# Establishing a Causal Chain: Why Experiments Are Often More Effective Than Mediational Analyses in Examining Psychological Processes

Steven J. Spencer, Mark P. Zanna, and Geoffrey T. Fong
University of Waterloo

The authors propose that experiments that utilize mediational analyses as suggested by R. M. Baron and D. A. Kenny (1986) are overused and sometimes improperly held up as necessary for a good social psychological paper. The authors argue that when it is easy to manipulate and measure a proposed psychological process that a series of experiments that demonstrates the proposed causal chain is superior. They further argue that when it is easy to manipulate a proposed psychological process but difficult to measure it that designs that examine underlying process by utilizing moderation can be effective. It is only when measurement of a proposed psychological process is easy and manipulation of it is difficult that designs that rely on mediational analyses should be preferred, and even in these situations careful consideration should be given to the limiting factors of such designs.



As the field of social psychology has developed, researchers have increasingly turned to the examination of psychological processes. This development is perhaps inevitable as the field matures—fewer new phenomena are being discovered and increasing attention is being paid to understanding known findings. We firmly believe that understanding psychological processes is fundamental to advancing the field. Nevertheless, over the years we have become concerned with what we see as an overemphasis on one particular way of examining psychological processes: regression models based on a seminal paper by Baron and Kenny (1986). To be sure such analyses have their proper place—and we plan to describe it. However, it is our contention that this analysis strategy is overused and has perhaps been elevated as the gold standard of tests of psychological processes and may even be seen in some quarters as the only legitimate way to examine them. The aim of this essay is to point to broader ways of understanding psychological processes in experimental contexts and to highlight the strengths and weaknesses of various approaches.

The study of psychological processes has a long history in social psychology. Most classic theories in social psychology theories have compelling accounts of such processes. For example, cognitive dissonance theory (Festinger, 1957) proposes that when people have two thoughts that are psychologically inconsistent, they experience an aversive arousal that motivates them to change one of the cognitions. Clearly, the aversive arousal is the mediating psychological process in this account. In those early days, social psychologists typically tested their theories by demonstrating an effect and then made plausible arguments about psychological processes. Much of the time these arguments were not even tested.

For example, Festinger and Carlsmith (1959) demonstrated that people paid $1 to lie about a boring task by saying it was fun came to believe their lie, whereas those paid $20 to lie did not. They argued that people in the $1 condition changed their attitude because the inconsistency between what they believed about the task (i.e., that it was boring) and what they said about the task (i.e., that it was fun) created an aversive state of arousal—cognitive dissonance. They also argued that people were motivated to alleviate this aversive state, and because it was easier to change one's beliefs than to take back one's actions, people resolved the inconsistency by changing their beliefs about the task (i.e., they thought the task was fun).

As the field developed, it became more common for statistical evidence to be offered for such hypothesized accounts of psychological processes. For example, research on the elaboration likelihood model of persuasion (Petty & Cacioppo, 1986) began to include measures of thought listing as an indicator of elaboration, and research on group polarization began to examine the number of arguments that were made in a discussion to examine this factor as an indicator of process (Burnstein & Vinokur, 1977). These authors demonstrated that a factor produced an effect and then they measured the proposed psychological process and demonstrated that the process occurred more often when the effect occurred and that the measurement of the process correlated with the measurement of the effect. For example, Burnstein and Vinokur (1977)

Steven J. Spencer, Mark P. Zanna, and Geoffrey T. Fong, Psychology Department, University of Waterloo, Waterloo, Ontario, Canada.

The writing of this article was supported by grants from the Social Sciences and Humanities Research Council of Canada and the Canadian Institute of Health Research. We thank Joanne Wood and Judy Harackiewicz for their helpful comments that provided much needed clarification of our arguments. In addition (without implying that they agree with our analysis), we thank Reuben Baron, David Kenny, Lee Fabrigar, and four anonymous reviewers for their generous comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Steven J. Spencer, 200 University Avenue West, Psychology Department, University of Waterloo, Waterloo, Ontario N2L 3GI, Canada. E-mail: sspencer@watarts.uwaterloo.ca

Journal of Personality and Social Psychology, 2005, Vol. 89, No. 6, 845–851

0022-3514/05/$12.00 DOI: 10.1037/0022-3514.89.6.845

demonstrated that when groups discussed issues they generated a larger number of persuasive arguments for their position and that this greater number of persuasive arguments was associated with group-polarization effects. Despite the increased examination of variables that measured psychological process at that time, little more than within-cell correlations were conducted in the way of providing evidence for such processes.

