Journal of Personality and Social Psychology
2005, Vol. 89, No. 6, 852–863

0022-3514/05/$12.00 DOI: 10.1037/0022-3514.89.6.852

When Moderation Is Mediated and Mediation Is Moderated

Dominique Muller
Université Paris Descartes, France

Charles M. Judd
University of Colorado at Boulder

Vincent Y. Yzerbyt
Catholic University of Louvain at Louvain-la-Neuve, Belgium

Procedures for examining whether treatment effects on an outcome are mediated and/or moderated have been well developed and are routinely applied. The mediation question focuses on the intervening mechanism that produces the treatment effect. The moderation question focuses on factors that affect the magnitude of the treatment effect. It is important to note that these two processes may be combined in informative ways, such that moderation is mediated or mediation is moderated. Although some prior literature has discussed these possibilities, their exact definitions and analytic procedures have not been completely articulated. The purpose of this article is to define precisely both mediated moderation and moderated mediation and provide analytic strategies for assessing each.

Keywords: mediation, moderation, mediated moderation, moderated mediation

There is by now a very large literature on the important topics of mediation and moderation (e.g., Aiken & West, 1991; Baron & Kenny, 1986; James & Brett, 1984; Judd & Kenny, 1981; Kenny, Kashy, & Bolger, 1998; Kraemer, Wilson, Fairburn, & Agras, 2002; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) and in general it is fair to say that the analytic strategies underlying the assessment of each are well understood (even if they are still developing). Both processes focus on a given treatment effect. The issue of mediation addresses how that treatment effect is produced. Mediational analyses attempt to identify the intermediary process that leads from the manipulated independent variable to the outcome or dependent variable. The issue of moderation focuses on factors that influence the strength and/or direction of the relation between the treatment variable and the dependent variable. Moderational analyses attempt to identify individual difference or contextual variables that strengthen and/or change the direction of the relationship between the treatment variable and the dependent variable.

In this growing literature, there have been occasional discussions of how the two issues of mediation and moderation might themselves be combined in theoretically interesting ways. Thus, in some of the early classic papers on mediation and moderation, James and Brett (1984) discuss moderated mediation, and Baron and Kenny (1986) devote a page to both moderated mediation and mediated moderation. More recently Wegener and Fabrigar (2000) also discuss these two topics. However, all of these treatments are relatively brief, and our perusal of the literature that has made reference to one or the other of these topics suggests that they are not clearly understood. Additionally, there exists no source that comprehensively articulates the definition of these two processes and the analytic models that underlie them.

It is our intention in the current article to accomplish this. In the initial section, we clearly define both mediated moderation and moderated mediation and provide examples of each. In the second section, we lay out the analytic models that are used to examine both mediated moderation and moderated mediation. As we show there, both of these processes rely on the same underlying models of the data, but both have different starting points and thus respond to distinct research questions. There is, however, a fundamental defining algebraic equality that underlies both processes. Accordingly, while we explicate the analyses for each, our argument is that these two processes are in some sense the flip sides of the same coin. In the third section, we use hypothetical data from the two examples developed in the initial section to illustrate both processes. We conclude by further discussing how these processes relate to each other and then by considering a variety of issues which our discussion has raised.

Defining Terms and Illustrating Processes

Initially we define terms in an abstract, generalized case. Subsequent to this, we provide more detailed substantive examples that hopefully will make clear why both processes, mediated moderation and moderated mediation, are theoretically important.

The abstract case initially focuses on four variables. First there is the manipulated independent variable, X_i. To keep things simple,

Dominique Muller, Université Paris Descartes, France; Charles M. Judd, University of Colorado at Boulder; Vincent Y. Yzerbyt, Catholic University of Louvain at Louvain-la-Neuve, Belgium.

This work was supported by a Lavoisier fellowship from the Ministère des Affaires étrangères, France, awarded to Dominique Muller and by National Institute of Mental Health Grant R01MH45049 awarded to Charles M. Judd. We express our gratitude to Olivier Corneille, Olivier Desrichard, Marcello Galucci, and David MacKinnon for their comments on a draft of this article.

Correspondence concerning this article should be addressed to Dominique Muller, Laboratoire de Psychologie Sociale, Université Paris-5 René Descarte, 71, avenue Edouard Vaillant, 92774 Boulogne Billancourt Cedex, France, or to Charles M. Judd, Department of Psychology, University of Colorado, Boulder, CO 80309-0345. E-mail: dominique.muller@univ-paris5.fr or charles.judd@colorado.edu

we assume it has only two values, indicating whether a participant was in the treatment or control condition, and its values are contrast coded (e.g., −1, +1 or −.5, +.5). We further assume that participants have been randomly assigned to one of these two levels, so that causal inferences can be made about the treatment effect.[1] The second variable is the outcome variable, Y_i, which is some measured response of the participants, presumed to be affected by the treatment. Next there is a mediating variable, Me_i, which is another measured response variable, also expected to be affected by the treatment variable. The mediational hypothesis, to be examined, is that this mediator is responsible for the causal effect of X_i on Y_i. Finally, there is a potential moderating variable, Mo_i, which is either some stable individual difference variable, assumed not to be affected by the treatment, or else some measure of the context or situation under which the treatment is delivered. We presume it is measured prior to the delivery of the treatment, and therefore, given random assignment to X_i, it is assumed that Mo_i and X_i are independent of each other.[2] In the general case, we assume that this moderator is continuously measured, although what we have to say also applies with a dichotomous, and hence contrast-coded, moderator. The moderational hypothesis, to be examined, is that the magnitude of the causal effect of X_i on Y_i depends on the moderator.

