Chapter twenty-five Mediation and Moderation

Charles M. Judd, Vincent Y. Yzerbyt and Dominique Muller

Our goal in this chapter is to provide an up-to-date and relatively comprehensive treatment of procedures for assessing mediation and moderation in social-personality psychology. Both of these processes enable researchers to ask questions of their data that extend the theoretical scope of inquiry beyond simply establishing some overall experimental effect or some simple relationship between two variables. That is, they both begin to enable researchers to arrive at a more comprehensive theoretical understanding of what produces an effect of interest by probing intricacies of that effect. As such, they are related but distinct analytic tools. They are related in the sense that they permit researchers to probe mechanisms underlying and limiting conditions for effects of interest. And yet, the questions they pose are fundamentally different, in ways that are often confused, and the underlying models are distinct.

Given their importance in developing a theoretical understanding of what produces an effect of interest, it is hardly surprising that the assessment of mediation and moderation is ubiquitous in social and personality psychology. As a result, the literature devoted to procedures for estimating and testing mediation and moderation is vast. While we cover what we consider to be the most important points in this literature, our intention is not to cover this literature exhaustively. Rather, our goal is to discuss basic analytic issues, complexities of interpretation and inference, and underlying assumptions and common pitfalls. Additionally, we provide citations to more in-depth and comprehensive treatments throughout.

The chapter is organized into four main sections. In the first short section we provide basic definitions of both mediation and moderation and illustrate the sort of theoretical questions that their assessment permits the researcher to address. Our emphasis here is on the theoretical definitions that underlie both mediation and moderation, rather than on the technical details of estimation and statistical inference. The second section of the chapter is devoted to a more in-depth treatment of mediation, including underlying assumptions, estimation, statistical inference, and power considerations. Coverage here includes both basic models assuming homogenous errors and more complex multilevel models that allow grouping and nonindependence of observations. The third section is devoted to a more in-depth treatment of moderation, again including underlying assumptions, estimation, statistical inference, and power considerations. And here too we discuss moderation analyses in the multilevel context, with nested nonindependent observations. In the final section we discuss the integration of these two processes, framed as either moderated mediation or mediated moderation. Again we discuss estimation issues for such models and theoretical interpretations and insights that they permit.

Defining Mediation and Moderation

In order to define mediation and moderation, we start with the presumption that research has established some relationship or effect of theoretical interest. For instance, a social psychologist may have conducted research to demonstrate that social projection – that is, people's tendency to consider that others have the same traits or show the same preferences as oneself – depends on others' group membership. Or a personality researcher may have explored ways in which a particular individual difference – say, extraversion – is related to the tendency to assume leadership roles in small-group settings.

Seldom, however, are researchers content with simply the demonstration of such a relationship or effect. To build a theoretical understanding of social behavior and individual differences more broadly, one must probe the mechanisms that underlie an effect and the limiting conditions for its occurrence. Understanding the mechanisms produces more refined assessments of what the effect really is and how it is produced. Understanding its limiting conditions informs the researcher about necessary and sufficient conditions for its occurrence. These two sorts of understandings – one of mechanisms and one of limiting conditions – are the concerns of mediation analyses and moderation analyses, respectively. That is, the goal of mediation assessment is to explore the underlying mechanisms responsible for an effect of interest, whereas the goal of moderation assessment is to explore the ways in which the magnitude of an effect of interest may depend on other variables.

While the questions addressed via the assessment of mediation and moderation are distinct, it is nevertheless the case that gaining knowledge of mechanisms and limiting conditions extends in similar ways one's theoretical understanding of an effect. If one really understands the mechanisms that produce an effect, then surely one gains insights into the necessary conditions to produce that effect. That is, if one understands the mechanisms, then it seems likely that one could turn off the effect by inhibiting those mechanisms. And if one really understands the conditions under which an effect is or is not produced, then surely one has gained some insight into the mechanisms responsible for an effect. So a full theoretical understanding of an effect of interest involves both understanding mechanisms (the question of mediation) and understanding limiting conditions (the question of moderation), and the knowledge gained from both of these assessments ultimately must converge.

Because of the fact that the understanding of mechanisms and the understanding of limiting conditions are theoretically intertwined and, in combination, give rise to a full theoretical understanding of the effect of interest, the theoretical questions asked by mediation and moderation procedures can be confusing. However, the analytic procedures for assessing mediation and moderation are different. The former set of procedures examines partial effects controlling for hypothesized mediators. The latter set of procedures examines interactions between the independent variable that produces the effect and some other moderating variable. The distinction in analytic procedures enforces the researcher to think clearly about whether he/she is probing mechanisms or limiting conditions.

Mediation

Basic Analytic Model

Suppose that a researcher wants to study the impact of an independent variable *X* on a dependent variable *Y*. Imagine that the independent variable has two levels – a treatment condition and a control condition – and that, in order to permit stronger causal inference, participants have been randomly assigned to one or the other of these conditions. In this context, the total linear effect *X* on *Y* is estimated by the slope in the following linear model[1]:

$$Y_i = b_{01} + cX_i + e_{1i}$$

The effect of *X* in this model is represented by the diagram in the top half of Figure 25.1.

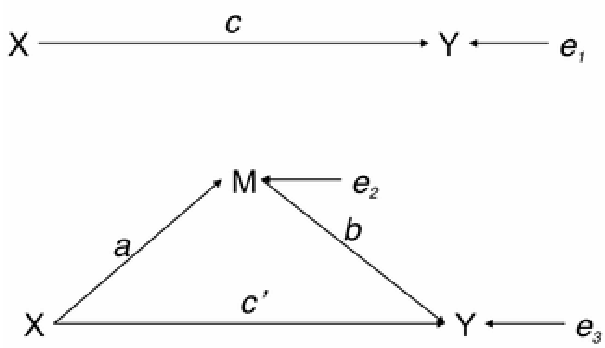

Figure 25.1. Basic mediation model.

In mediational analyses, the researcher is interested in finding the mechanism responsible for this *X* – *Y* relationship (Baron & Kenny, 1986; James & Brett, 1984; Judd & Kenny, 1981). Accordingly, the researcher generates hypotheses about one or more third variables that may be partially responsible for the observed total effect, *c*. The question will then be: Does part of the total effect go through the third variable, often called a mediator or an intervening variable?

To conduct the mediational analysis, one estimates the following two models:

$$M_i = b_{02} + aX_i + e_{2i}$$
$$Y_i = b_{03} + c'X_i + bM_i + e_{3i}$$

In the first of these models, *a* is the simple effect of *X* on the mediator. In the second, *b* is the partial effect of the mediator, controlling for *X*, and *c'* is the partial effect of *X* controlling for the mediator. These models are represented in the diagram at the bottom of Figure 25.1.

The fundamental equation of mediation expresses the total effect *c* as a function of the coefficients estimated in these two mediational models:[2]

$$c = ab + c'$$

What this equality tells us is that the total effect of *X* on *Y*, *c*, can be broken into two components, *a***b* and *c*′. The first of these components, *a***b*, is the indirect effect of *X* on *Y* via the mediator. This is the portion of the total effect that corresponds to the mediation via *M*. The second of these components, *c*′, is the residual direct effect of *X* on *Y* controlling for or "over and above" the mediator.

It should be noted that the term "direct" must be understood in relative terms, given that there may be other mediators that potentially explain this residual direct effect (Rucker, Preacher, Tormala, & Petty, 2011). Hence, in the case of two mediators *M1* and *M2*, the direct effect would be the residual effect of *X* on *Y* not explained by either *M1* or *M2*.

Let us illustrate mediation analysis as well as the underlying models by presenting a concrete example. In a social comparison study, pairs of participants, one of whom was in fact a confederate, performed an attentional task twice (Muller & Butera, 2007, Study 5). After the first round, participants were randomly given bogus feedback: Whereas half of them heard that they had outperformed the confederate (i.e., the downward comparison condition; DC), the other heard they had been outperformed by the confederate (i.e., the upward comparison condition; UC). The self-evaluation threat hypothesis suggests that the UC participants should feel more threatened in their self-evaluation than the DC participants. As a result, they should have fewer attentional resources left in order to process peripheral cues when completing the task. Because peripheral cues could either be selected to help or to hurt participants when dealing with the task, the difference between these two types of cues (called a cuing effect) should be reduced for UC participants. In their study, Muller and Butera did not measure the mediator (i.e., self-evaluation threat) but, in line with their hypotheses, they found a reduced cuing effect among UC than among DC participants.

Imagine now that we conduct a study to examine the hypothesis that self-evaluation mediates the impact of social comparison on participants' attentional resources (the data and SAS codes for this example are available at http://www.psp.ucl.ac.be/mediation/medmod/). To this end, we measure self-evaluation right after participants receive the bogus feedback but before they proceed to the second round of the attentional task. As mentioned earlier, the first model allows testing the effect of the independent variable (i.e., social comparison; contrast coded: DC = –0.5 and UC = 0.5) on the dependent variable (i.e., the cuing effect). This analysis reveals a larger cuing effect among DC (M = 69.33) than UC participants (M = 51.65). Given the coding we use, this translates into a significant negative slope, c = –17.67, $t(38) = 2.91$, $p < .01$. In the second model, we test the impact of the independent variable on the mediator (i.e., self-evaluation threat). This analysis reveals a smaller self-evaluation threat in DC (M = 4.40) than in UC (M = 6.32). Accordingly, this translates into a significant positive slope, a = 1.92, $t(38) = 3.62$, $p < .01$. In the last model, we regress the cuing effect on both the independent variable and the mediator. In line with our mediational hypothesis, this analysis reveals a significant slope for the mediator, b = –7.88, $t(37) = 5.80$, $p < .01$, such that (controlling for the independent variable) the higher the self-evaluation threat, the lower the cuing effect. This analysis shows that once we control for the mediator, the effect of the independent variable is no longer significant, c' = –2.51, $t(37) = 0.49$, $p = .63$. Finally, in line with the fundamental equation presented earlier (notwithstanding rounding errors), we note that the total effect c equals $a*b + c'$, as −17.67 = (1.92* –7.88) + (–2.51).

Both the analytical model and the preceding example take for granted that the researcher wants to investigate the mechanism underlying an observed experimental effect. Obviously, the initial step is thus to first establish that such an effect exists. Demonstrating this requires that the experiment have sufficient power to find the overall or total effect. Traditionally, the definition of mediation has taken for granted that there is a significant total effect, and the goal of mediation is then to at least partially account for the process that produces that effect (Baron & Kenny, 1986; Judd & Kenny, 1981).

In recent years, there has been increasing skepticism about the view that the impact of *X* on *Y* must be demonstrated before turning to a closer examination of the potential mediating role of a third variable (e.g., Shrout & Bolger, 2002). Relatedly, a similar confusion has surfaced regarding the exact conditions for mediation and whether or not a variable that "suppresses" a total effect should be called a mediator. In the following, we hope to clarify these issues.

One helpful way to think about this is to consider three key features in any situation targeted by a mediational hypothesis. A first feature concerns the presence or absence of a significant c – that is, the total effect of *X* on *Y*. A number of reasons may explain why c is not found in the particular data set examined by the researcher. It may be that such an effect simply does not exist. Alternatively, it may exist, but the experiment may not have had sufficient power to detect it. A second feature has to do with c' – that is, the direct residual effect. If in fact the mediator is playing some causal role in affecting *Y*, then c' should have a different value from c. Finally, the third feature is the indirect effect, $a*b$. When significant, this product points to the existence of a significant causal[3] flow between *X* and *Y* via the intervening variable, *M*.

A proper consideration of these three features allows us to define in unambiguous terms what for us corresponds to mediation, suppression, and the mere presence of an indirect effect.[4] An indirect effect can be said to exist whenever $a*b$ is significant, regardless of the values of c and c'. Both mediation and suppression presume that there is a significant indirect effect, but they imply additional considerations concerning the magnitudes of c and c'. Mediation for us implies the additional assumption that there is a significant total treatment effect to be explained by the mediational process, i.e., $|c| > 0$, and that this total effect, c, is larger in absolute value than c'. Finally, suppression in the context of a mediational model exists when there is a significant indirect effect and a significant c', a residual direct effect, that is larger in absolute value than the total effect, *c*. Suppression means that the intervening variable, when not controlled, is in fact dampening the total effect and that the inclusion of *M* in the model allows for the direct effect to be more fully revealed.