This practice changed substantially with the publication of an extremely influential paper by Reuben Baron and David Kenny in 1986. In this paper, the authors described the appropriate statistical analyses that allow one to test whether an independent variable (A) causes an effect on a dependent variable (C) through a mediating variable (B). There is little question that this paper addressed an important gap in knowledge, was cogently argued, and brought a much needed statistical sophistication to the field. Nevertheless, from our view it had the unintended side effect of providing a paradigm of what a social psychology experiment or set of experiments should look like if they are to be published in our top journals. (In fact, in a recent analysis, Quiñones-Vidal, López-García, Peñaranda-Ortega, and Tortosa-Gil (2004) reported that the Baron and Kenny, 1986, paper has been cited in *JPSP* more than any other one has—226 times.) Nowadays, it seems that the prototype of a top-flight manuscript in social psychology reports a study or two that demonstrate that the independent variable of interest (A) leads to the dependent variable of interest (C). Then an additional study or studies demonstrate through Baron and Kenny (1986) type of analysis that A influences C through a proposed mediator (B). For the purposes of this essay, we will call this sort of design a *measurement-of-mediation* design.

For example, Fein and Spencer (1997) followed this strategy when they proposed that threats to self-esteem (A) lead to increased discrimination against a stereotyped target (C) and that this effect is mediated through feelings of self-worth (B). In Study 1, they demonstrated that when people's self-images are affirmed, they discriminate less against a stereotyped target, and in Study 2 they demonstrated that when people's self-images are threatened they discriminate more against a stereotyped target. In Study 3, they measured state self-esteem (B) after threat (A) and demonstrated that threat reduced state self-esteem, which in turn predicted discrimination against a stereotyped target (C).

Although we think this sort of measurement-of-mediation design is not wrong per se and indeed that it is the right sort of design in some situations, we think that it is not the best way to study psychological processes in many situations. We also believe it prevents the study of some issues that do not lend themselves to this sort of analysis. In this article, we hope to set out some of the basic options for designing social psychological experiments that examine psychological processes and note their strengths and weaknesses.

Before we begin, however, we think it is important to make a crucial distinction between mediation as a theoretical analysis (which we refer to throughout the essay as psychological process) and mediation as a statistical analysis. Indeed, we feel it is the confusion between these levels of analysis that has led to the measurement-of-mediation design as the default paradigm. From a theoretical perspective arguing for mediation (i.e., arguing for a particular psychological process) is simply refining the causal chain in one's theory. If a theory proposes that one factor causes another factor, the identification of the process through which that causation occurs (i.e., the intervening cause) is often an increase in knowledge and an important refinement of the theory. A statistical analysis à la Baron and Kenny (1986), however, is not the only way to provide such evidence. Let us start with a couple of classic examples.

In a series of studies, Word, Zanna, and Cooper (1974) examined whether stereotypes could create a self-fulfilling prophecy through nonverbal behavior. Note the causal theory of psychological process they were testing: Stereotypes (A) lead to behaviors consistent with the stereotype (C) because of the nonverbal behavior on the part of those who hold the stereotype (B). In the first study, they had White participants interview either a Black or a White confederate and observed that the White participants displayed more distant nonverbal behavior to the Black confederate than to the White confederate (i.e., they showed a relation between A and B). In the second study they had White confederates interview White participants, but the confederates either treated them like the Blacks were treated or like the Whites were treated in Study 1. Word et al. (1974) found that those who were treated like the Blacks were treated in Study 1 did worse on the interview (i.e., they showed a relation between B and C).

In our view, this study provides strong evidence for the theoretically proposed psychological process even though it does not test for mediation statistically. In fact, we believe that this sort of design, because it utilizes the power of experiments to demonstrate causality, often does a better job of demonstrating the proposed psychological process than does the measurement-of-mediation design. The reason we make this claim is that by manipulating both the independent variable and the mediating variable we can make strong inferences about the causal chain of events. We argue that such designs should be understood as a powerful way to examine psychological processes.