It is important to note that, throughout our discussion (unless explicitly stated otherwise), we are presuming that all variables (with the exception of the outcome) have been centered at their mean. This assumption is made to increase the interpretability of various parameters in models that include interaction terms (Aiken & West, 1991; Judd & McClelland, 1989).[3]

Although there is some continuing discussion about the necessary and sufficient conditions to establish mediation (see below), we have chosen to adopt the original and classic approach as outlined by Judd and Kenny (1981) and Baron and Kenny (1986). According to this approach, to demonstrate mediation, one estimates three different models (typically using least squares estimators)[4]

$$Y = \beta_{10} + \beta_{11}X + \varepsilon_1 \tag{1}$$

$$Me = \beta_{20} + \beta_{21}X + \varepsilon_2 \tag{2}$$

$$Y = \beta_{30} + \beta_{31}X + \beta_{32}Me + \varepsilon_3 \tag{3}$$

and four conditions must be met:

1. In Equation 1, there must be an overall treatment effect on the outcome variable; that is, b_{11} is significant.[5]

2. In Equation 2, there must be a treatment effect on the mediator; that is, b_{21} is significant.

3. In Equation 3, there must be an effect of the mediator on the outcome controlling for the treatment; that is, b_{32} is significant.

4. In Equation 3, the residual direct effect of the treatment variable on the outcome (β_{31}) should be smaller (in absolute value) than the overall treatment effect in Equation 1 (β_{11}).

The following equality relationship exists among the parameters of these models (see, e.g., MacKinnon, Warsi, & Dwyer, 1995):

$$\beta_{11} - \beta_{31} = \beta_{21}{*}\beta_{32}$$

meaning that the difference between the overall treatment effect and the residual direct effect is equal to what is called the indirect effect via the mediator (i.e., $\beta_{21}{*}\beta_{32}$).

A continuing issue in the literature has focused on how one establishes Condition 4. Alternatives here, outlined by MacKinnon et al. (2002), involve testing whether the parameter difference on the left side of the above equality departs from zero or whether the product on the right side does so. An additional issue that continues to be debated concerns the necessity of Condition 1 (again, MacKinnon et al., 2002; Shrout & Bolger, 2002). We agree here with Shrout and Bolger (2002) who acknowledge that "experimentalists who wish to elaborate the mechanisms of an experimental effect need to first establish that the effect exists" (p. 430). We do, however, recognize that there may be grounds for disagreement here.

To demonstrate moderation, one estimates the following model:

$$Y = \beta_{40} + \beta_{41}X + \beta_{42}Mo + \beta_{43}XMo + \varepsilon_4 \tag{4}$$

where XMo is computed as the product of the treatment variable and the moderating variable. A test of the effect of that partialled product (i.e., the significance of b_{43}) is a test of the Treatment × Moderator interaction, asking whether the treatment effect varies in magnitude as a function of the value of the moderator.

In this context, having defined the relevant variables and the meaning of mediation and moderation, we are now able to define the two processes of central interest to this article: mediated moderation and moderated mediation. To clearly differentiate between them, we will define them in the most prototypic case, clearly contrasting the two. Subsequently, we will examine their interrelations and how one bleeds into the other.

The first of these, mediated moderation, can happen only when moderation occurs: the magnitude of the overall treatment effect on the outcome depends on the moderator. Given that the magnitude of the treatment effect depends on an individual difference or context variable, then the mediated moderation question is concerned with the mediating process that is responsible for that moderation. What is the process through which that overall moderated treatment effect is produced? An advantage of seeing me-

[1] In this regard we concur with recent arguments of Kraemer et al. (2002) who have the strong preference to confine discussion of mediation (and moderation) to experimental situations in which units are randomly assigned to the manipulated levels of the treatment variable.

[2] Again, we are consistent with Kraemer et al., (2002) in making these assumptions.

[3] In fact, the models we explicate do not depend on this centering assumption. The relevant variables could be deviated from values other than their means, with appropriate modifications in the interpretation of parameter estimates associated with lower-order predictor variables in models involving higher-order interactions. See Aiken and West (1991) and Judd and McClelland (1989).

[4] Throughout we will use β's to refer to unknown population parameters and b's to refer to their sample estimates. In all equations we omit the i subscript from all variables.

[5] Here and everywhere in the text, we use "overall effect" to refer to the direct plus the indirect effect of the independent variable on the outcome. Correspondingly, we use "residual direct effect" to refer to the direct effect of the independent variable on the outcome controlling for the mediator.

diation as a series of analytic steps (as outlined above) is that it permits one to recognize that there is more than one way by which an overall moderated treatment effect might be produced. We discuss these alternatives in the later section where we outline the analytic approach to mediated moderation.

Moderated mediation happens if the mediating process that is responsible for producing the effect of the treatment on the outcome depends on the value of a moderator variable. In other words, if the moderator is an individual difference variable, then it would mean that the mediating process that intervenes between the treatment and the outcome is different for people who differ on that individual difference. If the moderator is a contextual variable, then it would mean that the mediating process varies as a function of context. Note that this definition importantly implies mediation (at least for some people or in some contexts), as we defined it previously, but it does not imply any overall moderation of the treatment effect. And in fact, we will see that it is most convenient to clearly define moderated mediation in the prototypic case where there is no moderation of the X to Y effect. What varies as a function of the moderator is not the magnitude of the overall treatment effect on the outcome but the mediating process that produces it. Again, moderated mediation can happen in a number of different ways. We describe these in the section below where we discuss analytic models.

We now turn to illustrations of both mediated moderation and moderated mediation. These examples were chosen from existing social psychological research but they were constructed in such a way to illustrate both processes in prototypic cases. As such, they are reasonable models that might be examined in social psychological research, but they are deliberate oversimplifications.