It should be noted that this discussion focuses on situations in which only one mediator is examined. Matters get somewhat more complex when several mediators are examined. For instance, a third variable may be a suppressor variable and its inclusion in the model could actually reveal the existence not only of a direct effect but also of an indirect effect involving another intervening variable (Rucker et al., 2011).

Assumptions

As with any analysis, mediation analysis with unmanipulated mediators entails a number of important assumptions. A first assumption concerns the requirement that the relations among variables be linear. Of course, nonlinear transformations can be used in the analysis to model nonlinear relations. A second assumption is that the variables are measured both reliably and validly. A third assumption is that the errors or residuals in any one model are independent of each other or, equivalently, that there are no hidden nestings in the data that give rise to dependence. And a fourth and crucial assumption is that the aforementioned models have been correctly specified and that there are no correlated omitted variables that ought to be included in them. This assumption can be equivalently stated as the assumption that the errors or residuals in these models are uncorrelated with the predictor variables included in the models. We will discuss both the second and third assumptions at a later point in the chapter. For now we focus on the fourth assumption because we are convinced that in many applications of mediation analyses it is violated, with serious consequences.

All too often, from our point of view, one finds “mediational” analyses reported using cross-sectional data collected by measuring three variables, *X*, *M*, and *Y*, at roughly the same time. Even assuming no measurement errors, in the absence of any further information, the causal possibilities for why these three variables are related to each other are given by all the straight arrows (representing potential causal effects) and curved double-headed arrows (representing simple covariances induced by omitted variables) in Figure 25.2. This model obviously includes the possibility that *M* mediates the *X:Y* relationship: There is a straight arrow from *X* to *M* and another straight arrow from *M* to *Y*. But there are also reverse effects that may be responsible for the total covariation observed in the data. For instance, the effect of *X* on *M* may result from the impact of *X* on *Y*, which in turn affects *M*. And finally there are also omitted variables that are responsible for the covariation in the errors (the errors are that part of each variable not explained by the direct causal effects to it). If the results of mediational analyses are to provide unbiased estimates of true causal effects, then all of the arrows in Figure 25.2 with the exception of those posited by the mediating process (which we have labeled *r*, *s*, and *t*) must be zero. In other words, *a* will not equal *r* unless there is no reverse causal effect of *M* on *X* and unless there are no omitted common causes of both *M* and *X*. And the same holds for the other effects estimated in a mediational analysis.

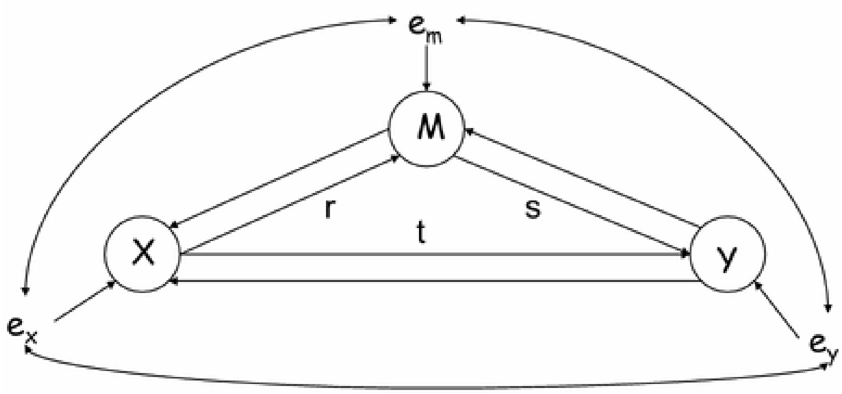

Figure 25.2. Three variable model showing effects responsible for total covariations.

At the beginning of this section we made the assumption that *X* was an experimentally manipulated independent variable, meaning that participants had been randomly assigned to its levels. The question is now what this buys us in terms of eliminating some of the effects and covariances of Figure 25.2 and thereby improving causal inferences from mediational analyses. With an experimental manipulation of *X*, the causal possibilities are contained in Figure 25.3. To be sure, many causal possibilities have been eliminated, but it is still the case that there are multiple reasons why the mediator, *M*, and the outcome variable, *Y*, covary. And if there is anything other than a direct effect of *M* on *Y*, then the mediational estimation will be biased (with bias in both the indirect effect and the residual direct effect).

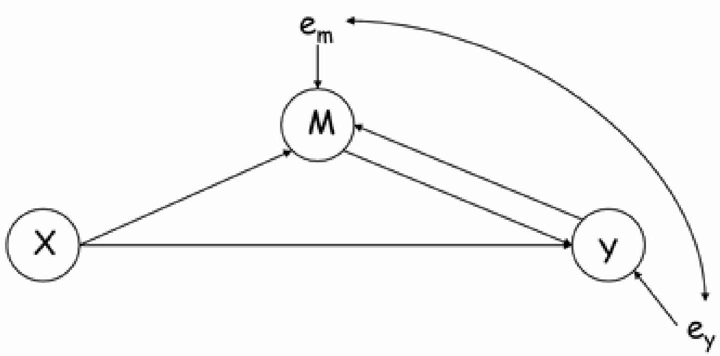

Figure 25.3. Three variable causal model with *X* manipulated.

What this means is that even with an experimental manipulation of the independent variable, *X*, mediational analyses will yield biased effects unless there is no potential of *Y* causing *M*, and unless there are no omitted common causes of both *M* and *Y*. We suspect that in most mediational analyses reported in the literature, even with experimental manipulations of *X*, the magnitude of the indirect effect via the mediator is substantially overestimated because the mediator and the outcome share omitted common causes. In a great many studies, the outcome and the mediator end up being measured by means of questionnaires, allowing for the intrusion of shared method variance. One of the very early treatments of mediation contained the following warning: The outlined analyses are "likely to yield biased estimates of the causal parameters…*even when a randomized experimental research design has been used*" (Judd & Kenny, 1981, p. 607, emphasis in the original). Unfortunately, this warning has gone largely unheeded.

MacKinnon (2008) summarizes additional considerations for the single mediator model. Among these, a crucial assumption is temporal precedence, because a mediational model ultimately refers to a causal sequence that must take place across time. As a matter of fact, the mediator model assumes that the treatment variable, *X*, comes before the mediator, *M*, which itself comes before the dependent variable, *Y*. This renders any mediational conclusions based on cross-sectional data highly problematic. The problem more often concerns the ordering of *M* and *Y* than the sequence involving *X*. Related to the issue of temporal precedence are two considerations, namely the level of the mediational chain and the measurement timing. The first concerns the specific steps that are selected for measurement in what may be a rather long and intricate causal chain. Depending on the focus of the researcher, the window used to examine the underlying causal chain may vary widely. The second is related to the correspondence between the timing of the measurement of the mediator and the dependent variable, on the one hand, and the true timing of the changes in the phenomena under examination, on the other. In many instances, changes in the mediator or in the outcome can occur long after the independent variable has been manipulated.

In Figure 25.4 we include a plausible causal model in the circumstance where *X* is an experimental manipulation and both *M* and *Y* are measured at two times points: time 1 at the same time that *X* is manipulated and time 2, somewhat later, when the effect of the treatment is thought to have been revealed. Again the mediational indirect effect is the effect of *X* on *M2* times the direct effect of *M2* on *Y2*. Even with such longitudinal data, this indirect effect will be estimated with bias if there is a reverse effect from *Y2* on *M2* or if there are omitted third variables responsible for the relationship between *M2* and *Y2*. This latter threat is reduced in magnitude somewhat because of the fact that earlier values of both variables are controlled, that is, *M1* and *Y1*. Assuming that other causal effects on these variables are unchanging over time, omitted and unchanging common causes will be effectively controlled by such longitudinal models (in the absence of measurement error).

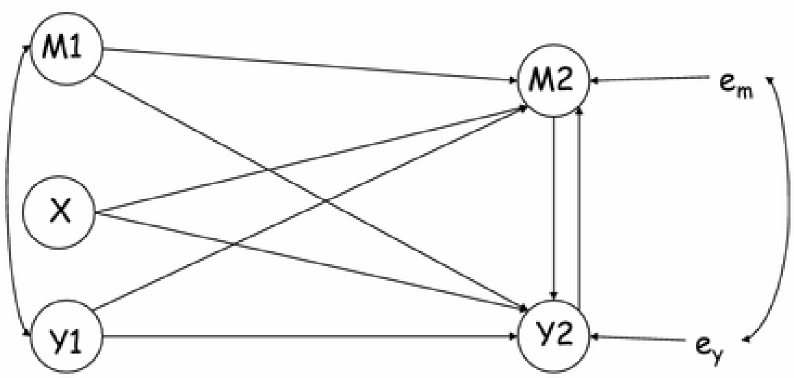

Figure 25.4. Causal possibilities in two-wave longitudinal mediation model.

Estimating and Testing Indirect Effects

So far, we provided the three basic equations underlying mediation analyses. These equations are sufficient to estimate and test the individual slopes in the estimated mediation models (although we later address limitations with these in the presence of measurement error). One important question remains how one should test the underlying indirect effect. There are basically three general approaches: the causal steps, the difference in coefficients, and the product of coefficients (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002).

First, one can test the indirect effect by estimating *a* and *b* and testing them individually against zero. The logic here is that if both steps of the indirect effect are significant, it means that the indirect effect is itself significant. In other words, if the two components of a product are significant, the product itself is significant.[5] Such a test has sometimes been referred to as the *a* and *b* joint significance test (Cohen & Cohen, 1983; Fritz, Taylor, & MacKinnon, 2012; MacKinnon et al., 2002).

Second, one can test the indirect effect by estimating and testing the difference between the coefficients c and c'. As we have seen, this test amounts to a test of whether the total X effect is different from the residual direct effect controlling for M. Because of the equivalence of $c - c'$ and $a*b$, this is conceptually similar to asking whether X indirectly influences Y via M.

Third, one can test the indirect effect by directly estimating and testing the product $a*b$. There are three main strategies for doing so. The first consists in testing the $a*b$ product by dividing it by an estimate of the standard error of a product of the two slopes. Baron and Kenny (1986) suggested computing this standard error by using a formula derived by Sobel (1982). The result of this computation is then compared with a normal distribution. As it turns out, this strategy is problematic because the product of two normally distributed variables does not typically have a normal distribution (Bollen & Stine, 1990; MacKinnon et al., 2002). A second strategy makes no assumption regarding the normality of the product distribution and relies on bootstrapping, a resampling method that consists of approximating the population distribution by sampling *with replacement* from the observed sample (Shrout & Bolger, 2002). The simplest bootstrapping technique, the percentile bootstrap, simply uses this bootstrap sampling distribution to provide a 95% confidence interval. The test of the indirect effect is statistically significant if this interval does not include 0. More elaborated bootstrapping techniques – the accelerated bias-corrected bootstrap and the bias-corrected bootstrap – involve corrections for potential biases in $a*b$ estimation and its standard deviation. The third strategy solves the normality issue by using numerical integration to estimate the distribution of this product. MacKinnon and colleagues developed a program called PRODCLIN for doing this (e.g., Fritz et al., 2012; MacKinnon, Fritz, Williams, & Lockwood, 2007). The program enables one to derive a 95% asymmetric confidence interval that can be used to estimate and test $a*b$, again with the indirect effect being statistically significant if 0 is not found in this confidence interval (for more information, consult http://quantpsy.org/sobel/sobel.htm).