This argument, however, does not mean that such designs (which for the purposes of this essay we will label as *experimental-causal-chain* designs) do not have drawbacks. They certainly do. One of the difficulties in implementing this design is that one has to be able to measure the proposed psychological process. Many processes do not lend themselves to easy measurement. A second difficulty is that one must be able to manipulate both the proposed independent variable and the proposed psychological process. Sometimes this too is problematic. Third, to convincingly argue for a proposed psychological process with such a design, one must be able to argue that the proposed psychological process as it is measured and as it is manipulated are in fact the same variable. For example, in the Word et al. (1974) study, the authors had to make it clear that the nonverbal behaviors they measured in Study 1 were in fact the same nonverbal behaviors they manipulated in Study 2. Finally, these designs do not allow an easy analysis of how much of the effect of A and C is explained by B. In many experimental contexts, however, researchers are much more concerned with establishing causality than determining the amount of variance explained in an effect. Nonetheless, when variance accounted for in a dependent variable is a primary concern, then experimental-causal-chain designs are not likely to be appropriate. Despite these drawbacks, we think that such experimental-causal-chain designs are underutilized in social psychology and should be given greater consideration as people plan to test their theories.

A second classic example of how compelling evidence for a proposed psychological process can be amassed without a statistical mediational analysis can be seen in Zanna and Cooper's (1974) demonstration of the role of an aversive state of arousal in cognitive dissonance. Note the causal theory of psychological process they were testing: Conflicting cognitions between attitudes and behavior (A) lead to attitude change (C) through their effect on an aversive state of arousal (B). In that study, they gave people high or low choice to write a counterattitudinal essay. They then gave them a pill (actually a placebo) and randomly assigned participants to get one of three types of information about the possible side effects of the pill. Participants were told the pill would make them feel tense or aroused, that it would have no side effects, or that it would make them feel relaxed or calm.

Participants given no information about the pill showed the typical dissonance effect: They changed their attitude in the direction of their essay when they believed they had high choice to write the essay, but not when they believed they had low choice. Participants who were told that the pill would be arousing, however, showed no evidence of dissonance reduction: Their attitudes were the same whether they believed they had high or low choice to write the essay. Finally, participants told the pill would be calming showed an especially large amount of attitude change when they believed they had high choice to write the essay.

Zanna and Cooper (1974) argued that these findings provided evidence that aversive arousal was the psychological process through which cognitive dissonance led to attitude change, and the field seemed to agree. We believe that such designs (which for the purpose of this essay we label as *moderation-of-process* designs)[1] are underappreciated for their ability to demonstrate psychological processes. In our view, such designs provide strong support for a psychological process if they meet two key assumptions: first, that the proposed moderating variable has an effect on the proposed psychological process (B), that is, there has to be evidence that the moderator variable does indeed affect the hypothesized psychological process and, second, that the only way that the proposed moderating variable affects the relation between the independent variable (A) and the dependent variable (C) is through its effect on (B), that is, there can be no alternative explanation for the observed pattern of moderation. For an elaboration of this argument see Sigall and Mills (1998).

In the Zanna and Cooper (1974) study, the misattribution literature (e.g., Storms & Nisbett, 1970) provided abundant evidence that instructions about pills (i.e., the proposed moderating variable) could affect people's understanding of their state of arousal (B) and therefore they had compelling evidence that the first assumption above was met. In addition, there were no plausible alternative explanations for the effect of the instructions about the pills on attitude change. Together these findings provide reasonably strong evidence that conflicting cognitions between attitudes and behavior (A) had their effect on attitude change (C) through their effect on an aversive state of arousal (B).

We should note the drawbacks of this design as well. First, this sort of design generally takes independent evidence that the moderating variable has the intended effect on the proposed psychological process (B). Such evidence is not always easy to come by, but in some cases—such as the Zanna and Cooper (1974) study and in other paradigms that manipulate cognitive load to affect controlled processing of information—such evidence is readily available. Second, such designs have to have strong evidence that the moderator only affects the proposed psychological process and not some other psychological process. For example, Zanna and Cooper (1974) needed to make the case that the instructions about the pill would affect the attributed source of arousal and not other variables such as impression management.