To illustrate mediated moderation, consider recent work by Smeesters, Warlop, Van Avermaet, Corneille and Yzerbyt (2003) on the role of "morality" versus "might" primes on cooperative versus competitive behavior in a prisoner's dilemma choice scenario. They argued that participant's social value orientation (proself vs. prosocial) would moderate the impact of such primes. Specifically, they showed that for participants who were more prosocial, the primes affected the choice of cooperative versus competitive behaviors ("morality" primes increased cooperation, compared to "might" primes) whereas for more proself participants, this difference was not found. Their explanation for the mediating process underlying this overall moderation was that the primes produce expectations about how the partner in the prisoner's dilemma game would behave. The "morality" prime induced expectations that the partner would cooperate; the "might" prime induced expectations that the partner would compete. And the researchers reasoned that these primed expectations would then be acted upon differently by prosocial and proself participants. Those who are more prosocial should attempt to match their behavior to what they expect from the partner: if they expect competition, their choice should be competitive; if they expect cooperation, their choice should be cooperative. But among more proself participants, competitive choices should predominate regardless of the expectations: if they expect competition, their choice should be competition; and if they expect cooperation, they should also compete in an attempt to exploit their partner's cooperative choice. In sum, the researchers anticipated that the effect of the prime on behavioral choice would depend on whether participants were prosocial or proself (overall moderation of the treatment effect). And they further predicted that this moderation would be mediated by expectations about the partner's behavior. Primes would induce expectations about partner's behavior, that they would either cooperate or compete, and participants who were prosocial would match their own behavior to their partner's whereas those who were more proself would compete regardless of their expectation.

To illustrate moderated mediation, we draw on research by Petty, Schumann, Richman, and Strathman (1993) that examined the role of positive mood in persuasion. They manipulated participants' mood (either positive mood or no mood induction) and subsequently exposed them to counterattitudinal persuasive information. They predicted that those in a positive mood would show more persuasion than those in the control condition. And overall they found this treatment effect. But they further were interested in understanding differences in what mediates this overall treatment effect as a function of a "need for cognition" individual difference variable. They argued that for those high in need for cognition, the mediating process producing more persuasion as a result of a positive mood would be one where the mood causes people to generate more positively valenced thoughts in response to the persuasive communication and then these thoughts in turn produce greater persuasion. On the other hand, they argued that for those participants who were low in need for cognition, the mediating process would not be through positively valenced thoughts. In other words, positive thoughts would mediate the mood—persuasion effect for those high in need for cognition, while for those low in need for cognition the same mood—persuasion effect would not be mediated by positive thoughts (or would be mediated by them less). This illustrates moderated mediation: a treatment effect is mediated differently as a function of some moderator variable. Need for cognition moderates the way in which the mood—persuasion effect is produced. Note that there is no overall moderation here: it is not the case that the mood—persuasion effect is larger or smaller for those who differ in need for cognition. Rather the mediating process for the same effect was hypothesized to be different.[6]

Analytic Models for Moderated Mediation and Mediated Moderation

There are three fundamental models that underlie both mediated moderation and moderated mediation.[7] We have already given the

[6] Petty et al. (1993) note that had there been an overall moderation of the mood effect by need for cognition, the assessment of moderated mediation would have been problematic. Finding a pattern by which, for instance, the effect of mood was absent for those high in need for cognition simply "precludes examination of any differential mediation of the positive mood effect for people high and low in NC because there was no effect of positive mood to mediate for the high-NC individuals." (p. 9) Although not intractable, we would agree that interpretational problems arise in talking about moderated mediation whenever the magnitude of the treatment effect on the outcome variable varies as a function of the moderator.

[7] The models that we give here appear somewhat different from those proposed by Baron and Kenny (1986, p. 1179) in their brief treatment of mediated moderation and moderated mediation. This is because in their example the moderator was a manipulated independent variable and it was then subsequently measured (as a manipulation check) to serve as a mediator. Our approach, utilizing conceptually different moderator and mediator variables and a single set of equations, is more parsimonious.

first of these as Equation 4, to assess moderation of the overall treatment effect:

$$Y = \beta_{40} + \beta_{41}X + \beta_{42}Mo + \beta_{43}XMo + \varepsilon_4 \quad (4)$$

This model allows the overall treatment effect of Equation 1 to be moderated by *Mo*. The second model allows the treatment effect on the mediator, in Equation 2, to be moderated:

$$Me = \beta_{50} + \beta_{51}X + \beta_{52}Mo + \beta_{53}XMo + \varepsilon_5 \quad (5)$$

And the third model is a moderated version of Equation 3, in which both the mediator's (partial) effect on the outcome and the residual effect of the treatment on the outcome, controlling for the mediator, are allowed to be moderated:

$$Y = \beta_{60} + \beta_{61}X + \beta_{62}Mo + \beta_{63}XMo + \beta_{64}Me + \beta_{65}MeMo + \varepsilon_6 \quad (6)$$

In all three models, we are making the same assumptions about all variables that we did earlier. Namely, there are two levels of X (contrast coded) to which experimental units have been randomly assigned, X and Mo are therefore uncorrelated, and both Me and Mo have been centered.[8] Given these assumptions, the slope parameters in these three models can be interpreted as defined in Table 1.

In Figure 1, we present the fundamental model that subsumes both mediated moderation and moderated mediation. The top panel in this Figure represents the overall treatment effect on the outcome, and the extent to which it is moderated. Accordingly, the overall treatment effect depends on the level of the moderator, to the extent that β_{43} departs from zero. The bottom panel presents the mediated model, allowing the treatment effect on the mediator to be moderated (to the extent that β_{53} departs from zero), the mediator's partial effect on the outcome to be moderated (to the extent that β_{65} departs from zero), and the residual direct effect of the treatment on the outcome (controlling for the mediator) to be moderated (to the extent that β_{63} departs from zero).

