily. Although the specifics of the story were wrong, its essence

This is an edited version of a lecture given at the Association for Psychological Science's annual convention, May 24, 2014. It appeared in the Times Literary Supplement, *July 18, 2014.*

was true; then, as now, there are far too many real stories of people in trouble being ignored by passersby who could have helped them—only now we can see them on YouTube. The accurate Genovese story, however, raises an additional social psychological lesson for students today: What cautions does it offer for thinking critically about an equally sensational crime today? What does it reveal about how societal sources of anxiety can generate an urban legend? America at the time was undergoing political assassinations, race riots, the Vietnam War, and rising crime rates—Kitty Genovese was one of 636 murders in New York that year. People were frightened. That headline resonated.

With Kitty Genovese, an oversimplified but emotionally compelling narrative led to good research. It often works the other way, of course: good research can lead to an oversimplified and emotionally compelling narrative. Consider the case of Walter Mischel's (dare I say) delicious marshmallow study of four-year-old children's ability to delay gratification. The part that got so much recent public attention was that children who had resisted temptation turned out years later to score as much as 210 points higher on their SATs than the most impatient children. Bingo!



The marshmallow study captured the public imagination just as Kitty Genovese had, but with a brighter, happier moral—one that suits our current cultural concerns. The marshmallow story is funny; we can all see those kids in our mind's eye—and try to imagine what we would have done, faced with a now-or-later choice. And it has such an all-American, Calvinist moral: delay gratification and heaven will be thine. I wasn't thinking very critically about this study until I read Matthew Bourne's analysis in the *New York Times* of how and why this story resonates in popular culture—as usual, by pruning away pesky details about the actual scientific findings. Mischel's original studies focused on 653 toddlers, all of them attending the Bing Nursery School at Stanford University, for children of professors and graduate students. The studies weren't originally designed to look at long-term outcomes; that idea occurred to Mischel much later, when he asked his own children, who had attended the Bing school, how his research subjects were faring in college. Of the original 653, he tracked down 185, of whom 94 provided their SAT scores.

So we have a problem with the representativeness of the sample, originally and in the follow-up. How would that affect the results? A 2012 article in *Cognition* by Celeste Kidd, Holly Palmeri, and Richard N. Aslin showed that some children were more likely to eat the first marshmallow when, by virtue of their previous history, they had reason to doubt the researcher's promise to come back with a second one. For children raised in an unstable environment, they wrote, "the only guaranteed treats are the ones you have already swallowed," while children raised in a more stable environment might be willing to wait a few more minutes, confident that the second treat will materialize. And not only the stability of the environment mattered: I'm an only child, so the effects of siblings didn't occur to me until a friend said, "try the marshmallow study on those of us who have three siblings. You grabbed what you could, or else one of them got it."

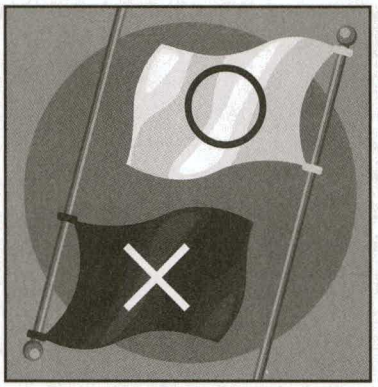
To offer these caveats about the way the Mischel findings are so often told is not to debunk the excellent original study but to make students more excited about it—to alert them not only to what it found, but also what it omitted. The study and its popular response illustrate how in science each finding generates new questions; the importance of critical thinking in coming up with alternative hypotheses; and the importance of not oversimplifying. An ability to resist temptation is one factor among many that shapes our lives, but if you are a kid from an unstable family, living in a tough environment, in poor health, delaying gratification may not be the best strategy at the time. It might not be a stable character trait. It might not even cross domains of experience.

Some of social psychology's great historical monuments—studies with a long influence—could not be replicated today: Muzafer and Carolyn Sherif's Robbers cave study, Stanley Milgram's obedience experiments, and Harry Harlow's wire and cloth mother studies.

Between 1949 and 1954, the Sherifs and their colleagues used a Boy Scout camp in Oklahoma to test their theories of the origin and reduction of intergroup animosity and prejudice. Although the camp had been specially arranged and was under the Sherifs' control, and the boys were randomly assigned to the two groups of Eagles and Rattlers, I had misremembered there being more data than there were.