Our analysis has led us to the following recommendations about experimental design summarized in Table 1. We argue that one important factor when designing an experiment that operationalizes and tests a psychological process is how easy it is to measure and manipulate this process. If it is relatively easy to measure and to manipulate the proposed process, then all other things being equal—and we acknowledge that the nature of the specific project and the art of designing studies to fit the topic might override these recommendations—the experimental-causal-chain design is probably the best bet. Some might argue that in this situation one could conduct multiple types of studies. Certainly if it is easy to both manipulate and measure the proposed psychological process, then all three types of designs are possible and, given unlimited time and resources, it might make sense to conduct them all. In our view, however, the experimental-causal-chain design represents the simplest and most straightforward way to examine the proposed process and is often the best strategy. Although using multiple methods to test a theoretical account would be ideal, we feel that in most situations requiring such multiple methods would be setting such a high standard that progress in the field might well be impeded. In our view, a properly implemented experimental-causal-chain design usually makes a quite compelling case for psychological process.

Unfortunately, we have often found that in many situations either manipulating or measuring (or both manipulating and measuring) the proposed psychological process is difficult. In such situations, the experimental-causal-chain design is difficult to implement. Other strategies are often more useful in such situations. For example, if you can easily manipulate the proposed psychological process, but it is difficult to measure it (e.g., it is an unconscious process that is difficult to measure or there are no reliable and valid measures of the construct), then the moderation-of-process design is probably the best bet. If it is difficult to measure or manipulate the proposed psychological process, then obviously more work has to be done on one of these fronts before the research can proceed. We argue that it is only when measurement of the proposed process is easy and manipulation of it is hard (e.g., a valid manipulation of the process is hard to create or manipulating the process would affect the nature of a process such as when a manipulation makes an implicit process explicit) that the measurement-of-mediation design is likely to be best.

Even when manipulating a proposed process is hard and measuring it is easy, however, the drawbacks of the measurement-of-mediation design must be taken into account. One major drawback of this sort of design (i.e., a design that utilizes a Baron and Kenny, 1986, type of analysis) is that the process not only needs to be easy to measure, but also measuring it must not interfere with how the

[1] We suspect that part of the confusion between moderation and mediation that Baron and Kenny (1986) noted was that social psychologists had become accustomed at the theoretical level to studying psychological processes by referring to them as mediation but they tested these processes with moderation at the statistical level (i.e., what we call *moderation-of process* designs). To avoid the sort of confusion between mediation and moderation noted by Baron and Kenny (1986), we feel it is important not only to distinguish between mediation and moderation at the statistical level, but also to distinguish between theoretical and statistical understandings of mediation.

Table 1
*Recommendations for Experimental Designs Based on Ease of Manipulating and Measuring the Proposed Mediator*

| Ease of measuring proposed process | Ease of manipulating proposed process: Easy | Ease of manipulating proposed process: Hard |
|---|---|---|
| Easy | Experimental-causal-chain design | Measurement-of-mediation design |
| Hard | Moderation-of-process design | No design is likely to work |

process leads to the effect of interest. Sometimes the very act of trying to measure the mediating process can either prevent the process from occurring or lead to the process occurring. For example, in studies that manipulate whether a concept is primed, measurement of the activation of the concept can in some instances prime the concept for all the participants and prevent the observation of an intended effect. In such instances, the measurement-of-mediation design is not a useful solution.

A second drawback is that in measurement-of-mediation designs the evidence that the mediator accounts for the relation between the independent variable and the dependent variable is essentially correlational. Whenever evidence for mediation is obtained in a measurement-of-mediation design, one must always consider whether there is a third variable that accounts for the observed relations. It is always possible that evidence of mediation is obtained spuriously because of the relation between the measured variable and the true psychological process.

A third drawback is primarily pragmatic. Measurement-of-mediation designs using the Baron and Kenny (1986) type of analysis often suffer from low power.[2] Therefore, it is often easier to demonstrate that an independent variable causes an effect than it is to demonstrate that a proposed mediator is affected by the independent variable and accounts for the effect of interest.

A fourth drawback of measurement-of-mediation designs can occur when researchers select measures of psychological process and outcome measures that are not theoretically distinct (i.e., the measures actually measure the same concept). In such instances, it is easy to obtain results that meet all the qualifications of a Baron and Kenny (1986) type of analysis, but such results do not provide evidence of psychological process. All such results demonstrate is that a manipulation affects two measures of the same outcome variable and that these two measures correlate with one another. It is therefore crucial, when evaluating measurement-of-mediation designs, to not only consider whether there is statistical evidence of mediation, but also whether there is discriminant validity between the measure of process and the outcome measure.