A fair amount of the mediation literature has been devoted to comparing these indirect effect tests (e.g., Fritz & MacKinnon, 2007; Fritz et al., 2012; MacKinnon et al., 2002). What this work shows is that the most powerful tests are probably the bias-corrected bootstrap tests (e.g., Fritz & MacKinnon, 2007). The major problem, however, is that this increase in power comes with a price in terms of

Type 1 error (Cheung, 2009; Fritz et al., 2012; MacKinnon, Lockwood, & Williams, 2004). The increase in Type 1 error, although not dramatic, is especially found when either *a* or *b* are 0 and their counterparts are large (i.e., when $a = 0$ and *b* is large or when *a* is large and $b = 0$; Fritz et al., 2012). Our own simulations show that the only test that does not suffer from this Type 1 error issue is the *a* and *b* joint significance test. This is of interest also because previous work showed that this test is statistically as powerful as the percentile bootstrap and the numerical integration tests (Fritz & MacKinnon, 2007; MacKinnon et al., 2002). The only remaining downside to this approach could be that this test does not provide directly a confidence interval for the indirect effect (MacKinnon et al., 2002). Although having such a confidence interval should not be seen as mandatory, we believe that when necessary one can still estimate indirect effect confidence intervals from either the percentile bootstrap or the numerical integration. Interestingly, MacKinnon and colleagues now suggest that *a* and *b* should be tested in addition to using *a***b* tests (Fritz et al., 2012). Obviously, one additional benefit of the *a* and *b* joint significance test, compared to all the other indirect effect tests, is that one can be reasonably confident that each step of the causal path really is significant. As a matter of fact, some data sets may lead to a significant *a***b* test while either the test of *a* or the test of *b* fails to reach significance, calling for some caution in interpretation and for possible replication.

To illustrate the test of the indirect effect, we can go back to the example inspired by the Muller and Butera's (2007) study. If, as we suggest, one relies on the joint significant test, the data reveal that both *a* and *b* were significant. Clearly, this test ensures that both *a* and *b* are reliable effects. In order to conduct the *a***b* test, we rely on Preacher and Hayes's (2008) macro. Specifically, the percentile bootstrap (using 5,000 resamples), for instance, gives a $CI_{95\%}$ of –27.94 to – 5.98. Because this confidence interval does not include 0, we can safely conclude that the indirect effect is significant.

Observed versus Latent Variable Models

More often than not, psychologists do not have direct access to their theoretical variables, but must rely on observed variables instead (e.g., Sigall & Mills, 1998). These observed variables necessarily contain errors of measurement – that part of an observed variable's variance that is not explained by the underlying theoretical construct. The presence of measurement error, as we mentioned earlier, can lead to substantial bias in the estimate of slopes in the mediational models. To reduce measurement errors, social and personality psychologists often measure their theoretical variables with more than one indicator. Researchers are then faced with two main options given that they want to reduce bias from measurement error by using these multiple indicators. First, they can average the multiple indicators and use this summary score as their proxy, their observed variable, in a regression model: This is the observed variable approach for dealing with measurement error (e.g., Ledgerwood & Shrout, 2011). This approach reduces measurement error but does not completely eliminate it (the degree to which such a composite still contains errors of measurement is given by Cronbach's α; Judd & McClelland, 1998; Schmitt, 1996). Second, researchers can use structural equation modeling to adjust for measurement errors. These models examine relationships among latent variables, adjusted for measurement errors, to test the theoretical model: This is the latent variable approach to estimating mediation (e.g., Ledgerwood & Shrout, 2011). Whatever the chosen model (observed or latent variables), researchers can then proceed with the mediation analysis of their choice presented earlier (MacKinnon, 2008).

To address the pros and cons of these two approaches for dealing with measurement errors, one needs to distinguish between accuracy and precision. Accuracy has to do with how good the model is with respect to estimating the various parameters without the biasing effects of measurement error. In other words, does the model estimate the effects accurately? Precision has to do with the sensitivity of statistical tests of those effects. In other words, does it detect an effect that is in fact present (i.e., does it have a low Type 2 error rate)?

An accurate model is one that provides parameter estimates that are as close as possible to the true theoretical relationships. Latent variable models are better suited for this purpose because they correct for measurement errors. This leads to more accurate estimates. In general, measurement error attenuates bivariate relationships and latent variable models, with multiple indicators, eliminate this attenuation.

The issues, however, become more complex in the context of a three-variable mediational model. For instance, if X is manipulated (hence there is no measurement error in X) and M is measured with error, the b path will be underestimated in an observed variable model (Hoyle & Kenny, 1999). But importantly, given that $c' = c - ab$, it also means that c' will generally be overestimated (given that b is attenuated). Note that this is just an illustration because matters can become even more complex when X is also measured (and therefore has measurement error; see Ledgerwood & Shrout, 2011). The point remains that latent variable models are better equipped to estimate the true values of the different paths. In a nutshell, latent variable estimates are more accurate.

From the preceding discussion one might surmise that more accurate estimates should also be more precise. For instance, if the effects are underestimated in observed variable models, one would expect them to be significant less often. Paradoxically, this is not what happens. Indeed, Ledgerwood and Shrout (2011) showed that although latent variable models provide more accurate parameter estimates, they also come with larger standard errors, which translate into less precise tests: one will be more likely to conclude that an effect is not there when in fact it exists in the population. Simulations revealed that this is true for both the *b* and *a***b* tests. To quote Ledgerwood and Shrout "the latent (vs. observed) variable approach produced estimates that are more accurate but less powerful, especially as reliability decreases and as effect size increases." (p. 1182). This leads these authors to suggest (apart from wisely encouraging the investment in more reliable measures) a two-step strategy by which one tests the indirect effect with an observed variable model strategy and later estimates this indirect effect with a latent variable model. We concur with this recommendation, particularly when estimating the size of the indirect effect is crucial.

Multilevel mediation

A recent and very important extension of mediational analysis concerns those cases in which the data are collected at more than one level (Krull & MacKinnon, 2001; MacKinnon, 2008; Schoemann, Rhemtulla, & Little, Chapter 21 in this volume). For instance, individuals may be grouped in some way – in work teams, classes, or other naturally occurring groups. When this is the case, if some of the variables in a mediational model are measured at the level of the individuals within those groups, then the assumption that errors or residuals in the models are independent is likely to be violated. This nonindependence arises because observations within a group are likely to be more similar to each other, on average, than are observations between groups, leading to a positive intraclass correlation due to group. Such dependence can seriously bias statistical inference procedures in tests of mediational models (Schoemann et al., Chapter 21 in this volume).

It is important to bear in mind that there are many plausible groupings that can give rise to dependence in data. In addition to individuals being clustered into groups, observations may be clustered within people, with multiple observations from each person. Such a situation is quite common, with repeated-measures designs that are frequently used by social psychologists. When there are only a few well-defined levels of independent variables of interest that differentiate these repeated measures, procedures that are variations of analysis of covariance with repeated measures can be used to examine issues of mediation within-participants (Judd, Kenny, & McClelland, 2001). In other situations, with more complex designs and numerous observations taken within persons, such as in the case of everyday experience data sets (Reis, Gable, & Maniaci, Chapter 15 in this volume), more general multilevel models are required.

To illustrate multilevel mediational models, we turn to an example from Pleyers, Corneille, Yzerbyt, and Luminet (2009). These authors were interested in factors responsible for evaluative conditioning. More specifically, they wanted to know whether the availability of cognitive resources (the independent variable) plays a role in the emergence of evaluative conditioning (the dependent variable) via its impact on contingency awareness (the mediating variable). Each participant was exposed to several unfamiliar consumption products (conditioned stimuli) that were consistently paired with one of a series of pictures known to elicit either negative or positive affective responses (unconditioned stimuli). All participants wore headphones during the presentation, over which half of them heard music while the remaining half heard numbers and were made cognitively busy by having to perform an auditory two-back task. Specifically, participants were instructed to press the spacebar as quickly as possible when they heard a number that was identical to a number they had heard "two places before" (for instance, if they heard the number "7" and before that they heard a "3" and before that a "7"). The authors then checked the extent to which participants evaluated each product in line with the valence of its associated picture and whether they were able to correctly associate each product with its specific picture. The independent variable, cognitive resources, thus varied between participants. However, both the mediator (contingency awareness) and the evaluative conditioning outcome (whether the specific product was evaluated congruently with the unconditioned stimulus with which it was paired) varied within participants. In short, this study examined how a level-2 variable (between participants) influences a level-1 variable via a level-1 mediator (within participants). Such a design is formally referred to as a 2–1–1 mediational design (Krull & MacKinnon, 2001).

A proper analysis of such data first requires ascertaining the impact of the level-2 manipulation on the level-1 outcome. This can be done with the following equations:

Level 1: $Y_{ij} = d_{0j} + e_{ij}$

Level 2: $d_{0j} = p_{00} + cX_j + u_{0j}$

In the first of these models, Y_{ij} is the extent to which the evaluation of the *i*th conditioned stimulus for the *j*th participant is congruent with its paired unconditioned stimulus (higher numbers mean a stronger evaluation in line with the affective response elicited by the unconditioned stimulus). The level-1 model is in essence estimated for each participant and, accordingly, the estimated intercept, d_{0j} represents the mean degree of evaluative conditioning for each participant and e_{ij} is the variation in that conditioning from product to product within each participant. In the level-2 model, the intercepts from level 1 (mean conditioning for each participant) are modelled as a function of the level-2 experimental condition to which each participant was randomly assigned, X_j. The intercept in this model, p_{00}, is the mean evaluative conditioning on average across participants and c is the degree to which the magnitude of evaluative conditioning depends on the experimental condition. This is the total or unmediated effect of the treatment. u_{0j} is random variation from participant to participant within experimental condition in the magnitude of evaluative conditioning.

The next step involves looking at the impact of the level-2 manipulation on the level-1 mediator. The relevant equations are the following[6]:

Level 1: $M_{ij} = d_{0j} + e_{ij}$
Level 2: $d_{0j} = p_{00} + aX_j + u_{0j}$

These two models are identical to the previous two except now the dependent variable at level-1 is the degree to which the participant is contingency aware for the individual conditioned stimulus (i.e., can he or she state the valence of the unconditioned stimulus with which it was paired during conditioning?). In the level-2 model the slope of X_j is a, which, parallel to the earlier terms we used for mediation, represents the effect of the treatment on the mediator.

The final step consists in looking at the joint impact of the level-2 manipulation and of the level-1 mediator on the level-1 outcome. The relevant equations are the following:

Level 1: $Y_{ij} = d_{0j} + bM_{ij} + e_{ij}$
Level 2: $d_{0j} = p_{00} + c'X_j + u_{0j}$

As can be seen, these models include one predictor at the within-subject level, M, and one at the participant level, X. The estimate of the b parameter is assessed at the within-subject level, level 1, because the mediator is assumed to be linked to the product, that is, evaluative conditioning with respect to a given product is expected to emerge only when there is awareness of the contingency between this product and the specific unconditioned picture with which it was paired. In the level-2 model c' estimates the residual direct effect of the treatment on the outcome, over and above the mediator.

Multilevel models have a complex structure, most notably because they include errors at multiple levels. As a consequence, the parameters of the model cannot typically be estimated by means of standard least squares methods used for single-level models. Instead, estimation is typically carried out using restricted maximum likelihood (REML) estimation. Having said this, the logic underlying the test of an indirect effect, whether one chooses to rely on $a*b$ or $c – c'$, remains the same even though these estimators will not be exactly equivalent in multilevel models as they are in single-level models. As noted by MacKinnon (2008), the nonequivalence between the two sides of the equation is not really problematic because the discrepancy is likely to be small and decreases as sample sizes increase (Krull & MacKinnon, 1999). Table 25.1 gives an overview of a multilevel analysis of Pleyers et al.'s (2009) data.

