Critically Evaluating Competing Theories: An Exercise Based on the Kitty Genovese Murder

Brad J. Sagarin
Northern Illinois University

Kimberly A. Lawler-Sagarin
Department of Chemistry
Elmhurst College

We describe an exercise based on the 1964 murder of Catherine Genovese—a murder observed by 38 witnesses, none of whom called the police. Students read a summary of the murder and worked in small groups to design an experiment to test the competing theories for the inaction of the witnesses (Americans' selfishness and insensitivity vs. diffusion of responsibility and pluralistic ignorance). A pretest–posttest assessment revealed a significant increase in students' ability to design experiments to test competing theories, and anonymous feedback revealed that the exercise encouraged students to appreciate the complexity of trying to explain real-world phenomena.

This article describes an exercise designed to teach students to evaluate competing theories for the same phenomenon. Encouraging students to evaluate critically both commonsense and social psychological explanations for real-world phenomena is an instructional challenge. Students faced with the task of explaining complex phenomena often seek a single "right" answer, particularly in introductory classes (Perry, 1970). Ideally, the undergraduate curriculum develops students' critical thinking and reasoning skills, moving students toward higher levels of cognitive development. At these higher levels, students realize that complex phenomena are often multiply determined and are seldom fully explained by a single theory (Kurfiss, 1994; Magolda, 1992; Pascarella & Terenzini, 1991; Perry, 1970).

Advances in scientific knowledge often require challenging accepted answers. To help advanced undergraduates make the transition to producers of scientific knowledge, we developed an exercise that revisits a historical event many psychology students first encounter in lower division psychology classes. However, whereas these lower division classes typically present the event with its accepted psychological explanation, this exercise encourages students to evaluate critically the accepted explanation and to design empirical tests that pit this explanation against a plausible alternative explanation. From the perspective of Bloom's (1956) taxonomy of higher education objectives, we designed the exercise to move students from the lower levels of knowledge and comprehension to the higher levels of analysis and evaluation.

We based the exercise on a tragic event often used to introduce research on the bystander effect: the 1964 murder of Catherine Genovese in Queens, New York. Most shocking was the fact that 38 witnesses observed the attack, which took place over 35 min, but none called the police (Gansberg, 1964). In the aftermath, commentators and social psychologists offered two competing explanations for the inaction of the bystanders. Commentators suggested that, "Americans were becoming a nation of selfish, insensitive people ... The Cold Society" (Cialdini, 2001, p. 113). Social psychologists, in contrast, suggested that the presence of so many witnesses inhibited helping due to diffusion of responsibility (i.e., the reduction in personal responsibility caused by the presence of other witnesses) and pluralistic ignorance (i.e., the inhibition of action in ambiguous situations caused by potential helpers observing the inaction of other onlookers; Cialdini, 2001; Darley & Latané, 1968; Latané & Darley, 1970).

The exercise had a fourfold purpose: (a) to motivate students to learn the details of the theory and research on the bystander effect, (b) to have students appreciate the complexity of trying to explain real-world phenomena, (c) to increase students' willingness to challenge accepted explanations for phenomena, and (d) to teach students to design experiments to test competing theories. We used anonymous feedback surveys and pretest–posttest assessments to evaluate the exercise's effectiveness in a senior-level laboratory in personality and social psychology course.

Method

Participants

Fifty-eight senior-level psychology majors (45 women, 13 men) from Northern Illinois University participated as part of a class activity.

Procedure

The instructor asked students to read the account of the murder of Catherine Genovese from Cialdini's (2001) *Influence: Science and Practice* (pp. 111–114). This excerpt summarized the event, the publicity that followed, the "cold society" explanation put forth at the time based on urban Americans' selfishness and insensitivity, and the accepted social psychological explanation based on diffusion of responsibility and pluralistic ignorance. Students read the text in class to ensure that all would be prepared to participate actively in the subsequent activities and discussions.

Next, the students worked in groups of 3 to 4 to extract the competing explanations for the inaction of the bystanders and to determine which explanation they found most compelling. The instructor then led a class discussion on these explanations. During this discussion, the instructor presented additional material on the bystander effect—the paradoxical phenomenon that the greater the number of bystanders who witness a person in need, the less likely it is that the person will receive help (Cialdini, 2001; Darley & Latané, 1968; Latané & Darley, 1970).