A fifth challenge of measurement-of-mediation designs is that they contain elements of both classic experimental designs with random assignment and correlational designs in which meeting the assumptions of multiple regression-based analyses is required. Although it is certainly a strength that measurement-of-mediation designs bring together both of these types of analyses, meeting the assumptions required for a multiple regression-based analysis can be difficult at times. Standard texts (e.g., Kmenta, 1997; Cohen, Cohen, West, & Aiken, 2003) highlight these assumptions, and care must be taken that these assumptions are met.

The final concern about measurement-of-mediation designs is that one must consider whether the independent variable of interest interacts with the proposed mediator. Many treatments of mediational analysis (e.g., Baron & Kenny, 1986; Judd, Kenny, & McClelland, 2001; Kenny, Kashy, & Bolger, 1998; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Wegener & Fabrigar, 2000) have discussed this issue, but we have found that it is useful to carefully consider its implication in the experimental context. Specifically we have found that in measurement-of-mediation designs it is useful to predict how each condition will affect the relation between the mediating variable and the dependent variable.

This approach is perhaps most clear when thinking about predicting the relation between mediators and dependent variables in control conditions. In experimental conditions, researchers usually make two predictions: first, that experimental manipulations will create psychological states and, second, that these psychological states will predict the dependent variables. In contrast, in control conditions, psychological states are not created, but baseline levels of these states may or may not be present. The crucial questions for researchers is whether baseline levels of these states are present and whether these states will predict the dependent variables in the same way that the manipulated levels of these states predict the dependent variables. Sometimes it is reasonable to expect such a relation; sometimes it is not.

For example, Son Hing, Li, and Zanna (2002) had aversive racists (i.e., those low in explicit, but high in implicit, prejudice against Asians) write an antiracism essay, then half of the participants wrote about two incidents in which they reacted more negatively to an Asian than they thought they should have (the hypocrisy condition) and the other half of the participants wrote nothing (the control condition). Son Hing et al. (2002) predicted that the hypocrisy manipulation would induce feelings of guilt and that this guilt would in turn predict bending over backward to avoid discrimination. This is indeed what they found. But what about the control condition? Should feelings of guilt in the control condition—that could just as easily result from cheating on one's boyfriend or girlfriend as from hypocrisy about racism—predict discrimination against Asians? As one might expect guilt in the control condition was unrelated to discrimination (cf. Zanna, 2004).

Statistically, this lack of correlation between guilt and discrimination in the control condition meant that a model of simple mediation as proposed by Baron and Kenny (1986) was not appropriate, but theoretically the results provided strong support for the proposed psychological process. In our experience, measurement-of-mediation designs often demonstrate more than

[2] Recently techniques have been developed that increase the power of measurement-of-mediation designs, such as Shrout and Bolger's (2002) bootstrapping method (see also Bollen and Stine, 1990). Such techniques may help overcome the problem of reduced power in measurement-of-mediation designs.

simple mediation, thus complicating the statistical analyses. (See Muller, Judd, and Yzerbyt in this volume on pp. 852–863 for a detailed description of these analyses.) We have learned, however, that these sort of statistical complications need to be weighed carefully against the theory that is being tested to understand their relevance. In this example, the result is not what is usually referred to as moderated mediation—in which mediation of an effect (e.g., hypocrisy produces guilt which, in turn, leads to bending over backward to not be prejudiced) occurs under one level of a variable (e.g., for younger participants), but not under another level of the variable (e.g., for older participants)—but rather is what Harackiewicz, Abrahams, and Wageman (1987) refer to as interactional mediation (see also Judd & Kenny, 1981). Setting the terminology aside, however, what we want to emphasize is that when conducting mediational analyses one must consider the relation of the mediator and the dependent variable in each condition—and determine whether the independent variable influences the relation between the mediator and the outcome.

What is crucial is not the form of the statistical analysis but rather whether the analysis supports the theoretically proposed account of psychological process. Perhaps another example will clarify this point. Davies and his colleagues (Davies, Spencer, Quinn, & Gerhardstein, 2002, and Davies, Spencer, & Steele, 2005) have conducted several experiments in which they have examined the causal theoretical argument that for women when stereotype threat is high watching stereotypic commercials (A) would lead to activation of gender stereotypes (B), which in turn would lead to behaviors related to stereotype threat (C). In testing this causal theoretical argument they used different types of control groups in the two series of studies. Operationally Davies et al. (2002) manipulated stereotype threat by exposing women to stereotypic or counterstereotypic commercials before they took a math test that was described as nondiagnostic. The control condition in these studies (i.e., the low stereotype threat condition) was when the women were exposed to the counterstereotypic commercials. In contrast, Davies et al. (2005) manipulated stereotype threat by describing a leadership task as being gender neutral or not. The control condition in these studies (i.e., the low stereotype threat condition) was when the task was described as gender neutral. Because the mediator was related to the outcome in the first set of controls, but not in the second set, standard mediation analyses were appropriate only in the first set of studies. Nevertheless, each set of studies provided strong support for the causal theoretical argument.