Table 25.1. ***Multilevel Mediation Analysis of Pleyers et al.'s (2009) Data (the analysis was performed using SAS PROC MIXED; data and SAS code are available at http://www.psp.ucl.ac.be/mediation/medmod/)***

Prediction of level-2 X on level-1 Y

Level 1: $Y_{ij} = d_{0j} + e_{ij}$

Level 2: $d_{0j} = 0.1482 – 0.1134\ X_j + u_{0j}$

(0.0423) (0.0423)

Prediction of level-2 X on level-1 M

Level 1: $M_{ij} = d_{0j} + e_{ij}$

Level 2: $d_{0j} = −0.1242 – 0.5665\ X_j + u_{0j}$

(0.0546) (0.0546)

Prediction of level-2 X on level-1 M and level-1 Y

Level 1: $Y_{ij} = d_{0j} + 0.1736\ M_{ij} + e_{ij}$

(0.0458)

Level 2: $d_{0j} = 0.1697 - 0.0151\ X_j + u_{0j}$

(0.0411) (0. 0483)

As can be seen in Table 25.1, the analysis of the data collected by Pleyers et al. (2009) corroborates the fact that $a*b = c - c'$ as $-0.5665 * 0.1736 = -0.1134 - (-0.0151)$**,** which gives $-0.0983 = -0.0983$. Table 25.1 confirms that all the necessary conditions for mediation are satisfied. As a matter of fact, c as well as a and b are significant. Moreover, the standard error of the product can be shown to be .0276. This means that the confidence interval ranges from –.0442 to –.1524. The fact that the confidence interval does not include 0 confirms the presence of a significant indirect effect. In other words, manipulating the availability of cognitive resources was able to decrease the amount of evaluative conditioning manifested for the various products, and this took place via the impact of cognitive load on participants' contingency awareness, as indexed by their ability to associate each product with the specific image with which it had been paired.

The multilevel approach is a highly flexible one that can be used in a wide variety of designs (Schoemann et al., Chapter 15 in this volume). So, although analytic strategies have been proposed for the examination of repeated measures design with within-subjects manipulations and within-subjects measures of the mediator and of the dependent variable (Judd et al., 2001), such a 1–1–1 design can best be approached from a multilevel perspective. Coming back to our evaluative conditioning example, such a situation would occur if, for every participant in the study, some random set of products had been presented under conditions of cognitive depletion whereas the remaining products had been seen in the absence of a secondary task. A multilevel analysis is of course desirable because the residuals associated with the responses of a particular individual are likely to violate the assumption of independence.

From Measured to Manipulated Mediators

In recent years, several authors have voiced a series of warnings with respect to the possible dividends deriving from a mediational analysis. Indeed, it would seem that mediational analysis has become so popular that it constitutes a mandatory step for any scientific contribution claiming to shed light on a particular psychological process at work in the context of some phenomenon of interest. But does this research really hold its promise? Unfortunately, the answer is not as positive as one would hope it might be. As we discussed earlier, a key problem derives from the fact that a statistical analysis of a set of correlations is taken to confirm the specific causal model put forth by the researcher.

As was noted by early proponents of mediational analysis (Baron & Kenny, 1986; Judd & Kenny, 1981), as well as in more recent contributions (Fiedler, Schott, & Meiser, 2011; MacKinnon, Krull, & Lockwood, 2000; MacKinnon et al., 2002), if a third variable M is indeed a mediator, a logical implication is that its inclusion in the model will reduce the relation between X and Y. At the same time, however, the finding that controlling for M reduces the relation between X and Y does not in fact imply that M is indeed a mediator. Said otherwise, whether a selected causal variable reflects a real cause or not cannot be determined statistically. To be sure, statistical mediation is a necessary condition if one wants to substantiate the conjecture that some third variable is a true mediator, but researchers ought to realize that it is not a sufficient condition.

Fortunately enough, the fact that no correlation pattern can actually prove whether some third variable is causally implicated in the emergence of an effect does not leave researchers without ammunition. Next to the measurement-of-mediation strategy, several other options allow one to evaluate a causal model whereby some independent variable is thought to set in motion a psychological process that, in turn, produces a given outcome. One prime candidate is the so-called experimental-causal-chain design (Spencer, Zanna, & Fong, 2005). The idea is actually quite simple: When an experimental manipulation, X, is shown to have an impact on some dependent variable, Y, and a specific psychological process, M, is thought to be at work, the researcher is encouraged to decompose the causal sequence into two pieces and conduct an experimental study on each piece. In essence, after demonstrating the c effect, the goal is to conduct two independent experiments addressing both the a and the b effects.

A nice illustration of this strategy comes from a study by Word, Zanna, and Cooper (1974). Building on earlier work on the so-called Pygmalion effect (Rosenthal & Jacobson, 1968), these authors reasoned that people's stereotypes (X) could create a self-fulfilling prophecy (Y) via their nonverbal behavior (M). They first ascertained the a relation by having white participants interview a black or a white confederate. As predicted, participants proved more distant in their nonverbal behavior when facing a black than a white confederate. In a second experiment, white confederates interviewed white unaware participants. As an experimental manipulation of the mediator, interviewers either adopted the distant nonverbal behavior observed with black interviewees in the first experiment or the less distant behavior encountered with white interviewees in the first experiment. In line with the hypothesis, whites did worse on this interview when they were treated like the blacks had been in Experiment 1.

Of course, the experimental-causal-chain design also has limitations. Obviously, the decomposition mandates that the proposed psychological process should be both easy to measure and easy to manipulate. Perhaps the most difficult issue concerns the equivalence between the process as it is measured and the process as it is manipulated. In some cases, it may be difficult to argue compellingly that the dependent variable in the experiment demonstrating the *a* relation is the same as the independent variable in the experiment demonstrating the *b* relation. Additionally, the fact that two experiments are conducted in isolation does not allow a proper determination of the amount of variance in the dependent variable accounted for by the independent variable. Keeping these drawbacks in mind, the experimental-causal-chain design still constitutes a powerful tool to uncover the key role that some intervening variable may play along some presumed causal chain.

As we have already discussed, the problem with measuring the mediator rather than manipulating it leaves open the possibility that there are important omitted common causes that can explain the covariation between the mediator and the outcome. There is, additionally, another potential problem with relying on the measurement of the mediator to establish mediation. In many circumstances, a mediator is difficult to measure or its measurement may alter the causal process, either by eliminating the impact of X on Y or by creating an effect (via, for instance, awareness) where none would be observed in the absence of measurement (Jacoby & Sassenberg, 2011; Spencer et al., 2005). Another strategy builds on the realization that mediation rests on the comparison between a factual state, that is, the influence of X on Y, and a counterfactual state, that is, the relation between X and Y when controlling M. In light of this analysis, a smart way to approach the question consists in creating a design in which one compares two factual states. According to Jacoby and Sassenberg (2011; see also Sigall & Mills, 1998), this can be done by means of the testing process by interaction strategy (TPIS), which boils down to an experimental manipulation of the mediator.

Interestingly, the TPIS approach does not require that an experimental effect of *X* on *Y* be observed in the first place. If there is an impact of the independent variable in the so-called standard condition, the TPIS would consist in interrupting the process by means of the moderating variable. An alternative plan is to amplify the effect or even reveal an otherwise masked effect by counteracting some suppressing variable. A classic study by Zanna and Cooper (1974) illustrates both processes. These authors had all participants behave counter to their initial attitudes. The implementation of *X* was straightforward enough: Whereas some were simply forced to do so, others were led to believe that they enjoyed freedom of choice. The extent of attitude change constituted the *Y* variable. The manipulation of *M* rested on three conditions. In the control condition, nothing special took place. In line with dissonance theory, free-choice (but not forced) participants experienced unpleasant arousal and changed their attitudes so that they were better aligned with their behavior. In the interruption condition, participants were initially given a (placebo) pill that they were told would cause arousal. This time, free-choice participants did not change their attitudes at all. In the amplification condition, participants were also given a pill but thought that it would cause relaxation. Now free-choice participants modified their attitudes even more than in the control condition. In sum, manipulating the specific way participants experienced the arousal resulting from their counterattitudinal behavior critically affected attitude change, demonstrating the mediating role of the arousal along with its interpretation.

Promising as the experimental approach may appear, it is of course not in and of itself a panacea to deal with mediation. Some obstacles remain (Bullock, Green, & Ha, 2010). A first issue concerns the isolation of *M* – that is, the experimental manipulation of *M* needs to target *M* and nothing else. A second limitation has to do with the successful manipulation of *M*. In other words, manipulating *M* by means of some variable *Z* is not necessarily equivalent to changing *M* by means of a manipulation of *X*. For instance, it may be difficult to be certain that the same people are affected. Finally, researchers ought to realize that there is a possibility for within-sample variation with respect to the indirect effect. In other words, the average indirect effect is potentially misleading. Interestingly enough, this raises the issue of moderated mediation in which the indirect effect varies in magnitude as a function of some other variable. But before we turn to this topic, we turn to a closer examination of moderation.

Moderation

This section focuses on testing and interpreting moderator effects. We start by defining moderation and discuss its relationship with statistical interactions. We then turn to the basic models used to test for interactions and moderator effects. Following this, a major section is devoted to how one interprets the results of these models and best practices for the presentation of moderator effects. We then turn to issues that complicate the search for moderator and interaction effects, focusing in particular on considerations of statistical power. Finally, we briefly discuss some additional issues and designs in which moderation takes on somewhat different forms.

Definitions and Basic Models

As defined earlier, the question of moderation is the question of whether the impact of some independent variable on the dependent variable varies in magnitude as a function of some third variable. Defined in this way, moderation implicitly assumes a causal model in which the independent variable is a cause of the dependent variable and the magnitude of that causal impact depends on some third variable. As such, moderation is not the same as an interaction between two variables. An interaction is the finding that the simple slope of one predictor variable in a linear model varies as a function of the value of another predictor variable. Interactions can exist in the absence of any causal effects of either predictor variable on the dependent variable.

As we will show analytically, interactions are symmetric: If the simple slope of X on Y varies as a function of Z, then the simple slope of Z on Y varies as a function of X. But when we speak of moderation, we are saying that the X – Y causal relationship is moderated by some variable Z. Because the causal effect that is moderated goes from X to Y, rather from Z to Y, we cannot say that X moderates the Z – Y causal effect. Thus, moderation implies an interaction, but an interaction is not sufficient to claim moderation.[7] Moderation is an interaction plus the additional strong assumption of a causal impact of an independent variable on a dependent variable that varies in magnitude. The viability of this strong assumption cannot be assessed or confirmed through any data analytic steps. Rather its plausibility depends on theoretical considerations and design variations that permit relatively strong causal inference (e.g., randomized experiments and certain longitudinal designs).

Analytically, a model that estimates moderation is a linear model that estimates the effect of the interaction of two predictor variables on the dependent variable. Interactions are included in models by including as an additional predictor the product of two other predictor variables that are also included in the model. As made clear by Cohen (1978), for the product to estimate the interaction, the two component predictor variables must be included in the model. Thus, the interaction between X and Z is estimated by the slope for the product predictor, b_3, in the following linear model:

$$Y_i = b_0 + b_1 X_i + b_2 Z_i + b_3 X_i Z_i + e_i \tag{25.1}$$

A test of the interaction is equivalently conducted by testing the null hypothesis that the slope for the interaction is zero, by testing whether this model has a significantly smaller sum of squared errors than the model that does not include the product predictor, or by testing the partial correlations of the product with the criterion (Y), controlling for the two component variables, X and Z (Aiken & West, 1991; Cohen, 1968; Judd, McClelland, & Ryan, 2009).

The slope for the product predictor in the following model, which does not include the two component variables as additional predictors, does not in general estimate the interaction:

$$Y_i = b_0 + b_1 X_i Z_i + e_i$$

To illustrate why it is that the slope for the product predictor in Equation 25.1 estimates the interaction, we can re-express that model as the "simple" relationship between either of the predictor variables and the outcome. Accordingly, the following re-expression represents the "simple" relationship between X and Y at various levels of Z:

$$Y_i = (b_0 + b_2 Z_i) + (b_1 + b_3 Z_i) X_i + e_i \tag{25.2}$$

We can think of these "simple" relationships as simple linear regression models between Y and X whose intercepts and slopes take on different values at varying values of Z. Thus (b_0 + $b_2 Z_i$) is the simple intercept of these various models and (b_1 + $b_3 Z_i$) is the simple slope of X. As always, the intercept tells us the predicted Y value when X equals zero, and these simple intercepts vary as a function of the values of Z. And the slope tells us the change in predicted values as X increases by one unit, again with the magnitude of these simple slopes varying as a function of the values of Z.