Table 1
Interpretation of the Slope Parameters in Equations 4, 5, and 6

Slope parameters	Interpretation of slope parameters
β_{41}	Overall treatment effect on Y at the average level of Mo
β_{42}	Moderator effect on Y on average across the two treatment levels
β_{43}	Change in overall treatment effect on Y as Mo increases
β_{51}	Treatment effect on Me at the average level of Mo
β_{52}	Moderator effect on Me on average across the two treatment levels
β_{53}	Change in treatment effect on Me as Mo increases
β_{61}	Residual direct treatment effect on Y at the average level of Mo
β_{62}	Moderator effect on Y on average within the two treatment levels and at the average level of Me
β_{63}	Change in residual direct treatment effect on Y as Mo increases
β_{64}	Mediator effect on Y on average within the two treatment levels and at the average level of Mo
β_{65}	Change in mediator effect on Y as Mo increases

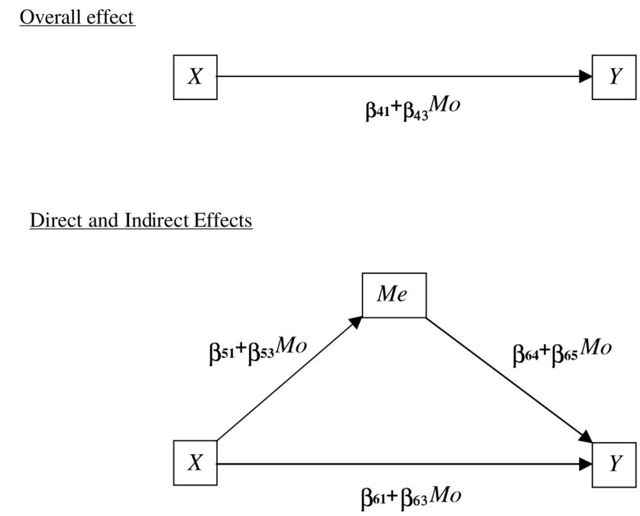

Figure 1. Models illustrating moderated mediation and mediated moderation.

From this Figure, the overall (moderated) treatment effect is $\beta_{41} + \beta_{43}Mo$. Using the estimated values for these parameters, one can use this expression to estimate "simple" overall treatment effects at particular levels of Mo. The (moderated) indirect effect, via the mediator, equals $(\beta_{51} + \beta_{53}Mo)(\beta_{64} + \beta_{65}Mo)$. And the residual (moderated) treatment effect equals $\beta_{61} + \beta_{63}Mo$. Again, the parameter estimates from these expressions can be used to estimate "simple" effects at particular levels of Mo.

Parallel to what we saw in the case of simple mediation (Equations 1–3), there exists a fundamental equality among the parameters of these models, focusing on the moderation of the indirect and residual direct effects. In the Appendix, we prove that at the level of the population parameters, making the assumptions we have made, the following equality holds:

$$\beta_{43} - \beta_{63} = \beta_{64}\beta_{53} + \beta_{65}\beta_{51} \quad (7)$$

This equality will not exactly hold in terms of parameter estimates, from a sample of data, unless the moderator is dichotomous and contrast coded (see Appendix). Nevertheless, this equality provides us with fundamental insights into the definitions of mediated moderation and moderated mediation.

We now turn to these definitions and what they imply about the three models we have just identified. Doing this provides us with analytic strategies for estimating both mediated moderation and moderated mediation. And the consideration of these processes, in the context of these models, will provide insights about the different ways in which each process might be produced.

Mediated Moderation

The above Models 4 through 6 are used to establish mediated moderation. Equation 7 establishes an equality condition on the

[8] When the parameters of these models are estimated in a sample of data, additional assumptions are necessary, of course, to derive their standard errors. Namely, residuals must be independent, normally distributed, and have a common variance.

parameters from these models. One could first focus on the two parameters on the left side of this equality, that is, β_{43} and β_{63}. With mediated moderation, there is overall moderation of the treatment effect, that is, $\beta_{43} \neq 0$, and the question then is whether the mediating process accounts for this moderation. If it does, then the moderation of the residual direct effect of the treatment should be reduced compared to the moderation of the overall treatment effect, that is, β_{63} should be smaller in absolute value than β_{43}. For this to be the case, we see on the right side of Equation 7 that there must be mediation and one or both of the indirect paths from the treatment to the outcome must be moderated. That is, either the effect of *X* on *Me* depends on the moderator ($\beta_{53} \neq 0$, and the average partial effect of *Me* on *Y* [β_{64}] is nonzero) and/or the partial effect of *Me* on *Y* depends on the moderator ($\beta_{65} \neq 0$, and the average effect of *X* on *Me* [β_{51}] is nonzero).

In light of this, to demonstrate mediated moderation in a sample of data, one estimates Models 4 through 6. In Model 4, we would expect b_{43} to be significant, indicating overall treatment moderation. In Models 5 and 6, either (or both) of two patterns should exist: both b_{53} and b_{64} are significant or both b_{51} and b_{65} are significant. And, as a result, the moderation of the residual treatment effect, b_{63}, should be reduced in magnitude (and may be nonsignificant in the case of what might be called "full" mediated moderation) compared to the moderation of the overall treatment effect.