Specifically, Davies et al. (2002) had men and women watch either stereotypic (high stereotype threat condition) or counterstereotypic (low stereotype threat control condition) TV commercials of women, measured their activation of gender stereotypes, and then measured their math performance on a nondiagnostic math test. They found that women who watched the stereotypic commercials showed stronger activation of the gender stereotype and worse performance on the math test than did women who watched the counterstereotypic commercials. In addition, in a Baron and Kenny (1986) type of analysis, they found that for women activation of the stereotype mediated the effect of the commercials on women's math performance.[3] Thus, the experiment found evidence of the theoretical causal argument with a relatively straightforward measurement-of-mediation analysis.

But things are not always so straightforward. In a similar set of experiments Davies et al. (2005) had men and women watch either the same stereotypic or neutral commercials, measured their stereotype activation, and then measured their desire to be a leader in a subsequent task. When women watched the stereotypic commercials in the high stereotype threat condition (i.e., when the leadership task was not characterized regarding gender and, thus, participants were allowed to construe the task as masculine), women activated the gender stereotype, and this activation predicted a reduced preference to be a leader. This pattern of results paralleled the findings in the Davies et al. (2002) studies—there was a significant within-cell correlation between activation of the stereotype and the outcome measure. When women watched the stereotypic commercials in the low stereotype threat control condition (i.e., when the leadership task was characterized as gender neutral and, thus, participants were not allowed to construe the task as gender related), however, they activated the gender stereotype, but this activation did not predict their desire to be a leader. This pattern of the results did not parallel the Davies et al. (2002) studies—that is, in this condition there was no within-cell correlation between activation of the stereotype and the outcome measure. Therefore, the pattern led to a more complicated statistical analysis. Thus, the experiment found evidence of the theoretical causal argument with a less-than-straightforward statistical analysis (i.e., the lack of simple mediation in a Baron & Kenny, 1986, type of analysis).

Note two things about these experiments: First, both studies provided support for the theoretically proposed model of psychological process despite the fact that they did not show the same pattern of results. Davies et al. (2002) suggested that for women under high stereotype threat the effect of watching stereotypic commercials on math performance is mediated by stereotypic activation, and Davies et al. (2005) suggested that for women under high stereotype threat the effect of watching commercials on the desire to be a leader is mediated by stereotype activation. What is different between the two studies is what is occurring in the control conditions (i.e., when stereotype threat is low). In Davies et al. (2002) women who activated the stereotype under low stereotype threat (i.e., when told the test was nondiagnostic and when exposed to counterstereotypic commercials) did poorly on the math test. In Davies et al. (2005), however, women who activated the stereotype under low stereotype threat (i.e., when told the task was gender neutral) did not show a decreased desire to be a leader. The different pattern of results in these studies may suggest something unanticipated about the different nature of the two low stereotype threat control conditions,[4] but the different pattern really says nothing about the theoretically proposed psychological processes that both studies were designed to test. Both

[3] Note that if men are included in this mediational analysis, moderated mediation is actually obtained. Both men and women activate the gender stereotype when they see stereotypic commercials, but this activation of the stereotype is only related to women's performance on the math test and not to men's.

[4] One explanation for the discrepancy in the results between the two studies is that when stereotype threat is low because a test is seen as nondiagnostic, then activation of the stereotype in this situation still produces some level of stereotype threat that predicts performance. When stereotype threat is low because a task is described as gender neutral, however, the link between stereotype activation and the task is severed and activation of the stereotype no longer predicts responses to the task.

studies suggest that for women activating the stereotype leads to stereotype-driven behavior when stereotype threat is high.

The second major point from the above example, however, is that understanding the correlations between the mediator and the dependent variable in each condition is crucial to conducting the analyses properly. Had Davies et al. (2005) examined simple mediation across conditions under the assumption that the mediator would show the same relation to the dependent variable in all conditions, then they would have found no evidence for the proposed psychological process, although such evidence existed. In our experience, more complicated statistical analyses are often required. Fortunately the Muller et al. (2005) article in this volume provides a useful framework for conducting such analyses.