Given this "simple" re-expression and the crucial centering issues that we will discuss in more detail later in this section, it is important to understand the meaning of the individual slopes in the model in the context of this re-expression.

b_0 estimates the intercept of the linear *Y*:*X* simple relationship when *Z* equals zero. b_1 estimates the slope of the linear *Y*:*X* simple relationship when *Z* equals zero. b_2 estimates the change in the intercept of the linear *Y*:*X* simple relationship as *Z* increases by one unit. b_3 estimates the change in the slope of the linear *Y*:*X* simple relationship as *Z* increases by one unit. It is this last parameter estimate that captures the *X* by *Z* interaction: To what extent does the simple slope between *Y* and *X* change as *Z* changes in value?

As outlined earlier, interactions are necessarily symmetric, although moderation is not. To demonstrate this symmetry, the model in Equation 25.1 can be equivalently re-expressed as the "simple" relationship between *Y* and *Z*:

$$Y_i = (b_0 + b_1 X_i) + (b_2 + b_3 X_i) Z_i + e_i \tag{25.3}$$

with symmetrically equivalent interpretations of the parameter estimates: b_0 estimates the intercept of the linear *Y*:*Z* simple relationship when *X* equals zero. b_2 estimates the slope of the linear *Y*:*Z* simple relationship when *X* equals zero. b_1 estimates the change in the intercept of the linear *Y*:*Z* simple relationship as *X* increases by one unit. b_3 estimates the change in the slope of the linear *Y*:*Z* simple relationship as *X* increases by one unit.

Given that interactions imply that the simple slopes for one predictor take on different values at various levels of another predictor, they imply nonparallel "simple" regression lines. Ordinal interactions are defined as interactions where all the simple slopes for one predictor have the same sign across all meaningful levels of the other predictor. Disordinal or crossover interactions are ones where the simple slopes have both positive and negative values across the meaningful levels of the other predictor.

To this point, we have said nothing about the scale of measurement of either the independent variable or its moderator. When predictor variables are categorical, these need to be coded numerically. Typical coding conventions include dummy coding and contrast coding (also known as effects coding). With two levels of a categorical predictor, the former coding convention uses values of 0 and 1 for the two groups while the latter uses values that sum to zero across the two groups (e.g., –.5 and +.5). Again the slope of the partialled product of two variables, regardless of whether they are continuously measured or coded categorical variables, will estimate their interaction.

If the moderator variable (*Z*) is categorical, analyses are frequently reported examining whether the magnitude of the *X*:*Y* relationship is different for the different groups, defined by the categorical levels of Z. Frequently this takes the form of testing whether the correlations between X and Y differ across the groups. Such a test is not in general the same as testing whether Z moderates the X:Y relationship, defining moderation as a statistical interaction. Interactions examine whether different simple slopes are needed for the different groups. Correlations in the different groups reflect not only those simple slopes but also the variances of X. It is entirely possible for the different groups to have different simple slopes but the same correlations. And the reverse is entirely possible as well. These two situations are illustrated in the graphs of Figure 25.5 (taken with permission from Whisman & McClelland, 2005). At the top (panel A) we have a situation where the two groups have different slopes but the same correlation. Thus in this case there is an interaction that would be undetected by a comparison of the two group correlations. At the bottom (panel B) there is no interaction – that is, the slopes in the two groups are the same – but one group has a larger correlation than the other does because of relatively greater variance of the X variable. The lesson is that one should test moderation as an interaction rather than by comparing the magnitude of correlations.

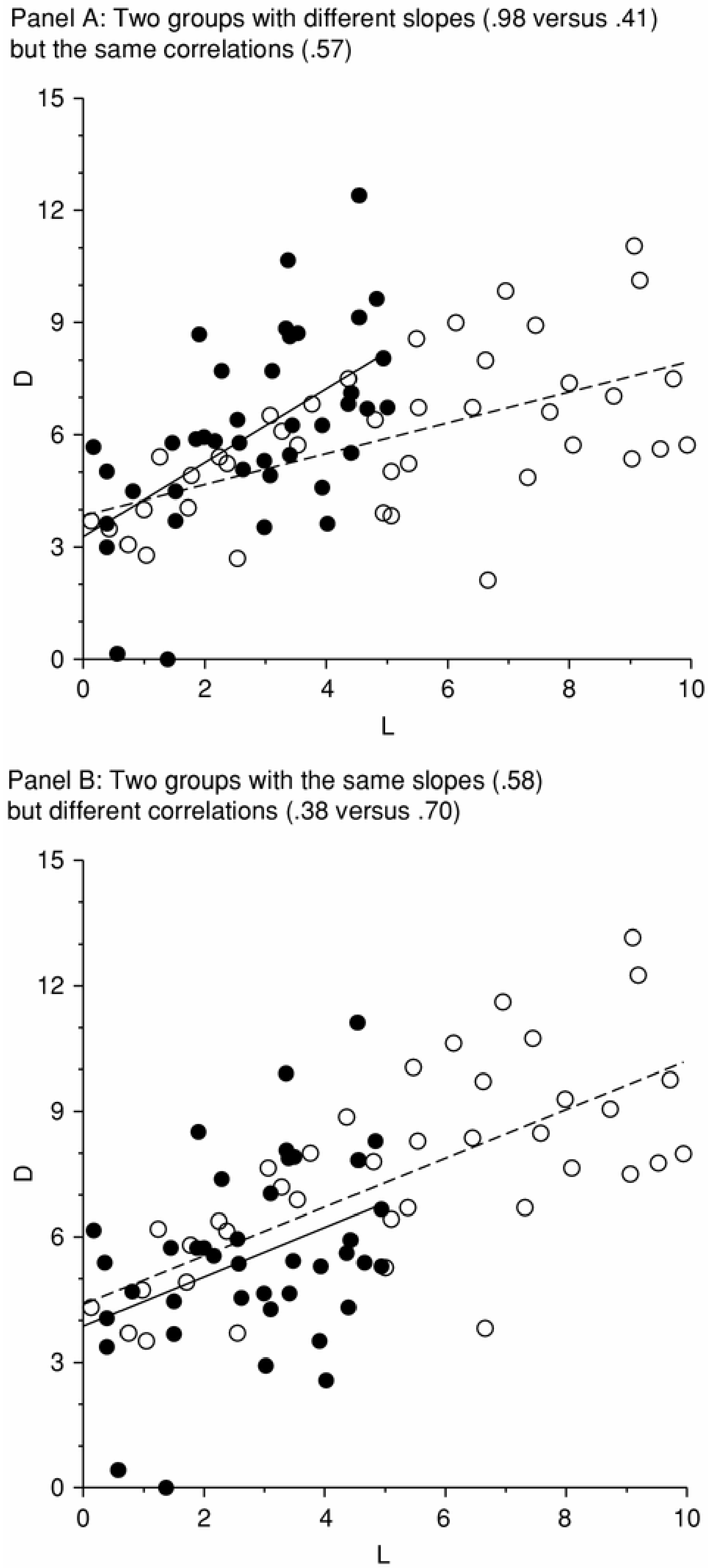

Figure 25.5. Panel A: Two groups with different slopes (.98 versus .41) but the same correlations (.57). Panel B: Two groups with the same slopes (.58) but

different correlations (.38 versus .70). Taken with permission from Whisman & McClelland (2005).

Interpretation and Presentation

Deriving and Plotting Simple Relationships. Once one has found a significant interaction, issues arise as to how that interaction should be interpreted, discussed, and displayed. An initial decision that needs to be made concerns which simple re-expression of the interactive model is theoretically most informative. Recall that there are two such simple re-expressions:

$$Y_i = (b_0 + b_2 Z_i) + (b_1 + b_3 Z_i) X_i + e_i$$
$$Y_i = (b_0 + b_1 X_i) + (b_2 + b_3 X_i) Z_i + e_i$$

If indeed the interaction is because of a moderation of a causal treatment effect – say, for instance, that *Z* moderates the impact of *X* on *Y* – then the choice of the more informative re-expression is easy: The interest is in *X* and how the simple relationship between *X* and *Y* depends on *Z* (i.e., the first of the preceding two re-expressions). In other cases, where there is not a clear causal model that can be assumed, there is no easy rule to follow in deciding in favor of one or the other of these re-expressions. One should try telling a theoretical story with them both and decide which is the more theoretically interesting and informative. Do you want to argue that the simple relationship between *Y* and *X* depends on *Z*? Or do you prefer to argue that the simple relationship between *Y* and *Z* depends on *X*? Both arguments would be correct, but one will generally make a more compelling story than the other.

For now, let us assume that the preferred interpretation is that the simple *Y*:*X* relationship depends on *Z*. Given this, one derives and plots simple *Y*:*X* linear relationship predicted by the model at different representative and theoretically meaningful values of *Z*. If *Z* is categorical, the choice of these values is easy: One wishes to plot the simple linear relationships for each of the groups defined by the *Z* categories. If *Z* is continuously measured, then the choice of the appropriate values of *Z* for these plots is less clear. One convention derives and plots simple regression lines at the mean of *Z* and at values of *Z* one standard deviation above and below the mean (Aiken & West, 1991). For instance, suppose that values of both *X* and *Z* vary between 1 and 5, and the mean of *Z* is 2.5 with a standard deviation of 1. And suppose the following are the parameter estimates from the interactive model:

$$Y_i = 2.0 + 0.5X_i - 0.5Z_i + 0.3X_i Z_i + e_i$$

The general form of the re-expression of this model is:

$$Y_i = (2.0 - 0.5Z_i) + (0.5 + 0.3Z_i)X_i + e_i$$

At the mean value of *Z* (i.e., 2.5), the simple *Y*:*Z* relationship is:

$$\begin{aligned} Y_i &= (2.0 - 0.5(2.5)) + (0.5 + 0.3(2.5))X_i + e_i \\ &= 0.75 + 1.25X_i + e_i \end{aligned}$$

And at *Z* values one standard deviation above and below the mean, the simple relationships are:

$$\begin{aligned} Y_i &= (2.0 - 0.5(3.5)) + (0.5 + 0.3(3.5))X_i + e_i \\ &= 0.25 + 1.55X_i + e_i \end{aligned}$$

$$\begin{aligned} Y_i &= (2.0 - 0.5(1.5)) + (0.5 + 0.3(1.5))X_i + e_i \\ &= 1.25 + 0.95X_i + e_i \end{aligned}$$

One then might plot these three simple models, with *X* on the horizontal axis (with values between 1 and 5) and three different lines, one for each simple relationship. Such a plot is given in Figure 25.6. Inspection of this plot leads to relatively clear interpretations: As *X* increases, predicted values of *Y* increase and this is more true at higher levels of *Z*. That is, the moderation of the *Y*:*X* relationship by *Z* is such that the positive relationship between *X* and *Y* becomes larger as *Z* increases.