Moderated Mediation

In the prototypic case of moderated mediation, there is an overall treatment effect (β_{41}) and the magnitude of this effect does not depend on the moderator ($\beta_{43} = 0$). However, the potency of the mediating process depends on the moderator. Accordingly, either the effect of the treatment on the mediator depends on the moderator ($\beta_{53} \neq 0$) or the partial effect of the mediator on the outcome depends on the moderator ($\beta_{65} \neq 0$), or both. Parallel to this, if the treatment effect on the mediator depends on the moderator ($\beta_{53} \neq 0$), then there must be a partial effect of the mediator on the outcome on average ($\beta_{64} \neq 0$), or if the partial effect of the mediator on the outcome depends on the moderator ($\beta_{65} \neq 0$), then there should be an overall treatment effect on the mediator ($\beta_{51} \neq 0$). In other words, at least one of the products on the right-hand side of the equality in Equation 7 must depart from zero.

Accordingly, if the right-hand side of the equality in Equation 7 departs from zero, then too must the left-hand side parameter difference. We have assumed that the first term in this difference equals zero (i.e., $\beta_{43} = 0$, there is no overall moderation of the treatment effect). This implies that the second term (β_{63}) differs from zero.

In sum, moderated mediation implies that the indirect effect between the treatment and the outcome depends on the moderator. That is, either the effect from *X* to *Me* depends on the moderator ($\beta_{53} \neq 0$, and the average partial effect of *Me* on *Y* [β_{64}] is nonzero) and/or the partial effect of *Me* on *Y* depends on the moderator ($\beta_{65} \neq 0$, and the average effect of *X* on *Me* [β_{51}] is nonzero). In either case, given no overall moderation of the treatment effect (i.e., $\beta_{43} = 0$), this implies that the residual direct treatment effect on the outcome, controlling for the mediator, is moderated (i.e., $\beta_{63} \neq 0$).[9]

In light of this, to demonstrate moderated mediation in a sample of data, one estimates Models 4 through 6. In Model 4, the prototypic case leads to the expectation that b_{41} is significantly different from zero, while b_{43} is not. In Models 5 and 6, either (or both) of two patterns should exist: both b_{53} and b_{64} are significant or both b_{51} and b_{65} are significant. A consequence is that the residual treatment effect should now be moderated, that is b_{63} may be significant. Although this is likely, we do not believe that the significance of b_{63} should be seen as a necessary condition for establishing moderated mediation.

Summary

In sum, for mediated moderation, there is overall moderation, produced by the mediating process, and when this process is controlled, the residual moderation of the treatment effect is reduced. What we have called prototypic moderated mediation is found when there is an unmoderated overall treatment effect, but the indirect effect of the treatment via the mediator is moderated. As a result, the residual treatment effect is likely to be moderated.

Data-Based Examples

In this section, we return to the two examples we used earlier to illustrate mediated moderation and moderated mediation, but this time with hypothetical data[10] and illustrative model estimates from those data.

Mediated Moderation Example

The example that we used to illustrate the definition of mediated moderation focused on cooperative behavior in a prisoner's dilemma situation (Smeesters et al., 2003). Participants were primed with either "might" or "morality" primes and then engaged in a one-trial prisoner's dilemma with a fictitious partner. Additionally, participants' social value orientation (from proself to prosocial) was measured. The researchers found an overall moderation of the effect of the prime by social value orientation: those participants who were more prosocial indicated they would more likely cooperate under the "morality" prime while they were more likely to compete under the "might" prime. On the other hand, more proself participants tended to choose competition regardless of the prime. It was then suggested that this moderation was produced by expectations about the partner. More specifically, the "morality" prime led to expectations that the partner would cooperate, while the "might" prime led to expectations of partner competition. And participants proself versus prosocial status was found to moderate the impact of these partner expectations on behavior: Prosocial participants attempted to match the behaviors expected from their

[9] In rare cases, it is possible that the effect of the treatment on the mediator is moderated in one direction while the effect of the mediator on the outcome is moderated in the opposite direction. In these cases, the overall moderation of the treatment effect may be the same as the moderation of the residual direct treatment effect.

[10] We used simulated data instead of real data to keep things parsimonious. Additionally, they are publicly available, as indicated below, so that the reader can recreate our analyses.

partner, while proself participants competed regardless of how they expected their partner to behave.

For this illustration, data were generated for 100 cases on four variables:[11]

(a) A dichotomous treatment variable indicating the type of prime received (*X*: referred to as PRIME): "might" priming (-1) versus "morality" priming ($+1$). Values on this variable were randomly assigned to the 100 cases.

(b) A continuous moderator variable (*MO*: referred to as SVO for social value orientation, with lower numbers for more proself participants and higher numbers for those who are more prosocial): randomly generated from a normal distribution with a mean of zero and a standard deviation of 1.35. This variable was centered in the sample. Sample values ranged from -3.369 to $+ 3.833$.

(c) A continuous mediator variable (*ME*: referred to as EXP for expectations about the partner's behavior, with higher numbers indicating that the partner was expected to be more cooperative). It was generated to be a function of prime, adding a random error component. Again, it was centered in the sample and had a standard deviation of 7.913.

(d) A continuous outcome variable (*Y*: referred to as BEH for behavior, with higher numbers indicating a greater probability of cooperative behavioral choices). This outcome variable was constructed to be affected by the prime, social value orientation, expectations, and the Mediator $\times$ Moderator interaction, plus a random error component.

As just described, these data were generated to be consistent with the mediated moderation hypothesis, that is, that the overall effect of the treatment (PRIME) on the outcome (BEH) would be moderated by social value orientation (SVO) and that this interaction would be due to the effect of the treatment on the mediator (EXP) and the moderation of the mediator effect on the outcome (BEH) by social value orientation (SVO).

Table 2 contains the univariate statistics and bivariate correlations for all four variables. Table 3 presents the regression models that estimate Equations 4 through 6 with these variables. Presented here are the unstandardized coefficients (*b*) and their associated *t* statistics. In these equations, there are product predictors included to estimate the interactions (PRIMESVO = PRIME*SVO; EXPSVO = EXP*SVO).