Illustrations in this article by Ástor Alexander

Most of the conclusions, Sherif wrote, "were reached on the basis of observational data"—confirmed by "sociometric choices and stereotype ratings." He said: "Observations made after several superordinate goals were introduced showed a sharp decrease in name-calling and derogation of the out-group common…in the contact situations without superordinate goals." (By the way, one pleasure of going back to read original studies is unexpected discoveries: The "name-calling" is so charmingly outdated! In 1948, boys "derogated" each other by saying things like "all of them are stinkers" and calling their enemies "smart alecks.") Sherif did provide some numbers and percentages and a few chi squares, but this was a field study, with all of the uncontrollable variables that field studies can generate. As "science" it would not meet today's standards. Was everything all hunky-dory for the Eagles and Rattlers afterward? The numbers of boys favorable toward the out-group improved, but the majority of boys in each group apparently maintained their hostility toward each other.



Yet Robbers Cave was and remains important for its central ideas: At the time, most psychologists did not understand—and most laypeople don't understand even today—that simply putting two competing, hostile groups together in the same room to, say, watch a movie won't reduce their antagonism; that competitive situations generate hostility and stereotyping of the out-group; and that competition and hostility can be reversed, at least modestly, through cooperation in pursuit of shared goals. That's the story of Robbers Cave: it was true then, and it's true now.

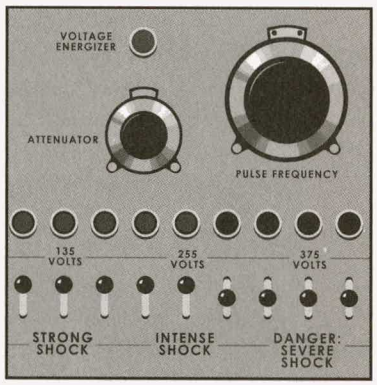
In fact, just as the Kitty Genovese story spurred bystander-intervention experiments, Robbers Cave generated a long line of experimental and field studies replicating the importance of superordinate goals. When Elliot Aronson went into the newly desegregated but hostile classrooms in Austin, Texas, where African American, Mexican American, and Anglo children were at war with each other, Sherif was part of his mental set, strongly influencing his design of the jigsaw classroom (organizing classroom activity to make students depend on each other to succeed). But Elliot did it right, using an experimental intervention and a control group. What a great coda to the Robbers Cave story—a direct link from Eagles and Rattlers, a made-up antipathy, to interethnic warfare in our schools, which is all too real and persisting.

Teaching the lessons of Stanley Milgram's experiments, of course, is far more complicated than Sherif's. Again, the cultural context of the times is crucial. In 1961, when Adolf Eichmann was claiming at his trial that he was "only following orders" in the murder of Jews during the Holocaust, Milgram began his effort to determine how many Americans would obey an authority figure when directly ordered to harm another human being. Milgram noted that he recruited not only undergraduate students, but also "factory workers, city employees, laborers, barbers, businessmen, clerks, construction workers, sales people, telephone workers." In the experiment, Milgram assigned his subjects to the role of "teacher" explaining that they were told that they were participating in an experiment on the effects of punishment on learning. The subjects were to read a list of paired words to the "learner" (a confederate working for Milgram), then present the first word of each pair again, upon which the learner was to recall the second word. Each time that the learner was incorrect, the teacher was to deliver an electric shock from a box with toggle switches in 15-volt increments that ranged from 15 volts all the way to 450 volts, and featured such labels as *Slight Shock, Moderate Shock, Strong Shock, Very Strong Shock, Intense Shock, Extreme Intensity Shock,* and *DANGER: Severe Shock, XXX.*

Some people hated the method and others the message, but the Milgram study has never faded from public attention and debate about it continues. In her book *Behind the Shock Machine*, Gina Perry, an Australian journalist, interviewed everyone she could find who was connected to the original study, along with Milgram's critics and defenders. She pored through the archives of Milgram's voluminous unpublished papers. Her goal was to argue that the experiments were flawed and unethical, in order to counteract what she considers Milgram's "bleak view of human nature."

Reinvestigations almost invariably yield some useful discoveries. Perry found violations of the research protocol: Over time, the man playing the experimenter began to drift off-script, urging reluctant subjects to keep going longer than he was supposed to.

To my own dismay, I learned that Milgram committed what researchers, even then, would have considered a serious breach of ethics: He did not fully debrief subjects at the end of each experiment. They *did* meet the "learner" to shake hands and be assured that he was fine, but they were not told that all those escalating levels of shocks were completely fake, because Milgram was afraid the word would get out and invalidate future participants' behavior. It was almost a year before subjects were mailed a full explanation. Some never got it; some never understood what the whole thing was about.



