from summer camps to glass ceilings: the power of experiments

Social science experiments on a few individuals from similar backgrounds can give rise to strategies for coping with social problems, ranging from intergroup conflict to women's inequality in the workplace. How does research on such narrow groups contribute to broad social understanding and insight?

Photo from Muzafer Sherif, The Robbers Cave Experiment, *courtesy of Wesleyan University Press.*

This photo from the Robbers Cave experiment illustrates the conflict that can emerge when two groups are given competing tasks. Original caption: "Rattlers displaying blue jeans captured in the raid on Eagles and inscribed, 'The last of the Eagles.'"

A man in torn clothes sprawls across an urban sidewalk. He moans softly. Pedestrians hurry by with no more than a worried glance. No one stops to help. Someone watching from afar might wonder at such uncaring behavior; surely some conscientious person would stop. Moreover, these pedestrians are all young adults wearing clerical garb, seminarians studying for the ministry. They are hurrying to the church to deliver sermons on the Good Samaritan. Why did they not stop? Researchers who staged this test found that seminary students did not stop because they worried about being late. Their personal obligation to keeping an appointment outweighed their general commitment to helping others.

Experiments such as this one startle us into new ways of understanding people. Although we tend to explain why people do what they do—or, in this case, not do—as an expression of personal character, experiments show that the context of events determines behavior to a significant extent. Experimental studies carry great weight in the social sciences, gaining acceptance in prestigious journals and, in a high-profile example, last year's Nobel Prize in Economics. Some experiment results also get exposure in popular media, generating prime-time news coverage and Hollywood films.

Many people who hear about these experiments—and some social scientists, too—wonder how experiments achieve their power to convince, especially when their results often defy common sense. Experiments usually feature contrived conditions and record the behavior of at most a few hundred participants, many of whom are college students. Yet the results can tell us a lot about society.

the robbers cave experiment and summer camp movies

A sociological experiment in the 1950s demonstrated the effectiveness of a now common strategy in which competing corporations form joint ventures that would appear to prevent one firm from gaining advantage over the other (much like the United States and Russia cooperating on the space station). In 1954, Muzafer Sherif, an early proponent of social science experiments, set up a summer camp near Robbers Cave State Park in Oklahoma to test theories about group conflict and how to avoid it. He believed that individuals develop a group identity when they work together toward a common goal. Groups become more cohesive and rigid when faced with competition from another group. This competition creates frustration, triggering hostility and conflict between the groups. Sherif thought a solution to the conflict might be found in the same process by which groups form: working toward a common goal. If hostile groups have to work together, then members might learn to see each other as part of a combined larger group, which would reduce their conflict.

A group of 22 boys—all white, middle-class and close to their 12th birthdays—came to the Robbers Cave summer camp. Sherif and his colleagues divided them into two teams, the Eagles and the Rattlers. Each team completed projects

Contexts, Vol. 2, Issue 4, pp. 42-49, ISSN 1536-5042 electronic ISSN 153-6052

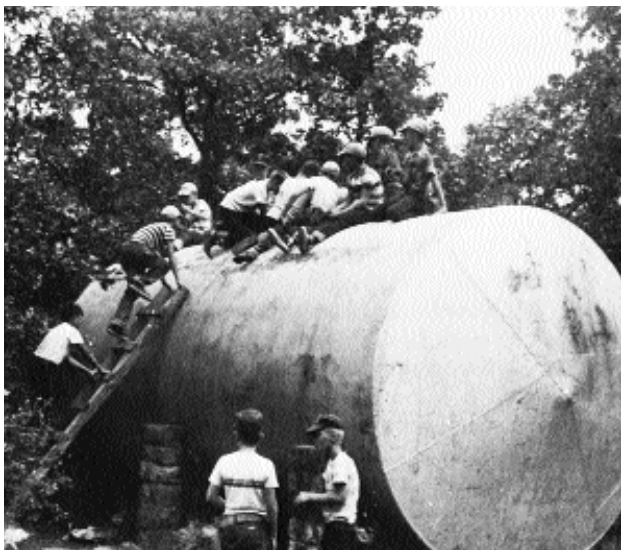

Photo from Muzafer Sherif, The Robbers Cave Experiment, *courtesy of Wesleyan University Press.*

A photo from the Robbers Cave experiment that shows the "Rattlers" and "Eagles" cooperating to achieve a common goal—restoration of the camp's water supply.

requiring the cooperation of members, such as building a diving platform at a swimming hole. In the second phase of camp activities, the two teams competed against each other in various contests. The results are familiar. Rivalry between teams generated hostility and even a little mayhem (exaggerated in subsequent summer camp movies), and threatened to spin out of control. Hostility emerged during the first contest—a baseball game. Boys in each group cursed members of the opposing group and called them names. At dinner, Eagles refused to eat with Rattlers. Later, the Eagles tore down the Rattlers' flag and burned it. The Rattlers retaliated by vandalizing the Eagles' cabin. A food fight erupted in the mess hall.