The results from Equation 4 find evidence for the predicted interaction between PRIME and SVO on the outcome variable, BEH. As predicted, "morality" primes are associated with a higher probability of cooperative behaviors than "might" primes, and this difference increases as participants are more prosocial and less proself. The treatment variable's (PRIME) effect on the outcome (BEH) is moderated by social orientation (SVO).

Table 2
Univariate and Bivariate Statistics for Mediated Moderation Example

Variable	PRIME (Treatment)	SVO (Moderator)	EXP (Mediator)	BEH (Outcome)
M	.000	.000	.000	58.171
SD	1.005	1.398	7.913	14.691
Correlations				
PRIME	—	.004	.342**	.314**
SVO		—	−.016	−.165
EXP			—	.385
BEH				—

Note. SVO = social value orientation; EXP = expectations about partner's behavior; BEH = behavior.
** $p < .01$.

The results from Equation 5 show an effect of the prime on the mediator, that is expectations about the partner's behavior (EXP). Moreover, this effect is not moderated by social orientation (the PRIMESVO coefficient is not significant). In order to help with the interpretation of the indirect effect in these data it is informative to calculate the simple effects of the prime (PRIME) on the mediator (EXP) at different levels of the moderator (SVO), as discussed earlier. Because the moderator varies continuously, we have calculated these simple effects at values of one standard deviation (1.398) above and below the mean SVO score:

$$\text{For High SVO } (+1\ SD)\text{: } b_{51} + b_{53}Mo = 2.692 + 0.089(1.398) = 2.816$$

$$\text{For Low SVO } (-1\ SD)\text{: } b_{51} + b_{53}Mo = 2.692 + 0.089(-1.398) = 2.568$$

The estimation of Equation 6 reveals a significant effect of the mediator (EXP) by moderator (SVO) interaction on the outcome (the EXPSVO coefficient is significant). We can again calculate the simple effect of EXP on BEH at values of SVO one standard deviation above and below the mean SVO score:

$$\text{For High SVO } (+1\ SD)\text{: } b_{64} + b_{65}Mo = 0.840 + 0.765(1.398) = 1.909$$

$$\text{For Low SVO } (-1\ SD)\text{: } b_{64} + b_{65}Mo = 0.840 + 0.765(-1.398) = -0.229$$

Taking the product of the two simple effects for each of the two values of SVO, we get the total indirect effects through the mediator (EXP) for each of these two values:

$$\text{For High SVO } (+1\ SD)\text{: } 2.816*1.909 = 5.377$$

$$\text{For Low SVO } (-1\ SD)\text{: } 2.568* -0.229 = -0.589$$

Accordingly, for all participants, regardless of social orientation, the cooperative prime increases expectations that the partner will cooperate. And such expectations lead to cooperative behavior on the part of the participants but only in the case of those who are high (prosocial) on social value orientation. It would seem that the moderation of the overall effect of the prime (by social value orientation) can be understood from the fact that the cooperative prime induces cooperative expectations among all participants, but those expectations translate into cooperative behavior only for those who have a cooperative social value orientation.

Consistent with this, Equation 6 also reveals that the residual direct effect of the prime on the outcome (BEH) is less (and actually not at all in this example) moderated by social value orientation, once the mediator (EXP) and its interaction with SVO

[11] The raw data are available at http://www.psp.ucl.ac.be/mediation/

Table 3
Least Squares Regression Results for Mediated Moderation Example

Predictors	Equation 4 (criterion BEH)		Equation 5 (criterion EXP)		Equation 6 (criterion BEH)	
	b	t	b	t	b	t
X: PRIME	4.580 (b_{41})	3.40**	2.692 (b_{51})	3.57**	2.169 (b_{61})	2.03*
MO: SVO	−2.042 (b_{42})	2.09*	−0.085 (b_{52})	−0.16	2.569 (b_{62})	3.54
XMO: PRIMESVO	2.574 (b_{43})	2.64**	0.089 (b_{53})	0.16	0.041 (b_{63})	0.05
ME: EXP					0.840 (b_{64})	6.05**
MEMO: EXPSVO					0.765 (b_{65})	7.91**

BEH = behavior; EXP = expectations about partner's behavior; MO = moderator variable; SVO = social value orientation; ME = mediator variable.
* $p < .05$. ** $p < .01$.

are controlled. Indeed, the coefficient associated with the PRIME * SVO interaction has been reduced from 2.508 (in Equation 4) to 0.041 (in Equation 6). Once we control for the mediator and allow the indirect effect via the mediator to be moderated, the residual direct effect of the prime on the outcome no longer depends on this moderator.

As we made clear earlier, mediated moderation may be produced in either (or both) of two ways. In the first way, the moderator affects the magnitude of the treatment effect on the mediator (and this is found in conjunction with a mediator effect in Equation 6). Alternatively, and this is the situation illustrated in our example, the moderator affects the magnitude of the mediator's partial effect on the outcome (and this is found in conjunction with a treatment effect on the mediator in Equation 5).

Moderated Mediation Example

We illustrate this analysis with data based loosely on the moderated mediation example we highlighted in the introduction (Petty et al., 1993). Recall that in this example the effect of positive mood (vs. control) on persuasion was thought to be more mediated by positive valenced thoughts in the case of individuals who are high in "need for cognition" than in the case of individuals low in "need for cognition." For this illustration data were generated for 100 cases on four variables:[12]

(a) A dichotomous treatment variable *(X*: referred to as MOOD): positive mood induction (+1) versus no mood induction (−1). Values on this variable were randomly assigned to the 100 cases.