Given the complexity of data analysis in measurement-of-mediation designs, we have found that it is critical to think about the within-cell correlations between the mediator and the dependent variable in each cell of the design. Will they be the same? Will they be different? What is the theoretical basis for the prediction of the relation in each cell of the design? We find thinking about these questions to be a useful strategy in interpreting measurement-of-mediation designs. As a simple rule of thumb, if you expect the relation between chronic levels of the mediator and the dependent variable in the control or baseline condition of your experiment to be similar to the relation of these variables in the experimental condition, then a standard Baron and Kenny (1986) simple mediation analysis may well be appropriate. If, however, you do not expect a relation between chronic levels of the mediator and the dependent variable, then a standard Baron and Kenny (1986) simple mediation analysis is inappropriate and unlikely to work. In such instances, the failure of such an analysis should not be seen as evidence against one's causal theoretical argument.

We have argued that measurement-of-mediation designs are most likely to be useful when a psychological process is easily measured but hard to manipulate. Even in these situations, however, the drawbacks and challenges of these designs often make them hard to utilize. Although we think that such designs have their place, we are concerned that their elevation as the gold standard to establishing a theoretical argument for psychological process may have needlessly thwarted progress in the field. In our view, these designs should be used sparingly—only when other easier to implement designs are not possible.

## Discussion

In testing for the evidence of psychological processes, we have emphasized that what should be most important is examining the theoretical arguments that are being made rather than focusing on the specific types of analyses that are being conducted. A hypothesis about psychological process is at base the refinement of a causal argument in which an intervening variable is seen as the effect of an independent variable and the cause of a dependent variable. In this article, we have tried to describe how a broader set of experimental designs than is often recognized can allow for the examination of such mediating hypotheses. Specifically, we argue that designs that utilize several studies to examine a psychological process as both an effect of the proposed independent variable and as a cause of the proposed dependent variable—what we call experimental-causal-chain designs—can often provide the most compelling case for a theoretical account of a psychological process. If the process can be both easily measured and manipulated, we feel this is usually the optimal strategy.

If a psychological process can be easily manipulated but is difficult to measure, then we recommend a design that examines this psychological process by manipulating the process to moderate the relation between the independent variable and the dependent variable—what we call a moderation-of-process design. Such designs (e.g., when cognitive load interferes with controlled processing of information) can provide compelling evidence of a proposed psychological process when there is compelling evidence that the operational manipulation of the process does indeed have the proposed theoretical effect and when alternative explanations for the effect of the manipulation on the relation between the independent and dependent variable have been ruled out.

In recommending these designs, we seek to emphasize the power of experiments in demonstrating causality. Experiments are effective in establishing cause and effect, and the specific case of establishing a mediator as the effect of an independent variable and the cause of a dependent variable is no different.

Despite our promotion of experimental designs to examine mediation, we recognize that in some situations it is difficult to manipulate the theoretically proposed psychological process. In such cases, we recommend that people consider a design in which they measure the proposed psychological process after the independent variable has been manipulated and examine whether this process can account for the effect of the independent variable on the dependent variable using regression-based analyses as suggested by Baron and Kenny (1986)—what we call measurement-of-mediation designs. In recommending these designs, we offer a number of cautions and suggestions to improve their effectiveness. As a matter of pragmatism, we feel these designs are harder to implement than are standard experimental designs.

We want to emphasize, however, that just because these designs are harder to implement does not mean they should be seen as more compelling accounts of a theoretical process. A design ideally makes it easy to demonstrate an important effect. Unfortunately, in our view, measurement-of-mediation designs often make finding evidence for psychological processes more difficult. It is of course tempting to be impressed when a finding that is difficult to obtain is observed, but when evaluating a theory what is critical is not how difficult it is to obtain the evidence, but rather how strongly the evidence supports the theory. Although we should be rightly impressed with the skill of our colleagues when they are able to develop a measurement-of-mediation design that is able to provide evidence for a causal theoretical argument, we should not confuse such admiration with evidence for their theory.