The experiment showed that hostility between groups develops spontaneously when individuals within a group work together and then compete as a team against another group. The final phase of the experiment showed how to reduce conflict. On a hot summer day, researchers disabled the water supply and asked volunteers to find the problem. Boys from both groups stepped forward, located the problem and worked together to solve it. Afterward, they all shared the water in a friendly manner. Finding water was important enough that it neutralized the groups' mutual antipathy, fostering cooperation and the beginning of trust.

An overarching cooperative task that requires the contributions of both groups for success reduced intergroup conflict. This principle is widely applied today, in contexts as distant as international relations, even though the experiment had nothing directly to do with such serious settings.

describing the world or testing theories

The logic of social experiments differs from that of other social research. Survey researchers, for example, try to describe a population of people by selecting a large, representative sample and then asking questions to determine respondents' attitudes and other characteristics. In contrast, experiments test theories rather than describe a population. That is, they test for evidence of a specific social process in a small sample of people, chosen to be as similar as possible. If a theory pre-

a reader's guide to social experiments

The hallmarks of good experimental research:

- A comparison between two groups as similar as possible but for one theoretically important difference (for example, undergraduate women assigned by coin flip to be team leaders or followers).
- Controlled conditions that allow the experiment to be repeated by other researchers.
- Follow-up studies that confirm the initial results and rule out competing explanations.
- A theory supported by experimental results that makes valid predictions in other contexts, spawning new research that reinforces the theory.

Pitfalls to avoid:

- Experimental results in one context cannot be simply exported to other contexts or cultures; they can support theories, which may then be used to make predictions for findings in other contexts.
- Ethical problems must be carefully considered. What effect might the research have on the lives of experiment participants?

STEADY P-TIME JOB

Working from own home; int. assignment doing charge acct. promotion for major Dept. Store. Good telephone personality required. No selling. Guar. hrly. Wage. Write Redwood City Tribune AD No. 499, include phone no.

Male college students needed for psychological study of prison life. $15 per day for 1-2 weeks beginning Aug. 14. For further information & applications, come to Room 248, Jordan Hall, Stanford U.

TELEPHONE CONTACTS

Sell security systems. New air cond. office, parking. Exc. location & understanding boss. $1.75 + comm. Call Mr. Nelson, 323-1339. Killmurray, 592-3950.

GET MORE OUT OF LIFE!

More money, more friends, more fun. Call now and learn about being an Avon Repre-

Photo courtesy of Philip G. Zimbardo, Stanford University

A newspaper advertisement used to recruit participants in the Stanford Prison experiment. The experiment tested the idea that normal American men would become brutal or be brutalized simply by taking on the role of a prison guard or prisoner.

dicts a particular result under certain conditions, experimenters then set up only those conditions. In this way, researchers can tell whether the predicted differences in behavior are produced by the conditions of the experiment instead of by individual differences among the participants.

Psychologist Philip Zimbardo's prison experiment at Stanford University is another famous example. He tested the theory that the brutal behavior of guards in prison camps (such as those in Nazi Germany) was a result of their being guards, rather than a result of their being individuals psychologically prone to act brutally. Zimbardo predicted that normal, mentally healthy, American men would become brutal or be brutalized simply because they became either prison guards or prisoners.

In the early 1970s, Zimbardo created a "prison" in the basement of the psychology building at Stanford. He selected only male Stanford undergraduates to participate, ruling out those with any prior psychological problems. He then randomly assigned the participants to be either prisoners or prison guards. The procedure is like flipping a coin. Heads and the participant becomes a guard, tails and he gets arrested. Random assignment helped to ensure that the two groups in the experiment—guards and prisoners—would be similar in other ways. Within a day of the prisoners' arrival, guards began acting brutally and prisoners showed signs of anxiety. Conditions rapidly deteriorated until the experiment had to be stopped. (Because social experiments directly change people's lives, extraordinary care must be taken to avoid causing harm. Some social experiments have the potential to be as dangerous as a clinical trial testing a new drug. Today, universities' Institutional Review Boards review proposed social experiments as stringently as they do medical and other scientific studies on people.)