The summary included descriptions of two specific experiments. In Darley and Latané (1968), a staged epileptic seizure elicited help 85% of the time from lone bystanders but only 31% of the time from groups of five bystanders. In Latané and Darley (1970), smoke seeping into the laboratory was reported to a confederate waiting outside the lab 75% of the time by a lone participant, 38% of the time by 3 participants, and 10% of the time by 1 participant sitting with 2 passive confederates. The summary also included some general findings from the bystander intervention research (not all of which were congenial to the diffusion of responsibility and pluralistic ignorance explanation): (a) in ambiguous situations, lone bystanders offer more help than multiple bystanders; (b) pluralistic ignorance is greatest in groups of strangers; (c) bystander inaction is worst in cities because cities are loud and distracting, multiple witnesses are more likely to be present, and the multiple witnesses are likely to be strangers; and (d) that very high levels of helping (90% to 100%) occur when the need is clear, regardless of the number of bystanders (Cialdini, 2001).

Students became highly engaged during the ensuing discussion, raising arguments in favor of each explanation. Those favoring the diffusion of responsibility and pluralistic ignorance explanation noted the large number of witnesses and the likelihood that individual witnesses had assumed that someone else had already called the police. Those favoring the "cold society" explanation pointed out that the 38 witnesses were not strangers, with one student noting that the witnesses called each other during the attack, although none called the police. The students further pointed out that Genovese's cries for help likely made this a situation of unambiguous need.

The instructor then asked students to break into small groups and design an experiment that would test the competing theories. The instructor circulated among groups, answering questions and offering suggestions. The instructor asked groups that had completed their design to describe the pattern of results that would support each explanation—a question that helped determine whether the design independently tested both theories.

Many groups created 2 × 2 factorial designs, with one independent variable testing each theory. The groups varied considerably, however, in the operationalization of the independent variables and the design of the situation of need. For example, among groups that incorporated a rural versus urban independent variable, some used participants' hometown or current place of residence, whereas others randomly assigned participants to witness an emergency in one location or another.

When all groups had completed their experimental design, each group presented its design to the class, with the other students providing constructive criticism. The class evaluated each design based on its ability to distinguish between the competing explanations. After each presentation, the rest of the class commented on positive design features such as particularly clever methodologies and pointed out problems such as confounds and limitations such as quasi-experimental designs. The instructor also discussed weaknesses of alternative (e.g., two-cell, urban/one-witness vs. rural/multiple-witnesses) designs. The exercise required one-and-a-half 75-min class periods.

Evaluation

Pretest. One week prior to the class activity, students completed a pretest to assess their ability to design experiments to test competing theories for the same phenomenon. The pretest presented students with a scenario involving competing theories and asked students to design an experiment to test the competing theories (see Appendix).

Posttest. Two weeks after the class activity, students took the final exam for the class. The first question presented a scenario similar to the scenario that appeared on the pretest (see Appendix) and asked students to design an experiment to test the competing theories presented in the scenario.

Feedback survey. One week after the class activity, students completed an anonymous feedback survey. The survey contained two open-ended questions ("What did you learn from doing the exercise?" and "How did you feel about this exercise as a way to discuss the topic of the Kitty Genovese murder?") and one yes–no question ("Would you recommend that I use this exercise in future semesters?").

Results

Two graduate students coded participants' pretest and posttest answers on a 4-point scale ranging from 1 (*an incorrect answer*), 2 (*a mostly incorrect answer with some correct elements*), 3 (*a mostly correct answer with some incorrect elements*), to 4 (*a correct answer*). The mean of the coders' ratings comprised the final scores ($\alpha = .75$).

Among the 40 students who completed the pretest and posttest, posttest scores ($M = 3.50$, $SD = 0.59$) were significantly higher than pretest scores ($M = 2.62$, $SD = 0.69$), $t(39) = -7.31$, $p < .001$ (two-tailed, paired-samples t test). In another comparison, 45.0% of participants had a completely correct answer (i.e., received scores of 4 from both coders) on the posttest, compared to 10.0% on the pretest. Similarly, 87.5% of participants had a mostly correct answer (i.e., received scores of 3 or 4 from both coders) on the posttest, compared to 25.0% on the pretest. These results suggest that students became substantially better at designing experiments to test competing explanations after the class activity. A replication in a subsequent semester of the same course ($N = 21$) that reversed the pretest and posttest scenarios produced similar results (posttest $M = 3.29$, $SD = 0.63$; pretest $M = 2.68$, $SD = 0.86$), $t(20) = -3.02$, $p = .007$ (two-tailed).