For critics like Perry, these flaws are reason enough to kick Milgram off his pedestal and out of our textbooks. I disagree. I think we must continue giving his experiments the prominent position we do, and for the same reason we originally did. When I first read about Milgram's experiments in grad school, I remember thinking, "Very clever, but what do they contribute? Wasn't Nazi Germany evidence enough of obedience to authority?" But that was Milgram's point: in the early 1960s, Americans—and American psychologists—deeply believed in national character. Germans obeyed Hitler, it was widely assumed, because obedience was in the German psyche: look at all those high scores on the Authoritarian scale. It could never happen here.

Elliot Aronson tells the following story in his memoir, *Not by Chance Alone*. When he was at Harvard in 1960, his first year out of grad school, he gave a guest lecture in Gordon Allport's class. Allport, the grand old man of social psychology, introduced him as a "master of mendacity" because of the dramatic experiments on cognitive dissonance that Elliot was already becoming famous for. Elliot was mildly insulted, naturally, and in talking with Allport afterward he defended the use of "deception" in high impact experiments as "not lying, but theater." Allport replied: "Why do you guys go through all that rigamarole? Why don't you just ask the participants what they would do?" This from Gordon Allport! Elliot tried to explain that most people cannot predict or account for their own behavior with any degree of accuracy, but Allport was unpersuaded.

A month or two later, Elliot went to Yale to give a colloquium and met Milgram for the first time. Milgram described the experiment he was planning and laid out its basic design. Elliot said, "Wow, I'll bet a sizeable number of people dole out more intense shocks than they themselves would ever have predicted." Even he, however, never dreamed that two thirds would go all the way.

For me, reading Perry's criticisms made it all the clearer why the Milgram experiments deserve their prominence. "Deep down, something about Milgram makes us uneasy," she writes. There is indeed something that makes everyone uneasy: the evidence that situations have power over our behavior. This is a difficult message, and most students—indeed, most people—have trouble accepting it. "I would never have pulled those levers!" they cry. "I would have told that experimenter what a... stinker...he is!" Perry insists that people's personalities and histories influence their actions. But Milgram never disputed that fact; his own research found that many participants resisted. Milgram wrote: "There is a tendency to think that everything a person does is due to the feelings or ideas within the person. However, scientists know that actions depend *equally* on the situation in which a man finds himself." Notice the "equally" in that sentence; many critics, like Perry, don't.

One of the original subjects in the experiments, called Bill, tried to explain to Perry why the experiments were so valuable, and why he did not regret participating, although he went to the end. He hadn't thought about the experiment for 20 years, he said, until he began dating a psychology professor. She, thrilled to have met a living link to the experiment, asked him to speak to her class. "Well," Bill tells Perry, "you would have thought Adolf Hitler walked in the room. I never really thought about it that way, you know?" Bill told the students, who were silently sitting in judgment on him: "It's very easy to sit back and say, 'I'd never do this or that' or 'Nobody could ever get me to do anything like that.' Well, guess what? Yes, they can."

That, of course, is the moral of the story. But the wall of hostility that Bill felt from the students means that they, like Gina Perry, weren't *getting it*. They were reading about the experiment, seeing the films, and still not understanding that they themselves might have been Bill.

Perhaps one way to help the Milgram medicine

go down is to show how it generated research on the psychology of the minority who resisted. The obedience studies might shock or depress students who think they provide a "bleak view of human nature," but these experiments of majority behavior *also* launched research into the conditions under which a brave minority becomes more likely to dissent, blow the whistle, disobey, and otherwise resist tyranny. That is, Milgram's work spurred investigation into the fuller human story: the bleak and the inspiring, the conformist and the rebel.

I turn now to Harry Harlow's classic experiments, conducted throughout the 1950s and 1960s, on the importance of contact comfort. Harlow took infant rhesus monkeys away from their mothers and raised them with a "wire mother", a forbidding construction of wires with a milk bottle connected to it, and a "cloth mother," a similar construction but one covered in foam rubber and terry cloth. At the time, it was widely believed (by psychologists, if not mothers) that babies become attached to their mother simply because mothers provide food. But Harlow's baby monkeys ran to the terry-cloth mother whenever they were frightened or startled, and clinging to it calmed them down. They went to the wire mother only for milk, and immediately abandoned it. Every intro class and textbook tell the story of the wire and cloth mothers, with those heartbreaking photos of infant monkeys clinging to their cloth mother when a scary moving toy was put into their cage. Wasn't this discovery, like Milgram's, something "we all knew"—in this case, that infants need contact comfort even more than they need food if they are to flourish? Didn't we have enough data from René Spitz's and John Bowlby's observations of abandoned infants warehoused in orphanages?