The Stanford prison experiment helped shift thinking away from blaming German culture for the Holocaust and toward the social conditions that promote brutal behavior. The study received much media attention and was made into a popular German movie, *Das Experiment*. Ironically, the film version concluded that the solution to brutality is for individuals to take personal responsibility for their actions. But a solution that follows more consistently from the study itself is to construct social situations that discourage brutality. (see "Making Sense of the Senseless: Understanding Genocide," *Contexts*, Spring 2003.)

Why was this experiment so influential? It said nothing directly about German behavior during the Holocaust. Rather, it tested a theoretical prediction that a coercive setting can induce brutal behavior. A good experiment subtly shifts the burden of scientific proof, challenging other researchers to show whether a social process demonstrated in the experiment operates differently in a complex, naturally occurring setting. Simple experiments are convincing in part because they demonstrate a difference in the behavior of people in contrasting situations. Simplicity helps build agreement; most people observing the results of the Stanford and Robbers Cave experiments would interpret their meanings similarly. Controlled conditions also allow other researchers to repeat the experiments to see if the same results occur, perhaps using slightly different procedures. Good experiments can in these ways extend theories and produce new knowledge.

Of course, no single study, theory or method, no matter how good, establishes a scientific fact. Instead, science synthesizes different kinds of research from a variety of researchers to reach its conclusions. An experiment such as Zimbardo's Stanford prison makes a simple yet forceful statement that builds on earlier and inspires later research pointing

Photo courtesy of Philip G. Zimbardo, Stanford University

"Simulated" prisoners and guards. Student guards in the Stanford Prison experiment were free, within limits, to do whatever they thought was necessary to maintain law and order in the prison, and to command the respect of the prisoners. Within a day of the prisoners' arrival, guards began acting brutally and prisoners showed signs of anxiety. Conditions rapidly deteriorated until the experiment had to be stopped.

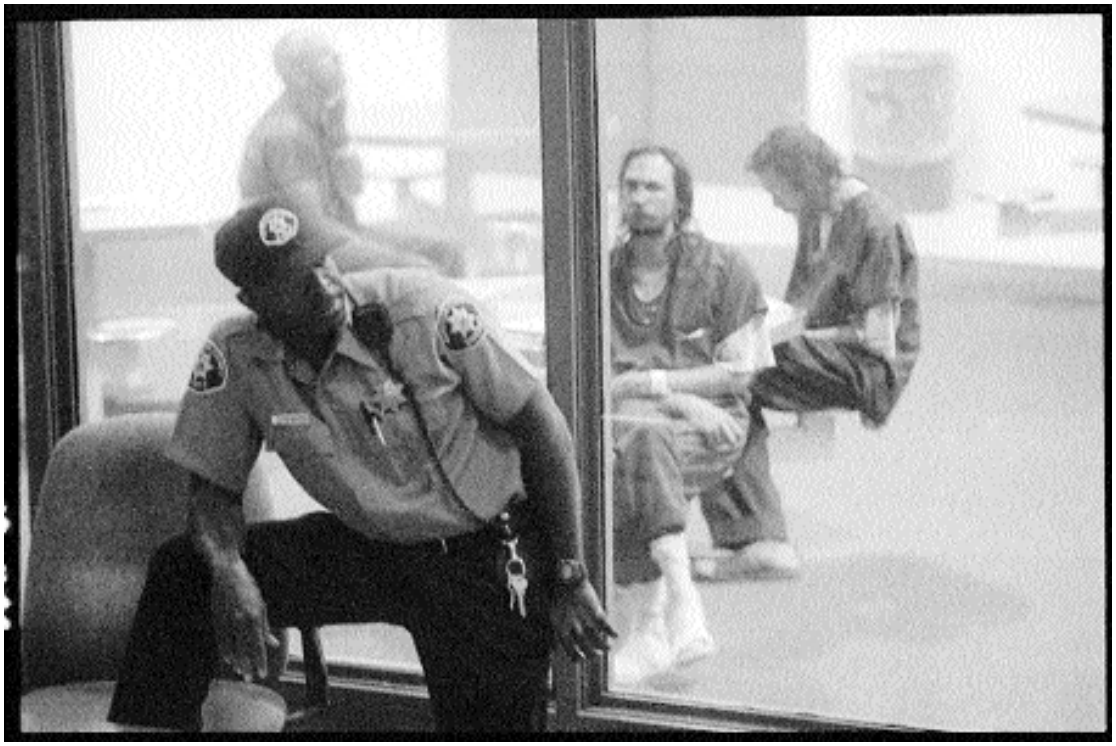

Photo by David Bacon <http://dbacon.igc.org/>

Real prisoners and guards in the Alameda County Jail at Santa Rita. The idea tested in the Stanford Prison experiment—that coercive settings can induce brutal behavior—has been used to help understand such behavior not only in prisons, but in many other circumstances of forced incarceration, including state-supported violence such as the Holocaust.