(b) A continuous moderator variable (*MO:* referred to as NFC for need for cognition scores with higher numbers for higher need for cognition): randomly generated from a normal distribution with a mean of zero and a standard deviation of 1.35. This variable was centered in the sample. Sample values ranged from −4.82 to + 3.07.

(c) A continuous mediator variable (*ME*: referred to as POS for positive valenced thoughts, with higher numbers indicating more positive valenced thoughts). This is the mediator. It was generated to be a function of the treatment variable and its interaction with NFC, adding in a random error component. Again it was centered in the sample and ranged from −18.05 to + 21.74.

(d) A continuous outcome variable (*Y*: referred to as ATT for attitude change, with higher numbers indicating more attitude change). This outcome variable was constructed to be affected by the treatment variable and the mediator, plus a random error component.

As just described, these data were constructed following the theoretical model of Petty et al. (1993), so that there was an overall treatment effect on attitude change, unmoderated by need for cognition. For those high in need for cognition, the treatment variable affected the mediator (positive valenced thoughts) more than for those low in need for cognition, while this mediator affected the outcome equally for all participants. Unlike Petty et al. (1993), we assumed that need for cognition scores were continuously measured rather than dichotomizing the variable at its median.

Table 4 contains univariate statistics and bivariate correlations for all four variables. Table 5 presents the regression models that estimate Equations 4 through 6 with these variables. Presented here are the unstandardized coefficients (b) and their associated t statistics. In these equations, product predictors are included to estimate the interactions (MOODNFC = MOOD*NFC; POSNFC = POS*NFC).

The results from Equation 4 indicate an overall effect of the treatment, MOOD, on the outcome variable, ATT. This effect is not moderated by need for cognition, NFC.

In Equation 5, the mediator, POS, is the criterion. Here, there is a significant effect of MOOD and a significant MOOD × NFC interaction. This significant interaction is indicative of moderated mediation, in that it means that the magnitude of the indirect effect of MOOD, via the mediator, varies in magnitude as a function of NFC. As in the previous example, it is useful to again calculate the simple effects of MOOD on POS at one standard deviation (1.405) above and below the NFC mean:

$$\text{For High NFC } (+1\ SD)\text{: } b_{51} + b_{53}Mo = 4.336 + 1.256(1.405) = 6.101$$

[12] The raw data are available at http://www.psp.ucl.ac.be/mediation/

Table 4
Univariate and Bivariate Statistics for Moderated Mediation Example

Variable	MOOD (Treatment)	NFC (Moderator)	POS (Mediator)	ATT (Outcome)
M	.000	.000	.000	1.98
SD	1.005	1.405	8.322	16.79
Correlations				
MOOD	—	−.023	.521**	.405**
NFC		—	.068	.110
POS			—	.629**
ATT				

Note. NFC = need for cognition; POS = positive valenced thoughts; ATT = attitude change.
* $p < .05$. ** $p < .01$.

$$\text{For Low NFC } (-1\ SD)\text{: } b_{51} + b_{53}Mo = 4.336 + 1.256(-1.405) = 2.571$$

For someone well above the mean on need for cognition, positive mood, compared to control, leads to a considerably higher POS score. On the other hand, for someone well below the mean on the moderator, there is less of an effect of MOOD on POS.

From the Equation 6 results, importantly we see that there is a significant effect of the mediator (POS) on the outcome. Moreover, this is not moderated by need for cognition (the POSNFC coefficient is not significant). In order to help interpretations we can however, again calculate the simple effects of the mediator on the outcome at values of NFC one standard deviation above and below the mean:

$$\text{For High NFC } (+1\ SD)\text{: } b_{64} + b_{65}Mo = 1.248 - .036(1.405) = 1.197$$

$$\text{For Low NFC } (-1\ SD)\text{: } b_{64} + b_{65}Mo = 1.248 - .036(-1.405) = 1.299$$

Taking the product of the two simple effects, once at one standard deviation above the mean NFC and once at one standard deviation below the mean, we get the total indirect effects at the two values:

$$\text{For High NFC } (+1\ SD)\text{: } 6.101*1.197 = 7.303$$

$$\text{For Low NFC } (-1\ SD)\text{: } 2.571*1.299 = 3.340$$

Equation 6 also reveals, as might be expected, that the residual direct effect of the treatment (MOOD) on the outcome is moderated, once the mediator is controlled. That is, there is a significant effect of the MOOD × NFC interaction. Since the overall effect of the treatment on the outcome does not vary as a function of NFC (from Equation 4), and since the indirect effect via the mediator does vary as a function of NFC, then it is the case that the residual direct effect, controlling for the mediator, is moderated by NFC. Again, we can calculate the two simple residual treatment effects at the two levels of NFC (plus and minus one standard deviation from the mean):

$$\text{For High NFC } (+1\ SD)\text{: } 1.480 - 2.169(1.405) = -1.567$$

$$\text{For Low NFC } (-1\ SD)\text{: } 1.480 - 2.169(-1.405) = 4.527$$

These results as a whole make very clear that the indirect effect, via the mediator, is much higher when NFC is high rather than low, while the residual direct effect is much higher when NFC is low rather than high. This pattern is what is expected under prototypical moderated mediation.

This example illustrates a case where there is moderated mediation because the effect of the treatment on the mediator depends on the moderator. As a result, the overall magnitude of the indirect effect via the mediator depends on the moderator. Alternatively, as we made clear previously, moderated mediation could also happen when the effect of the mediator on the outcome depends on the moderator, producing a moderated indirect effect in a different manner.