We should be explicit that our analysis is restricted to experimental contexts. In nonexperimental situations, the experimental strategies that we have suggested by definition are not an option. In such nonexperimental settings, we believe that mediational analyses as suggested by Baron and Kenny (1986) are often the best option for examining psychological process. We should also be explicit that we are not in any way criticizing the original Baron and Kenny (1986) paper or elaborations of this method of doing analyses (Judd, Kenny, & McClelland, 2001; Kenny, Kashy, & Bolger, 1998; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Shrout & Bolger, 2002). As we have said, this perspective has been cogently argued and addressed an important need in the field. We do think, however, that measurement-of-mediation designs have been overused and the Baron and Kenny (1986) type of analysis often has been misapplied.

Finally, we believe that it is important to keep psychological processes in their proper perspective. In our view, any healthy science should have room for discovery and the development of new theories as well as explanation and the refinement of existing theories. Theoretical arguments about psychological processes by their very nature tend to emphasize explanation and the refinement of existing theories. In evaluating the theoretical contribution of theories, let us rightly value the place of psychological processes in social psychology, but let us also value the discovery of new phenomena and the development of new theories.

In addition to their theoretical contribution, the practical contribution of theoretical accounts of psychological processes should also be considered. At their worst, accounts of psychological processes (we will not even say theoretical accounts here) can become an infinite regress of ever-finer intervening causes and the worst sort of navel gazing. For example, some accounts of psychological process propose mediators that are conceptually very similar—perhaps indistinct—from either the independent variable or dependent variable being examined. It may be easier to find statistical evidence for such accounts, but their theoretical and practical significance is suspect. At their best, however, theoretical accounts of psychological process can provide important insights that allow us to intervene to make the world a better place. It is such psychological processes that should be valued, and a number of different experimental approaches should be seen as valid ways of testing such ideas. What we propose is that people consider a broad range of possible designs when examining psychological processes and select the design that is best suited to the particular problem being studied.

## References

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 115–140). Oxford: Blackwell.

Burnstein, E., & Vinokur, A. (1977). Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology, 13,* 315–332.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). All consuming images: How demeaning commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28,* 1615–1628.

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology, 88,* 276–287.

Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through negative evaluations of others. *Journal of Personality and Social Psychology, 73,* 31–44.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology, 58,* 203–210.

Harackiewicz, J. M., Abrahams, S., & Wageman, R. (1987). Performance evaluation and intrinsic motivation: The effects of evaluative focus, rewards, and achievement orientation. *Journal of Personality and Social Psychology, 53,* 1015–1023.

Judd, C. M., & Kenny, D. A. (1981). *Estimating the effects of social interventions.* Cambridge, England: Cambridge University Press.

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods, 6,* 115–134.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology.* (Vol. 1, 4th ed., pp. 233–265). New York: McGraw-Hill.

Kmenta, J. (1997). *Elements of econometrics* (2nd ed.). Ann Arbor, MI: University of Michigan Press.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7,* 83–104.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When mediation is moderated and moderation is mediated. *Journal of Personality and Social Psychology, 89,* 852–863.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to persuasion.* New York: Springer-Verlag.

Quiñones-Vidal, E., López-García, J. J., Peñaranda-Ortega, M., & Tortosa-Gil, F. (2004). The nature of social and personality psychology as reflected in *JPSP*, 1965–2000. *Journal of Personality and Social Psychology, 86,* 435–452.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422–445.

Sigall, H., & Mills, J. (1998). Measures of independent variables and mediators are useful in social psychology experiments: But are they necessary? *Personality and Social Psychology Review, 2,* 218–226.

Son Hing, L. S., Li, W., & Zanna, M. P. (2002). Inducing hypocrisy to reduce prejudicial responses among aversive racists. *Journal of Experimental Social Psychology, 38,* 71–78.

Storms, M. D., & Nisbett, R. E. (1970). Insomnia and the attribution process. *Journal of Personality and Social Psychology, 16,* 319–328.

Wegener, D. T., & Fabrigar, L. R. (2000). Analysis and design for nonexperimental data: Addressing causal and noncausal hypothesis. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 412–450). New York: Cambridge University Press.

Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology, 10,* 109–120.

Zanna, M. P. (2004). The naïve epistemology of a working social psychologist (or the working epistemology of a naïve social psychologist): The value of taking "temporary givens" seriously. *Personality and Social Psychology Review, 8,* 210–218.

Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attributional approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology, 29,* 703–709.

Received June 23, 2004
Revision received July 1, 2005
Accepted July 21, 2005 ■