to a conclusion. Eventually, we better understand the social processes underlying a problem and can attempt a practical intervention. Experiments also can be used to directly assess the effectiveness of alternative social policies.

arresting domestic violence: experimenting with social policy

In 1981, police in Minneapolis changed the way they responded to reports of domestic violence. Before 1981, police officers had the discretion to arrest the person who committed the assault, order him (or her) to leave the home for a short period or provide on-site counseling. Advocates expressed concern that police were treating episodes of domestic violence too leniently, thereby failing to deter future assaults. Lawrence Sherman and Richard Berk designed an experiment to test whether making an arrest in a domestic violence case deterred future assaults better than the other two options of separating the couple and counseling.

The experiment had important implications for public policy, but it also addressed a long-standing dispute between two theoretical traditions in criminology. Deterrence theory holds that punishment discourages future criminal behavior. This school of thought maintains that suspects who are arrested will be less likely to commit another assault than those who are separated or counseled. A second theoretical tradition, known as labeling theory, suggests that when individuals are arrested, they become stigmatized as criminals by both society and in their own eyes. Their new self-image as a criminal then increases the likelihood of subsequent criminal behavior. (Labeling theory is the reason that names of juvenile offenders are kept out of the media except for serious offenses.) If labeling theory is valid, then those arrested for domestic violence actually would be more likely to commit another assault.

During the Sherman-Berk experiment, whenever Minneapolis police officers responded to a domestic violence call, they determined which procedure—arrest, separation or counseling—to follow by random assignment. Researchers tracked the behavior of suspects in the study for six months following the domestic violence incident. Results showed a deterrent effect for arrest and no evidence for labeling theory. That is, suspects who had been arrested were slightly less likely to commit another assault during the subsequent six months than were those who had been separated or counseled.

Although the deterrence effect of arrest was small, the experiment had a large effect on public policy. Arrest in domestic violence cases became the preferred procedure in many police departments and 15 states passed mandatory arrest laws. Meanwhile, debate over implications for social theory continued. During the next decade, other researchers repeated the experiment in several other police jurisdictions. The new results were more complicated. Arrest deterred suspects who were employed, perhaps because arrest is more serious for those who have a lot to lose. For unemployed suspects, arrest had the opposite effect, as predicted by labeling theory. They were more likely to commit a subsequent assault than the unemployed men who had been separated or counseled. The theoretical advance was exciting, but it left policy implications unresolved. In practice, police officers are still uncertain whether making an arrest will be beneficial in a domestic violence case. More systematic research could better equip police and judges to make such critical, sometimes life-and-death decisions.

We may need a system that produces public policies in a way similar to the system of clinical trials that produces new medical drugs. None of the alternatives available to the police in the Minneapolis experiment was new. But we do not have an organized system to formulate new policies, test them, and then compare them to alternative policies in controlled experiments. Such a system is worth considering. It might lead to more effective public policy the way that our system of developing new drugs has led to more effective medicine.

why do some groups score low on standardized tests?

Low intelligence seems the obvious explanation for low scores on a mental ability test. But what if something besides intelligence determines test scores? In the 1990s, psychologist Claude Steele's experiments yielded the startling discovery that scores on standardized tests depend not only on students' ability to answer, but also on what they expect the consequences of their test scores to be. Students who are stereotyped as having low ability may underperform when they are apprehensive about getting a low score.

Steele and his colleagues conducted a simple experiment. They gave a difficult standardized test—like the college SAT but harder—to a group of Stanford students. Instructions for taking the test varied. Some students, selected at random, were told the test results could be used to compare their performance to that of other students. Some students were told the test was only to familiarize them with similar tests they would encounter at the university. When students were told the tests were just for familiarization, black students scored about the same as white students of similar academic attainment. But when students thought they were going to be com-

Photo by David Bacon <http://dbacon.igc.org/>

A series of experiments has demonstrated that students' performance on a test is profoundly affected by what they understand to be the test's purpose and significance. This helps explain discrepancies between racial groups on standardized exams.

pared, black students scored lower than did comparable white students—as is common on standardized tests.