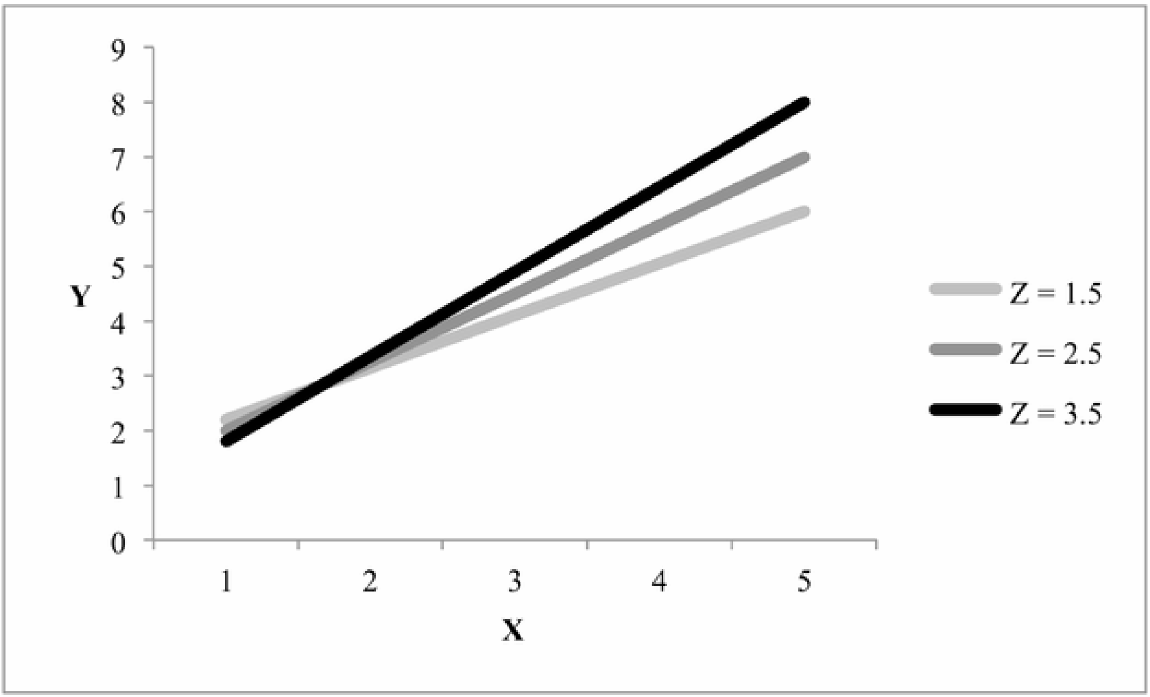

Figure 25.6. Simple Y:X relationships at values of Z of 1.5, 2.5, and 3.5.

There is of course nothing sacred about the choice of Z values at which one chooses to plot these simple relationships. If some theory motivated the desire to see the simple relationship when Z equals 2, one could easily derive and plot this simple relationship in addition to or instead of those at the values of the mean and plus one and minus one standard deviations.

Testing Simple Slopes. A significant interaction tells us that simple slopes vary significantly across the range of values of the moderator variable. But it does not tell us about whether particular simple slopes, at particular values of the moderator variable, are significantly different from zero. Often the further question of whether particular simple slopes are significant is of interest, even though significant simple slopes should not be seen a requirement for interpreting significant interactions (Judd, McClelland, & Culhane, 1995; Rosnow & Rosenthal, 1989).[8]

Given the interactive model and its re-expression in the terms of the Y:X simple relationship:

$$Y_i = b_0 + b_1 X_i + b_2 Z_i + b_3 X_i Z_i + e_i$$

$$Y_i = (b_0 + b_2 Z_i) + (b_1 + b_3 Z_i) X_i + e_i$$

it is apparent that b_1 estimates the simple Y:X slope when Z equals zero. And a statistical test of whether this coefficient differs from zero accordingly provides a test of the simple X slope at that particular value of Z. This then provides a general solution for testing the simple slope of X at any value of Z that might be of interest: One simply deviates Z from that value, recomputes the product term, and tests the X coefficient in the interactive model.

Consider Z' defined as Z deviated from some value c of interest: $Z' = Z - c$. If one now estimated the interactive model using Z' instead of Z (and importantly recomputing the product predictor so that it is XZ'):

$$Y_i = b_0 + b_1 X_i + b_2 Z_i' + b_3 X_i Z_i' + e_i$$

which, re-expressed in terms of the Y:X simple relationship, yields:

$$Y_i = (b_0 + b_2 Z_i') + (b_1 + b_3 Z_i') X_i + e_i$$

Accordingly, b_1 now estimates the simple slope of X when Z' equals zero, which it will of course when $Z = c$. In other words, by deviating Z around different values, the slope of X in the model that includes the product predictor takes on different values and these values are the simple X slope at whatever value of c has been used in computing the deviated Z, Z'. And a test that the slope of X differs from zero provides an inferential test of whether the simple slope of X when $Z = c$ differs from zero.

Most commonly, and as recommended by many (e.g., Aiken & West, 1991; Cohen, Cohen, West, & Aiken, 2003; Judd et al., 2009), the value used for c is the mean of Z. This is what is commonly referred to as mean-centering the predictor. Under mean-centering, the slope associated with X will equal the simple slope of X when at the mean of Z in the context of the interactive model that allows the simple slope of X to vary across the values of Z.

Other routinely used values of c include values one standard deviation above and below the mean of Z, as mentioned earlier, thus permitting one to estimate and test simple slopes of X when Z equals those values. But, again, there is nothing sacred about these routinely used values of c. If there are other theoretically meaningful values of Z at which it is important to test whether the simple X slope differs from zero, then the interactive model should be reestimated while deviating Z from those values.

Although the aforementioned approach represents a straightforward procedure for testing the simple slope of X at different values of Z that are of interest, Preacher, Curran, and Bauer (2006) provide more general procedures that enable the researcher to examine the range of values of Z across which the simple slope of X is and is not significant, given an interactive model. They have implemented their approach in a highly useful Web-based application that is available at http://quantpsy.org/interact/mlr2.htm (see also Hayes & Mathes, 2009).

It is worth noting one interesting side consequence of what we have just discussed. We have been considering an additive transformation of one of the variables, $Z' = Z - c$, involved in an estimated interactive model. And what we have shown is that as c takes on different values, the estimated coefficient for X – that is, b_1 – varies, because it generally will equal the simple X slope when Z' equals zero or when $Z = c$. The other estimated slopes in the model, however – b_2 and b_3 – remain constant. Hence, we are left with the curious result that an additive transformation of one of the component predictor variables involved in an interactive model affects the estimated value of the partial slope for the other variable that is a component of the product interaction, but it has no effect on the estimated partial slope of the component variable that is transformed. Additionally, it has no effect on the estimated slope of the product predictor.[9]

Tests of "Main Effects" in Interactive Models. In light of the preceding discussion, it should be clear that the estimated slopes of the component variables (b_1 and b_2) do not in general estimate what the literature devoted to the analysis of variance (ANOVA) refers to as main effects. In the ANOVA literature, a *main effect* represents the effect of one experimental factor on average across the levels of other crossed experimental factors. An estimated slope of a component variable in an interactive model is in general a "simple" slope for that component variable when the other component variable equals zero. If one centers the two component variables on their means (and computes the product predictor as the product of those centered components), then the slope of each component variable will estimate the "simple" slope at the mean value of the other component variable. But even this is not conceptually the same thing as a main effect in the analysis of variance literature. In general, main effects of component variables cannot be defined when those variables are measured more or less continuously and when they are involved in an interaction. An interaction by definition means that there is no "overall" or "main" effect of a component variable. An interaction by definition means that the effects of one component variable vary as a function of the other one. Accordingly, it is generally a mistake to refer to main effects in moderated regression models. One can certainly estimate and test "simple" effects at the mean value of other component variables, and in many cases these will be very similar to average affects of one variable across the levels of the other variable. But strictly speaking, they are not that; rather they are "simple" effects at the average value of the other variable. Certainly if one fails to center component variables around their means, then slopes associated with those component variables provide nothing resembling what the ANOVA literature defines as main effects.

Standardization. Authors frequently report standardized slopes (or betas) in regression models rather than slopes with the variables in their raw metrics. In general we are not enamored of such practices, for reasons that we briefly discuss but that are reviewed in detail elsewhere (see Turkheimer & Harden, Chapter 8 in this volume; also Cohen, 1990; Tukey, 1969). In simple regression, estimated standardized regression coefficients equal estimated correlations. Hence, as Figure 25.5 illustrated, they are affected by both the magnitude of the raw or unstandardized slope and the variability of the predictor variable. Accordingly, the metrics used in deriving standardized slopes are sample-specific, with the result that standardized slopes will vary from sample to sample, even when raw slopes do not.

In interactive models, these issues are complicated further, as explained by Aiken and West (1991), Freidrich (1982), and Whisman and McClelland (2005). The problem arises from the fact that the product of two standardized variables is not itself standardized (i.e., it will not in general have a mean of 0 and a standard deviation of 1). Accordingly, one might estimate an interactive model and then attempt to interpret the standardized regression coefficient associated with the product predictor. But that slope might be something rather different than if one first standardized the *Y*, *X*, and *Z* variables, computed the product of the two standardized predictors, and then regressed *Y* on standardized *X*, standardized *Y*, and their product. Because of this, we find efforts to interpret standardized slopes in interactive models relatively misleading and at best uninformative. Of course, when a linear regression program outputs those standardized estimates and tests them, those inferential tests are identical to tests of unstandardized or raw regression slopes. But the standardized slopes themselves are interpreted only with difficulty.

Difficulties of Detecting Interactions

The primary challenge in conducting research where moderation is hypothesized is one of assuring adequate statistical power to test interactions. Of course, statistical power is an issue in the conduct of all research (Cohen, 1988), or at the very least should be. Cohen provided guidelines for power consideration when testing "small," "medium," and "large" effects. Effect sizes calibrated in this manner are most typically expressed in *d* units (the ratio of a mean difference to the pooled within group standard deviation). A more general effect size estimate that can be used to calibrate effect sizes is the squared partial correlation, or PRE (Judd et al., 2009), according to which "small," "medium," and "large" effects correspond to PREs of .02, .13, and .25 respectively. According to Cohen's power tables, to have adequate power (i.e., $1 - \beta = .80$; $\alpha = .05$), one would need sample sizes of 392, 55, and 26 to detect, respectively, "small," "medium," and "large" effects, assuming the absence of measurement error.

This general conclusion about statistical power is complicated in the testing of interactions by two issues that are particularly pernicious in this case. First, unreliability in the measured variables substantially reduces power, and the unreliability of product predictors is a multiplicative function of the unreliability of its component variables (Busemeyer & Jones, 1983; Cohen et al., 2003). Thus, if two variables have reliabilities of .80, the reliability of their product would equal only .64. And as the reliability of a predictor declines, the power needed to test its effect, given some true effect size, is substantially reduced. Aiken and West (1991) estimate that as the reliability of a predictor declines from 1.00 to .80, the sample sizes necessary for acceptable power levels are likely to double. Accordingly, with unreliability of a product predictor being the product of the unreliabilities of its components, sample sizes needed for tests of interactions are likely to be substantially greater than those given earlier.

A second power problem concerns the restriction in range of predictors. In the case of any predictor variable, if it does not vary very much, it is difficult to find an association between it and some outcome variable. This is why in most experimental research, one goes to some length to ensure that the experimental manipulation is substantial enough (i.e., that the difference between a control condition and the treatment condition is large). Outside of the experimental laboratory, it is often difficult to sample respondents who have substantial variability on important predictor variables. Assuming most measured predictor variables have a somewhat unimodal quasi-normal distribution, their variances are typically very substantially less than what the variances would be if everyone was found only at one or the other extreme value (as we typically construct the distribution to be in experiments).

Thus, whatever factors lead to a restriction of range of a predictor (or, more accurately, a reduction in its variability) decrease the power to find an effect of that predictor. In the case of a product predictor, McClelland and Judd (1993) show that its variance is a multiplicative function of the variance of its components, much like its reliability. Hence factors that restrict or limit the variance of the component variables restrict or limit the variance of the product predictor even more. McClelland and Judd (1993) compare the relative efficiency or power to detect an interaction given various joint distributions of the component variables. In the best-case scenario, all of the observations are located at the most extreme four corners of the joint distribution of *X* and *Z* (very high on *X* and very high on *Z*, very high on *X* and very low on *Z*, etc). This is the ideal design for detecting an interaction, given some constant *N*, and is the basis for a 2 × 2 crossed experimental design with manipulations that maximize the variance of the two independent variables. In real-world settings with measured, rather than manipulated, independent variables, it is exceedingly unlikely to encounter such a joint distribution. More likely would be a roughly bivariate normal joint distribution, with most observations clustered near the joint means of the two component variables. McClelland and Judd (1993) showed that such a joint bivariate normal distribution requires 17 times the number of observations to have the equivalent power to detect an interaction compared with the optimal four-corners design.