Well, no, apparently we didn't. As Deborah Blum describes in *Love at Goon Park*, most American psychologists at the time were under the influence of either behaviorism or psychoanalysis, two apparently opposite philosophies that nonetheless shared a key belief: that the origin of a baby's attachment to the mother was through food. Behaviorists believed that healthy child development required positive reinforcement: Baby is hungry; hunger drive satisfied, baby becomes conditioned to associate mother with food; mother and breast are equated. Interestingly, that was the Freudian view as well: no mother need be present, only a breast. "Love has its origin," Freud wrote, "in attachment to the satisfied need for nourishment." Why would cuddling be necessary? For the eminent behaviorist John Watson, cuddling was coddling.

But whereas Milgram's findings need constant reiteration in every generation, there is nothing surprising in Harlow's any more. One might say that the very success of his research has made teaching it unnecessary: no one would argue against Harlow's findings, as many students always want to do with Milgram's. Adult humans could choose to walk out of Milgram's experiment at some point, and a third of them did. But the monkeys were captives, tortured by their isolation. In recent decades, psychologists have learned that "torture" is not an exaggeration to describe the experience of isolation for any primate. And to torture *infants?* It's horrible. But the fact that so many people think it is horrible now—and didn't then—is an extraordinary story for teachers to tell. How has it happened that we have extended the moral circle to include other primates?

In 1973, as a young editor at *Psychology Today*, I interviewed Harlow. I walked through his lab with our photographer, Rod Kamitsuka, and looked aghast at a roomful of monkeys, each cowering alone in its own cage, electrodes on their heads. When Rod took a picture of one, it became wildly excited and fearful, careening around its tiny cage trying to escape. Rod and I were horrified, but Harlow was amused by us. "I study monkeys," Harlow said, "because they generalize better to people than rats do, and I have a basic fondness for people." I asked him what he thought of his critics who said that taking infants from their mothers was cruel and that the results did not justify the cruelty. He replied: "I think I am a soft-hearted person but I never developed a fondness for monkeys. Monkeys do not develop affection for people. And I find it impossible to love an animal that doesn't love back." Today, that sounds like lame moral reasoning: the fact that animals don't love us is no justification for torturing them.

When I revisited Harlow's work, however, I was reminded of how many pioneering discoveries he made, most of them lost in the telling of the main story of contact comfort. But he also demonstrated that monkeys use tools, solve problems, and learn

and explore because they are curious or interested in something, not just to get food or other rewards. He demonstrated the importance of contact with peers, which can even overcome the detrimental effects of maternal deprivation. Harlow created a nuclear family unit for some of the monkeys and found that under those conditions, rhesus males became doting fathers—something they don't do in the wild.

Harlow was hardly the first to demonstrate the power of "mother love," the necessity of contact comfort, and the devastation that ensues when an infant is untouched, unloved, neglected. Was experimenting with monkeys, by raising them in isolation with only wire or cloth mothers and causing them anguish that no observer could fail to see, essential to make the same point that Bowlby and Spitz had done? I don't know. *What Harlow did, like Milgram, was to make his case dramatic, compelling, and scientifically incontrovertible.* The evidence was based not on anecdote or observation, however persuasive, but on empirical, replicated data. As Blum shows, that's what it took to begin to undermine a scientific worldview in which the need for touch and cuddling—physical expressions of mother love—had been so deeply ignored.

When I was first thinking about this topic, I was prepared to argue that it is time to jettison Harlow, given that his findings no longer surprise nor serve to dissuade students of a deeply held belief. Perhaps that judgment reflects my ineradicable memory of seeing those helpless, suffering baby monkeys. But in revisiting his work, I changed my mind. We *should* keep him, we *should* discuss his discoveries, while expanding our story of what they mean. Harlow's work is a great contribution for the story of psychology: it shows not only how we thought about mothers, but also how we thought about monkeys. It shows how dominant psychological perspectives influence our lives—in his day, behaviorism or psychoanalysis; in our day, genetics and brain—seeping into the questions we ask and the studies we do. The take-home message for students is not "look how much smarter, kinder, and more ethical we are today than those guys were" but rather: (1) where would we be without these classics? what do they teach us about humanity that *made* them classics? and (2) What is happening in today's culture that affects the questions scientists are asking now—and the answers they get? Where might our own mistakes and biases lie—we, with all our Institutional Review Boards and informed consents, where are *our* failings of ethics and methods? The classics are living history, and we are not at the end of history—by any means. S