My colleagues and I conducted subsequent experiments showing that Steele's theory was not limited to particular racial groups, but applied to any stigmatized group. We randomly assigned white university undergraduates to be treated as an advantaged "majority" or disadvantaged "minority," by telling some students that their left- or right-handedness made it unlikely that they would be able to contribute to a group project, and also that other group members might resent their inability to contribute. Then, we gave the students a standard test of mental ability, explaining that the results of the test would be used to assign them to group positions such as "supervisor," "analyst," or "menial" in the group project. We found that students' test scores were substantially lower if they were treated as a disadvantaged "minority" for as little as 20 minutes.

The line of research begun by Claude Steele now includes many studies by different researchers. They show that when black and white students take the same standardized test, different expectations for the consequences of the test—not differences in mental ability—determine whether white students have an advantage. That is, while the best mental ability tests do a fair job of determining differences in cognitive skills among otherwise similar individuals, differences in test scores between racial and ethnic groups are created by social conditions rather than by the groups' mental abilities.

Applied programs based on this research show promise for increasing the academic performance of disadvantaged students. One surprising detail is that the performances of the best black students suffer the most. The threat of fulfilling a negative stereotype is felt most keenly by black students with the potential to excel; it is they who worry most about the potential backlash from their competition with white students. This may explain why remedial programs to improve academic performance of weaker students have not closed the gap between blacks and whites generally. Honors programs that encourage black students to undertake accelerated studies may have more effect, because promising black students have more academic ability than their grades and test scores suggest. Claude Steele helped develop a successful program to improve the performance of incoming minority students at the University of Michigan that emphasizes high academic standards, affirming students' ability to achieve those standards, and building trust that successful minority students can be accepted in the academic community.

how can women attain status equal to men at work?

Social experiments can also suggest strategies individuals can use to improve their lives. Status Characteristics Theory explains how individuals attain influence in work groups: people who are expected to contribute more to the group gain more influence in the group and receive greater rewards from the group. That is, expected contributions often count more than actual contributions. Individuals expected to perform well are more often followed by the group and rewarded accordingly. For example, a woman may make a brilliant suggestion that guarantees a successful project, but her suggestion may be ignored until a respected male coworker endorses it. He then gets the credit.

Research using the theory confirms that people expect men to contribute more to group success than women and that men do have more influence in decision making. Men get more credit for the group's successes and less blame for the group's failures. And when group members are evaluated, men get higher performance ratings and bigger rewards. To achieve the same level of rewards, women must work harder and contribute more than men. Status Characteristics Theory can also explain the familiar strategies women have used to break through to positions of influence in the workplace. Traditionally, they have out-competed men, following a masculine model that includes demonstrating competence through hard work and aggressive, even ruthless, competition. Successful women sometimes feel that they have sacrificed too much of themselves by following "male" strategies.

In the early 1980s, Cecilia Ridgeway conducted experiments using this theory that produced remarkable results for professional women struggling for career advancement under a glass

the hawthorne experiment

In the late 1920s and early 1930s, a Western Electric Company assembly plant near Chicago was the site of a series of studies aimed at developing scientifically based strategies for increasing worker productivity.

One experiment led to a concept called the "Hawthorne Effect." The researchers took a small group of female workers away from their peers, and placed them in a separate room so the experimenters could study the effect of changes in lighting, work procedures and break times on their productivity. It came as no surprise that improved lighting increased the workers' productivity, at least at first. But when the experimenters lowered the lighting to earlier levels, productivity continued to increase. Similar results after changing other aspects of the workers' environment led researchers to a conclusion that has since become known as the Hawthorne Effect: Workers increased their efforts because they were getting attention from the researchers, and because they bonded together as members of a prestigious "special" group.

Though legendary in its implications, the experiment has been criticized for design flaws and for confounding key variables. For example, two members of the study group were replaced mid-experiment with two new workers selected for their industriousness and cooperativeness. Simultaneous investigations by other sociologists revealed that workers who bonded strongly could unite to suppress work effort as well as speed it up.

Despite such shortcomings, reports of the Hawthorne experiment were used with enthusiasm by advocates of the human relations approach to workplace management. They felt that the results of the experiment challenged the scientific management perspectives that had shaped the Hawthorne studies in the first place. As a concept, the Hawthorne Effect—which posits that many interventions work, whatever they are, simply because people respond to being studied—also has been applied to a range of situations, such as student achievement in experimental schools, community organizing and military campaigns. Such applications confirm the power of relatively small experiments to stimulate thinking about issues of great importance, both for sociologists and for the larger public.