Thus, when the component variables that are expected to interact are measured variables and when the distribution of each is roughly normal, one will generally have very low power to test an interaction, compared to the optimal design with all observations at the most extreme values of the component variables. As the range of the component variables is reduced, the variance of the product predictor is reduced even more drastically, resulting in a very substantial loss of power. This explains why significant interactions, which are relatively ubiquitous in experimental research, are reported only infrequently using survey or correlational data, unless the sample sizes are exceedingly large (e.g., greater than 500). This then provides some guidance for sampling strategies if one wishes to argue an interactive hypothesis, given measured rather than manipulated independent variables. Rather than sampling randomly, the more powerful alternative is to sample purposively, oversampling the extreme four corners of the joint distribution of the predictors. Some might object that then one moves away from a sample that is truly representative of the population, which of course is true. But, as always, there are multiple simultaneous and often conflicting goals in research. If one wishes to find significant interactions, then oversampling observations at the extreme four corners of the joint distribution is the most powerful approach.[10]

From the preceding discussion readers might draw two erroneous conclusions. The first would be that given a random sample of observations and a hypothesized interaction between two measured predictor variables, one should restrict the analysis to observations that are toward the extreme four corners of the joint distribution. But throwing out observations, regardless of where they are in the joint distribution of the two predictors, will always result in a decrease in statistical power for testing interactions. Given a constant *N*, it is easier to detect interactions by oversampling the extreme four corners.

A second erroneous conclusion that might be drawn from the fact that interactions are typically found with more power in experiments than with measured predictors is that one should dichotomize predictor variables – using median splits, for instance – and conduct analysis of variance instead of treating measured predictors continuously. There is now a large literature showing that dichotomizing continuous predictors will not result in increases in statistical power (Cohen, 1983; Irwin & McClelland, 2003; MacCallum, Zhang, Preacher, & Rucker, 2002; Maxwell & Delaney, 1993).[11]

Historically, there have been some who have blamed low power in detecting interactions on multicollinearity between the product predictor variable and its components. In fact, this is not a factor because the collinearity of a product with its component variables is a function of the coding of the component variables, being substantially reduced when the components are centered around their mean. And, as already discussed, such centering has no impact on the test of the interaction in moderated regression models.

Multilevel Interactive Models

In the section of this chapter devoted to mediation, we discussed multilevel models suitable for data where there are nestings of observations that induce dependence (i.e., multiple observations from each participant, observations grouped in classrooms, etc.). Here we provide a short treatment of moderation in such circumstances as well.

As an example, let us turn to a study by Toma, Corneille, and Yzerbyt (2012, Study 3). These authors were interested in social projection and how the probability of success influences people's tendency to project in cooperative settings. They invited participants into the laboratory and asked them first to self-describe on a series of positive and negative personality traits. Next, participants were told to imagine that they and a partner were involved in a cooperative task allegedly taking place at a software company. Depending on conditions, they learned that the probability of success of the task, and thus of gaining access to a much-desired outcome, was either low (they were informed that, in the past, some 5% of the teams succeeded) or high (95% of the teams succeeded). Finally, participants were presented with the same list of traits as before and asked to describe their partner as well as to indicate the valence associated with each trait.

These data are inherently multilevel, with individual traits that are rated being the first level and the participants being the second. At level 1 – within each participant – there are three variables: participants' ratings of their partner, their rating of themselves, and their rating of the trait's valence. At level 2 – between participants – is the manipulated variable: the probability of success of the cooperative task. Toma et al. (2012) expected social projection – that is, that people would generally perceive their partner as being similar to themselves – but that this tendency would be less marked when the probability of success of the cooperation task is thought to be low. The authors also hoped that this pattern would not be affected by the valence of the traits.

For the sake of this presentation, we simplify the analysis somewhat by looking only at social projection for the negative traits (see Table 25.2). At level 1, we consider, that for each one of the j participants, the rating of their partner on the i traits (Y_{ij}) should be predicted by their self-rating on the same traits (X_{ij}).

$$Y_{ij} = b_{0j} + b_{1j}X_{ij} + e_{ij}$$

Table 25.2. ***Multilevel Moderation Analysis of Toma et al.'s (2012, Study 2) Data (the analysis was performed using SAS PROC MIXED; data and SAS code are available at http://www.psp.ucl.ac.be/mediation/medmod/)***

Level 1: $Y_{ij} = b_{0j} + b_{1j}X_{ij} + e_{ij}$

Level 2: $$\begin{aligned} b_{0j} &= a_{00} + a_{01}Z_j + u_{0j} \\ b_{1j} &= a_{10} + a_{11}Z_j + u_{1j} \end{aligned}$$

After substitution, we thus have:

$$Y_{ij} = a_{00} + a_{01}Z_j + a_{10}X_{ij} + a_{11}Z_jX_{ij} + u_{0j} + u_{1j}X_{ij} + e_{ij}$$

In the main analysis, this gives:

$$\underset{(0.133)}{Y_{ij} = 3.836} + \underset{(0.133)}{0.046Z_j} + \underset{(0.065)}{0.293X_{ij}} + \underset{(0.065)}{0.163Z_jX_{ij}} + u_{0j} + u_{1j}X_{ij} + e_{ij}$$

When participants expect success of the cooperation, this becomes:

$$Y_{ij} = \underset{(0.187)}{3.882} + \underset{(0.133)}{0.046Z_j} + \underset{(0.089)}{0.456X_{ij}} + \underset{(0.065)}{0.163Z_jX_{ij}} + u_{0j} + u_{1j}X_{ij} + e_{ij}$$

When participants expect success of the cooperation, this gives:

$$Y_{ij} = \underset{(0.188)}{3.900} + \underset{(0.133)}{0.046Z_j} + \underset{(0.093)}{0.129X_{ij}} + \underset{(0.065)}{0.163Z_jX_{ij}} + u_{0j} + u_{1j}X_{ij} + e_{ij}$$

As can be seen, each participant has an intercept and a self-rating slope, estimating the impact of the characterization of the self on the characterization of the partner. Greater projection of self-ratings onto partner ratings should be indicated by greater slopes.

At the second level, Toma et al. (2012) modeled both the intercepts and the slopes as a function of the probability of success of the cooperative task, contrast-coded (–1, +1) as Z_j:

$$b_{0j} = a_{00} + a_{01}Z_j + u_{0j}$$
$$b_{1j} = a_{10} + a_{11}Z_j + u_{1j}$$

In these level-2 models, the *a*'s are estimated slopes and intercepts and the *u*'s are level-2 errors or residuals. The first model is modeling the mean rating[12] of the partner as a function of the probability of success, with a_{01} estimating the degree to which the mean rating of the partner differs between the two experimental conditions. The second of these level-2 models is modeling social projection: to what extent is the rating of the partner a function of participants' self-rating? The intercept in this second model estimates the mean level of social projection, averaging across participants, and the slope a_{11} estimates the degree to which social projection depends on the experimental manipulation. It is thus this last slope that corresponds to the critical multilevel interaction, that is, the tendency of self-ratings (a level-1 variable) to predict the partner ratings (a level-1 variable) as a function of the manipulated probability of success of the cooperation task (a level-2 variable).

What may seem surprising here is that it is the slope of the level-2 predictor in this second level-2 model that captures the critical interaction, when our expectation up until now has been that slopes of product predictor variables estimate interactions. But it is easy to show that Toma et al. (2012) were in fact modeling a product predictor. In the following we have substituted the level-2 estimated models into the level-1 model:

$$\begin{aligned} Y_{ij} &= b_{0j} + b_{1j} X_{ij} + e_{ij} \\ &= (a_{00} + a_{01} Z_j + u_{0j}) + (a_{10} + a_{11} Z_j + u_{1j}) X_{ij} + e_{ij} \\ &= a_{00} + a_{01} Z_j + a_{10} X_{ij} + a_{11} Z_j X_{ij} + u_{0j} \\ &\quad + u_{1j} X_{ij} + e_{ij} \end{aligned}$$

Thus, what we ultimately have is a model of social projection as a function of the probability of success, the self-ratings, and the interaction between probability of success and self-ratings. However, this multilevel model differs from the earlier interactive models in that we now have three random-error terms rather than just a single one. First, there is random variance from participant to participant in the mean rating given to the partner (u_{0j}); second, there is random variance from participant to participant in the effect of self-ratings ($u_{1j}X_{ij}$); and finally, there is random variance in individual observations from the participants (e_{ij}). It is the presence of these multiple random-error terms that accommodates the hierarchical nature of the data, allowing for heterogeneity of variance at the different levels.

Table 25.2 presents the output of the PROC SAS analysis of a simplified version of Toma et al.'s (2012) data. It can be seen that self-ratings globally predict partner ratings, a_{10} = 0.293, t = 4.53, $p < .0001$. Importantly, the coefficient associated with the critical interaction term is also significant, a_{11} = 0.163, t = 2.53, p = .016. In line with the authors' predictions, follow-up analyses, looking at the simple effects of self-ratings on partner ratings in each of the two experimental conditions, confirm that the impact of self-rating proves highly significant when participants expected the cooperation to succeed, a_{10}= 0.456, t = 5.11, $p < .0001$, whereas there is little trace of social projection when the probability of success was low, a_{10} = 0.129, t = 1.39, p = .17.

In what we have just examined, one of the interacting variables was measured at level 1 and one at level 2. If they are both measured at the same level, either level 1 or level 2, then their interaction would be modeled as a simple product predictor either at level 1 or level 2.

Moderated Mediation and Mediated Moderation

Having now discussed mediation and moderation, we briefly turn to a consideration of models in which both processes are at work. In the case of mediated moderation, Z moderates the overall or total effect of X on Y and the researcher wants to show that some mediator, M, mediates this moderation. The researcher thus wants to show that the moderation is mediated. To illustrate, imagine a persuasion researcher who has shown that attitude change in response to a persuasive communication depends on the interaction of the number of persuasive arguments and their quality: More arguments leads to more persuasion, but only when those arguments are of high quality. It seems reasonable that this interactive effect might be mediated by the depth of processing of the persuasive message. That is, more arguments lead to more in-depth processing, which in turn leads to greater persuasion, but the first link here, from more arguments to more in-depth processing, is mainly found for high-quality arguments.

In the case of moderated mediation, there is an overall treatment effect of X on Y, and the researcher wants to show that the mediation of this treatment effect is different (i.e., moderated) at the different levels of a moderator, Z. The researcher thus wants to show that the mediation is moderated. As an illustration here, imagine a researcher who is interested in the effects of mere exposure on liking. The mediational argument underlying mere-exposure effects might be that more frequent exposure to an object leads to a sense of familiarity, which in turn may lead to greater liking. But the researcher argues that the degree to which this mediational chain may hold should depend on the novelty of the object because the sense of familiarity with the object cannot increase much for objects that are not novel. Thus, the hypothesized mediational path, from exposure to familiarity to liking, is moderated by novelty.

Interestingly, although the starting questions are different, the basics of mediated moderation and moderated mediation are the same: Both rely on the same underlying models and both imply that the indirect effect (i.e., $a*b$) of the treatment on the outcome via some mediator is moderated by some other Z variable. In other words, the magnitude of $a*b$ depends on Z. Where the two differ, however, is in whether one starts by presuming moderation of the overall or total treatment effect and wants to find a mediator for this moderation or whether one starts by presuming an overall effect and wants to show this overall effect is mediated to a larger extent at different levels of the moderator (Muller, Judd, & Yzerbyt, 2005).

To examine either mediated moderation or moderated mediation, the following models are estimated:

$$
\begin{aligned}
Y_i &= b_{10} + b_{11}X_i + b_{12}Z_i + b_{13}X_iZ_i + e_{1i} \\
M_i &= b_{20} + b_{21}X_i + b_{22}Z_i + b_{23}X_iZ_i + e_{2i} \\
Y_i &= b_{30} + b_{31}X_i + b_{32}Z_i + b_{33}X_iZ_i + b_{34}M_i \\
&\quad + b_{35}M_iZ_i + e_{3i}
\end{aligned}
$$

In the following, we assume that in all models *Z* has been centered on its mean.