Of the 44 students who completed the feedback survey, 41 (93.2%) recommended using the exercise in future semesters. Two (4.5%) recommended against using the exercise in future

semesters, and one (2.3%) circled both "yes" and "no." These 3 students believed they had covered the material in other classes, although the ambivalent student noted that other classes had not discussed the "cold society" explanation.

The open-ended answers of the 41 students who recommended using the activity in the future revealed the following: On the question regarding what they learned from the exercise, 21 students (51.2%) mentioned learning about the topic (e.g., "I learned the concepts of the cold society and the diffusion of responsibility, and that it would be interesting to actually perform a study on this matter," "Topic of bystander effect is a lot more complicated than I had thought"). Ten (24.4%) mentioned learning how to operationalize theoretical concepts and to design experiments, specifically in ways that test two competing theories (e.g., "This exercise really made me think about how to formulate an experiment"). Six (14.6%) reported learning how different theories can explain the same event (e.g., "I learned the two components of the bystander effect and that neither bystander or cold society fully explain").

On the question regarding how they felt about the exercise as a way to discuss the topic of the Kitty Genovese murder, 34 students (82.9%) thought it was "very effective," "interesting," "[a] good exercise," and so on. Some appreciated the nonlecture format, particularly the group work (e.g., "I thought talking with others was a good way to generate ideas and gain a better understanding of the topic," "enjoyed having a small group to discuss and then getting everyone's ideas").

Conclusions

This article describes an exercise designed to teach students how to evaluate critically and test competing theories to explain complex real-world phenomena. Although the Kitty Genovese murder provides a thought-provoking event appropriate for social psychology classes, the approach could be adapted easily for other classes, as long as the selected event has two or more plausible explanations. Ideally, students should have encountered the event and its accepted explanation in prior classes, as a previously encountered event and explanation will maximize the impact of the exercise on students' willingness to challenge accepted answers.

Instructors might choose to save class time by having students read the background material outside of class. Alternatively, instructors could expand the activity by having students modify their experimental designs after receiving feedback from other groups. We suggest that instructors retain the collaborative experimental design part of the exercise, as students found the group activities engaging, and results suggested that it was successful in teaching students how to design experiments to test competing theories.

References

Bloom, B. S. (1956). *Taxonomy of educational objectives.* New York: Longmans, Green.

Cialdini, R. B. (2001). *Influence: Science and practice* (4th ed.). Boston: Allyn & Bacon.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8,* 377–383.

Gansberg, M. (1964, March 27). 37 who saw murder didn't call the police. *The New York Times,* pp. 1, 38.

Kurfiss, J. (1994). Intellectual, psychosocial, and moral development in college: Four major theories. In K. A. Feldman & M. B. Paulsen (Eds.), *Teaching and learning in the college classroom* (pp. 139–162). New York: Ginn.

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Englewood Cliffs, NJ: Prentice Hall.

Magolda, M. B. B. (1992). *Knowing and reasoning in college: Gender-related patterns in students' intellectual development.* San Francisco: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research.* San Francisco: Jossey-Bass.

Perry, W. G., Jr. (1970). *Forms of intellectual and ethical development in the college years.* New York: Holt, Rinehart & Winston.

Appendix

Pretest Scenario

An advertising agency produces an automobile ad that shows an attractive couple driving a car while a narrator describes the features of the car. When the agency tests the ad, they discover that consumers rate it as very persuasive. Unfortunately, the agency does not know whether the ad is persuasive because of the attractive couple or the description of the car's features. The agency hires you to design an experiment to test these two competing theories: (a) the car ad is persuasive because of the attractive couple, or (b) the car ad is persuasive because of the description of the car's features.

Posttest Scenario

A television studio produces a new comedy show that stars a talking cat named Checkers. Unfortunately, viewers rate the show as not funny. The studio decides to add a laugh track (the sound of an audience laughing after the jokes) and to broadcast the show after *The Simpsons,* the highly successful sitcom. Now, viewers rate the show as very funny, but the studio does not know whether the show is funny because of the laugh track or because it was broadcast after *The Simpsons.* The studio hires you to design an experiment to test these two competing theories: (a) the new show is funny because of the laugh track, or (b) the new show is funny because it was broadcast after *The Simpsons.*

Notes

1. We thank Evelyn Comber, Melissa Commisso, and Amy Martin for their help coding the pretest and posttest measures and Anne Britt, Evelyn Comber, and Greg Light for their comments on an earlier draft of this article.

2. Send correspondence to Brad J. Sagarin, Northern Illinois University, Department of Psychology, DeKalb, IL 60115; e-mail: bsagarin@niu.edu.