AT&T photo, courtesy of AT&T

The main room of the Western Electric Company's Hawthorne assembly plant, 1927. Workers in this area of the plant constituted the control group for some of the Hawthorne experiments on worker productivity.

AT & T photo, reprinted with permission of AT&T

The subjects of the "Hawthorne Experiment." These female employees of the Western Electric plant worked in a separate room (shown here) so that researchers could test the effects of different experimental treatments on their productivity.

ceiling. Ridgeway realized that people value not only the ability of a person to contribute, but also whether that person is motivated by a desire to help the group; they would not expect a person who is competent but selfish to contribute much of value. Ridgeway proposed that, because of gender stereotypes, however, people expect that even selfishly motivated men will contribute to the group, but expect contributions from women only when women demonstrate that they care about the group.

Ridgeway conducted an experiment to test this theory. Four team members worked together to reach a decision. One of the team members—secretly collaborating with the experimenters—made comments that were either group-motivated ("It is important that we cooperate") or self-motivated ("I want to win points for myself"). As predicted, in the self-motivated condition, male collaborators had more influence over the groups' decisions than female collaborators. In the group-motivated conditions, however, women collaborators' influence increased while the men's stayed at about the same high level as when they appeared selfish. Put another way, group- motivated women had as much influence as equally competent men regardless of the males' motivations.

The results suggest a strategy to succeed at work that women could use as an alternative to the competitive male one. Demonstrated competence is primary. Assertiveness also helps, but the focus on ruthless competition may be unnecessary for women's success. Instead, emphasizing a concern for other group members and the importance of working together to accomplish group goals can help competent women achieve recognition for their contributions. Future research in actual workplaces will help refine an effective strategy.

from theory to practice

The power of experiments flows from their use to test general theories. Sherif's Robbers Cave experiment tested a theory that explains how cooperation forms within groups and competition develops between them. Ridgeway tested her theory that influence in groups flows from the expectations people have about the ability and motivation of group members to contribute to group success.

Alone, a social experiment only demonstrates some phenomenon in one restricted context. But when experiments test theories, and their results lead to more tests in wider contexts, as well as other research with other methods, then we gain knowledge capable of transforming society. The experiments described have inspired lines of research with the potential to increase cooperation among competing organizations, decrease domestic violence, reduce the racial gap in academic success, and remove the glass ceiling limiting women in business. They successfully made the leap from small groups to helping us understand society at large. ■

recommended resources

American Sociological Review. "Employment, Marriage, and the Deterrent Effect of Arrest for Domestic Violence: Replications and Re-Analyses," 57 (1992): 679-708. Three research articles analyze follow-up studies to the original Sherman and Berk experiment on police responses to domestic violence.

Berger, Joseph and Morris Zelditch, Jr., eds. *Status, Rewards and Influence*. San Francisco: Jossey-Bass, 1985. An overview of research on status processes in task groups.

Cohen, Bernard P. *Developing Sociological Knowledge: Theory and Method*. Chicago: Nelson Hall, 1989. A classic text that describes social science as a reciprocal process in which research tests theory and theory develops through the interpretation of research.

Lovaglia, Michael J. *Knowing People: The Personal Use of Social Psychology*. New York: McGraw-Hill, 2000. An accessible overview of social psychological research useful for individuals in their personal lives.

Lovaglia, Michael J., Jeffery W. Lucas, Jeffrey A. Houser, Shane R. Thye and Barry Markovsky. "Status Processes and Mental Ability Test Scores." *American Journal of Sociology* 104 (1998): 195-228. A research article demonstrating the adverse effect of a negative stereotype on the standardized test scores of white students.

Ridgeway, Cecilia. "Status in Groups: The Importance of Motivation." *American Sociological Review* 47 (1982): 76-88. An important research article that suggests a way for women to achieve equal status in the workplace.

Sherif, Muzafer, O. J. Harvey, B. Jack White, William R. Hood and Carolyn W. Sherif. *The Robbers Cave Experiment: Intergroup Conflict and Cooperation*. Middletown, CT: Wesleyan University Press, [1961] 1988. The classic study on the origins of conflict between groups and a method for bringing together competing groups.

Steele, Claude M. "Thin Ice: 'Stereotype Threat' and Black College Students." *Atlantic Monthly*. August, 1999. An accessible overview of research and applied programs for reducing the disadvantage of black college students.