In the first of these models, b_{11} estimates the total effect of *X* on *Y* at the average level of *Z* and b_{13} estimates the degree to which that total effect is moderated by *Z*. In the terminology of mediation, which we gave in the first part of this chapter, b_{11} is equivalent to *c*, the total effect, allowing that effect to be moderated by *Z*. Note that in the context of a mediated moderation, although b_{11} is equivalent to *c*, one is primarily interested in b_{13} and is seeking to explain, via mediated moderation, what the mediating process is that is responsible for the moderation of the overall effect of *X* on *Y*.

In the second model, b_{21} estimates the effect of *X* on the mediator, *M*, at the average level of *Z* and b_{23} estimates the degree to which that effect is moderated by *Z*. In the terminology of mediation given earlier, b_{21} is equivalent to *a*, the first portion of the indirect effect, again allowing that effect to be moderated.

In the third model, b_{31} is the residual direct effect of the treatment on the outcome at the average level of *Z* and b_{33} estimates the degree to which that residual direct effect is moderated. In the earlier terminology, b_{31} is equivalent to *c'*, allowing this effect to be moderated. Again, note that in the context of mediated moderation, the parameter one is primarily interested in is b_{33}, asking whether the overall moderation of the effect of *X*, b_{13}, is reduced once one controls for the mediating process (and its moderation).

And finally, also in the third model, b_{34} is the partial effect of the mediator on the outcome controlling for the treatment at the average level of *Z* and b_{35} estimates the degree to which that effect is moderated. Again, in the earlier terminology b_{34} is equivalent to *b*, allowing this effect to be moderated.

The resulting mediation models are portrayed in Figure 25.7. The top diagram represents the total effect (the first of the models above), allowing that total effect to be moderated. The bottom diagram represents the second and third models above, allowing all possible effects in this mediational model to be potentially moderated. Earlier, when discussing mediation, we presented the fundamental mediational equality $c - c' = a*b$, with the effects in this equality defined as in Figure 25.1. As shown by Muller et al. (2005), there is a similar equality that holds for the mediated moderation and moderated mediation model of Figure 25.7, although now the effect that should be reduced (in the case of mediated moderation) or increased (in the case of a prototypical moderated mediation; see Muller et al., 2005) is not b_{11} (the conceptual analog to c) but b_{13}. Hence, assuming that X is a dichotomous treatment variable that has been contrast-coded (and Z is centered), the equality underlying mediated moderation and moderated mediation is $b_{13} - b_{33} = (b_{23}b_{34}) + (b_{21}b_{35})$. What this equality shows is that the overall moderation of the treatment effect, b_{13}, differs from the moderation of the residual treatment effect on the outcome, b_{33}, as a function of the degree to which the indirect effect is moderated. And in considering whether the indirect effect is moderated, there are two components to consider: whether the effect of the treatment on the mediator is moderated, b_{23}, times the average effect of the mediator on the outcome, b_{34}, and whether the effect of the mediator on the outcome is moderated, b_{35}, times the average effect of the treatment on the mediator, b_{21}.

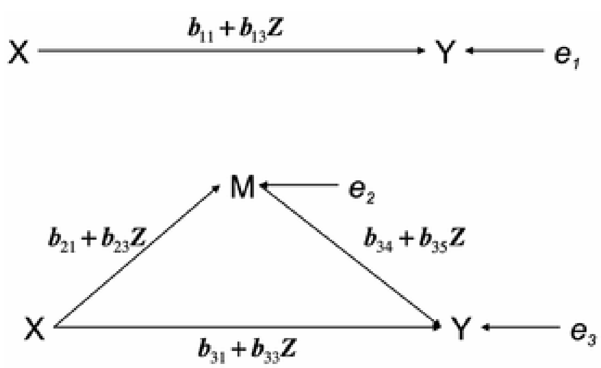

Figure 25.7. Models for mediated moderation and moderated mediation.

Importantly, what this equality further shows is that the moderation of the indirect effect can happen in two ways. First, it may be that the treatment effect on the mediator is moderated (and the mediator affects the outcome). Second, it may be that the mediator's effect on the outcome is moderated (and the treatment affects the mediator). And of course, both of these may be true simultaneously. For us (and others; see also Preacher, Rucker, & Hayes, 2007), this distinction between which component of the indirect effect is moderated is an important theoretical distinction. If there is moderation of the indirect effect via a mediator, then it may be the case that the treatment effect on the mediator is moderated, or it may be the case that the mediator effect on the outcome is moderated.

In sum, because we see this distinction as critical, we suggested that to claim mediated moderation or moderated mediation, in addition to a significant b_{13} (in the case of mediated moderation) or a significant b_{11} (in the case of moderated mediation), researchers need to find either b_{23} and b_{34}, or b_{21} and b_{35}, conjointly significant (Muller et al. 2005). Although we do not see it as mandatory (see the Mediation section), one may also want to test whether the overall indirect effect is moderated. To do so, interested readers could refer to the extensive work by Preacher et al. (2007), who provide such tests in the context of bootstrapping techniques.

Conclusion

For social and personality psychologists, the techniques for assessing mediation and moderation have become very important tools that are widely used throughout the discipline. Although their use is not without pitfalls, and these have sometimes seriously limited what one can conclude from such analyses, we are convinced that these are very valuable tools. Their widespread use will continue for the foreseeable future. What we hope to have provided in this chapter is a relatively accessible but thorough guide for the use of these tools and, in so doing, to have clarified underlying assumptions, ongoing controversies, and areas of ambiguity where further work is warranted.

Readers who are familiar with the literature we have reviewed will be aware that some of our definitions, arguments, and suggestions are at variance with definitions, arguments, and suggestions advocated by others whom we highly respect. In our view, this divergence is exciting because it suggests that the last word remains to be written about the wise and appropriate use of mediation and moderation analyses. While these tools are already well developed and widely used, we are convinced they will continue to be refined so that their application will only become more precise and fruitful.

Social and personality psychologists now have at their disposal a wide range of very sophisticated methodological tools that were not in existence some thirty or forty years ago. They should take great pride in these advances. Included in these are analyses to assess mediation and moderation, methods of inquiry that were seldom thought about or practiced only a few decades earlier. Indeed, in these areas, it is social and personality psychologists who have been leading others in the refinement and use of these tools. These are tools of great potential, and their further refinement will continue to be one of our great contributions.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

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. *Sociological Methodology*, 20, 115–140.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology*, 98, 550–558.

Busemeyer, J. R., & Jones, L. (1983). Analysis of multiplicative comfination rules when the causal variables are measured with error. *Psychological Bulletin*, 93, 549–562.

Cheung, M. W. L. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods*, 41, 425–438.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, 70, 426–443.

Cohen, J. (1978). Partialed products are interactions; partialed powers are curve components. *Psychological Bulletin*, 85, 858–866.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7, 249–253.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 304–312.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression /correlation analyses for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, 47, 1231–1236.

Friedrich, R. J. (1982). In defense of mutliplicative terms in multiple regression equations. *American Journal of Public Health*, 26, 797–833.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18, 233–239.

Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 47, 61–87.

Hayes, A. F., & Mathes, J. (2009). Computational procedures for probing interactions in OLS and logist regression: SPSS and SAS implementations. *Behavior Research Methods*, 41, 924–936.

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195–222). Thousand Oaks, CA. Sage.

Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, 40, 366–371.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression*. Newbury Park, CA: Sage.

Jacoby, J., & Sassenberg, K. (2011). Interactions do not only tell us when, but can also tell us how: Testing process hypotheses by interaction. *European Journal of Social Psychology*, 41, 180–190.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, 69, 307–321.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602–619.

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.

Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp.

180–232). New York: McGraw-Hill.

Judd, C. M., McClelland, G. H., & Culhane, S. E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology*, 46, 433–465.

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2009). *Data analysis: A model comparison approach* (2nd ed.). New York: Routledge.

Kraemer, H. C., Kiernan, M., Essex, M., & Kupfer, D. J. (2008). How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychology*, 27, S101–S108.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, 23, 418–444.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, 36, 249–277.

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology*, 101, 1174–1188.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19–40.

MacKinnon, D. (2008). *Introduction to statistical mediation analyses*. Hillsdale, NJ: Lawrence Erlbaum.

MacKinnon, D., Fritz, M., Williams, J., & Lockwood, C. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, 39, 384–389.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, 1, 173–181.

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.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39, 99–128.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, 113, 181–190.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114, 376–390.

Muller, D., & Butera, F. (2007). The focusing effect of self-evaluation threat in coaction and social comparison. *Journal of Personality and Social Psychology*, 93, 194–211.

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

Pleyers, G., Corneille, O., Yzerbyt, V., & Luminet, O. (2009). Evaluative conditioning may incur attentional costs. *Journal of Experimental Psychology: Animal Behavior Processes*, 35, 279–285.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437–448.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879–891.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42, 185–227.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.

Rosnow, R. L., & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin*, 105, 143–146.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, 5, 359–371.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 350–353.

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.

Sobel, M.E. (1982). Asymptotic confidence intervals for indirect effects in structural equation modeling. In S. Leinhart (Ed.), *Sociological methodology 1982* (pp. 291–312). San Francisco: Jossey-Bass.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89, 845–851.

Toma, C., Corneille, O., & Yzerbyt, V. Y. (2012). Holding a mirror up to the self: Egocentric similarity beliefs underlie social projection in cooperation. *Personality and Social Psychology Bulletin*, 38, 1259–1271.

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work. *American Psychologist*, 24, 83–91.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin, 109*, 524–536.

Whisman, M. A., & McClelland, G. H. (2005). Designing, testing, and interpreting interactions and moderator effects in family research. *Journal of Family Psychology*, 19, 111–120.

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

[1] Throughout this chapter, for notational simplicity, we express models in terms of parameter estimates rather than the parameters themselves. Estimates may be generated by different estimating procedures in the context of different assumptions about the variables in the models. Most typically they will be least-squares estimates. But, in the case of latent variable models, logistic models involving dichotomous outcomes, or mixed models involving hierarchical levels, estimates will generally be obtained by some maximum likelihood estimation procedure.

[2] Importantly this equality holds regardless of rescaling of the component variables. Hence, it is also found with standardized estimates.

[3] Establishing that this represents a *causal* flow requires a set of very specific assumptions that we detail later in this section.

[4] In the following, when we talk of suppression, we are doing so within the confines of the causal model underlying mediation, in which some variable M is affected by X and in turn affects Y. Suppression has a broader meaning, i.e., whenever controlling for a third variable augments an effect of interest, outside of the specific causal model that we are assuming (e.g., Tzelgov & Henik, 1991).

[5] It is possible that some tests of the *ab* product may yield nonsignificant results even when the two component slopes are significant. This may happen when inappropriate assumptions are made in testing the *ab* product (i.e., that its sampling distribution is normal).

[6] Because this mediator is dichotomous rather than continuous, one should be doing estimation of these models using a logistic link function or logistic regression. To keep things simple, however, we chose for this example to act as if the mediator was continuous.

[7] Our definition of moderation represents a departure from Baron and Kenny's (1986) definition where they equate it with a statistical interaction. We explicitly assume a particular causal model in defining moderation. As such, our definition agrees more closely with recent work devoted to moderation in the context of

randomized trials in health outcomes (e.g., Kraemer, Kiernan, Essex, & Kupfer, 2008), where moderation assumes the existence of some causal effect that is moderated.

[8] The interaction tests whether simple slopes vary, and it is possible that they significantly vary even though particular simple slopes do not differ from zero.

[9] This final result will not in general be the case if one simply included the product predictor but not its component variable as predictors. It is for this reason that the slope of the product predictor estimates the effect of the interaction only when it is a partialled product, controlling for the two component variables.

[10] Of course, one cannot practically oversample the four corners of the joint distribution if one does not already know respondents' values on those variables. Typically, however, one can identify demographic variables that are likely to covary with the variables, and then the oversampling might be based on those demographics.

[11] Also note that researchers sometime dichotomize continuous predictors and argue that it is fine as long as their effects are significant. Because dichotomizing continuous predictors can sometime increase Type-1 errors (Maxwell & Delaney, 1993), we do not recommend this kind of reasoning.

[12] This assumes that the self-ratings have been centered at level 1 around the participant's mean.