Table 5
Least Squares Regression Results for Moderated Mediation Example

	Equation 4 (Criterion ATT)		Equation 5 (Criterion POS)		Equation 6 (Criterion ATT)	
Predictors	*b*	*t*	*b*	*t*	*b*	*t*
X: MOOD	6.813 (b_{41})	4.415**	4.336 (b_{51})	6.219**	1.480 (b_{61})	.957
MO: NFC	1.268 (b_{42})	1.117	.767 (b_{52})	1.496	.356 (b_{62})	.366
XMO: MOODNFC	−.691 (b_{43})	−.609	1.256 (b_{53})	2.450*	−2.169 (b_{63})	−2.112*
ME: POS					1.248 (b_{64})	6.613**
MEMO: POSNFC					−.036 (b_{65})	−.279

Note. ATT = attitude change; POS = positive valenced thoughts; MO = moderator variable; NFC = need for cognition; ME = mediator variable.
* $p < .05$. ** $p < .01$.

Integrating the Two Processes

Although our purpose in this article has been to clearly define both mediated moderation and moderated mediation, it should by now be apparent that ultimately they both rest on the same analytic models and the same fundamental equality, given in Equation 7:

$$\beta_{43} - \beta_{63} = \beta_{64}\beta_{53} + \beta_{65}\beta_{51} \tag{7}$$

We can conceptualize a continuum of analytic cases, varying in the relative magnitude of the difference on the left side of this equality (and thereby varying as well as a function of the terms on its right side). The continuum is defined by the relative magnitude of the difference between the absolute values of both β_{43} and β_{63}: $|\beta_{43}| - |\beta_{63}|$.[13] In the middle of this continuum lies the situation in which this difference equals zero. In this case, neither mediated moderation nor moderated mediation can be said to occur. Below this point lies the range of cases in which $|\beta_{43}| - |\beta_{63}| < 0$ and above it lies the range of cases in which $|\beta_{43}| - |\beta_{63}| > 0$.

Our consistent definition of mediated moderation rests on this difference being greater than zero: There is moderation of the overall treatment effect in Equation 4 (β_{43}), and this is reduced in magnitude once the moderation of the indirect effect is controlled (β_{63}).

One might then be tempted to say that the other process, moderated mediation, occurs whenever the difference is smaller than zero, that is, whenever it is the case that the moderation of the residual treatment effect (β_{63}) is greater than the moderation of the overall treatment effect (β_{43}). But we have been very careful throughout our discussion to always talk about moderated mediation in the "prototypic" case when there is no overall moderation, that is, when $\beta_{43} = 0$. Our reasons for this insistence on defining the "prototypic" case of moderated mediation can now be made clear. If one defines moderated mediation without the assumption that $\beta_{43} = 0$, then we would suggest that any point along the continuum we have just defined (other than zero), including points where $|\beta_{43}| - |\beta_{63}| > 0$, can be seen as moderated mediation, depending on the theoretical intentions of the researcher. That is, if moderated mediation consists solely in identifying whether a mediating process between the treatment and the outcome is moderated, then any time the terms on the right or left side of Equation 7 depart from zero, this can be said to occur.

This continuum of values for $|\beta_{43}| - |\beta_{63}|$ thus provides conceptual insights about the processes we are examining. When this difference equals zero, then neither mediated moderation nor moderated mediation can be said to occur. When mediated moderation is hypothesized this difference must be positive. On the other hand, when moderated mediation is hypothesized this difference will be negative if $\beta_{43} = 0$, which is the case for what we have called prototypic moderated mediation. If one defines moderated mediation as occurring whenever the mediating indirect effect is moderated, relaxing the restriction of the prototypic case that $\beta_{43} = 0$, then moderated mediation can be said to occur regardless of the sign of $|\beta_{43}| - |\beta_{63}|$, so long as it is not zero. Accordingly, under this more relaxed definition, every case of mediated moderation could be called moderated mediation, but the reverse is not true.

In light of this conclusion, it is important to emphasize that the relative magnitude of the $|\beta_{43}| - |\beta_{63}|$ difference does not by itself determine whether one is examining a case of mediated moderation or moderated mediation. Other issues are involved. For instance, as we have made clear, an overall treatment moderation is assumed in the case of mediated moderation (i.e., $\beta_{43} \neq 0$).

Ultimately, we would suggest that whether reference is made to mediated moderation or moderated mediation depends on the theoretical goals of the researcher, having examined the relevant models. If there is overall moderation of the treatment effect and if the models are examined in order to determine what is the process responsible for this overall moderation, then the analysis is in the service of mediated moderation goals. If, on the other hand, one suspects that the process mediating a treatment effect depends on a particular moderator, then the analysis is in the service of moderated mediation goals. And this may be the case regardless of whether there is or is not moderation of the overall treatment effect. In other words, while we have defined the prototypic moderated mediation case as occurring when the overall treatment effect is unmoderated, ultimately moderated mediation is the goal whenever the analysis is undertaken to understand how a mediating process is moderated, regardless of whether the overall treatment effect is itself moderated.

Remaining Issues and Conclusion

Overall Significance Testing

Our focus throughout has been on specifying the conditions necessary to establish mediated moderation and moderated mediation. To do this, we have tended to focus on the values of population parameters and only occasionally have we discussed their estimates, their standard errors, and traditional significance testing. Following the advice we have given on these matters, our approach to testing for mediated moderation and moderated mediation may seem rather piecemeal. For instance, in the case of both processes, we argue that at least one of the two indirect effects (from the treatment through the mediator to the outcome) should be significantly moderated, while the other indirect effect should be significant on average. Additionally, in demonstrating mediated moderation, we argued that overall moderation should be significant. In outlining what we have called prototypic moderated mediation, we suggested that the overall treatment effect (i.e., b_{41}) should be significant, overall moderation should not be found (i.e., b_{43}, *ns*), but the residual effect of the treatment on the outcome may be significant (i.e., b_{63}).

One might be tempted to argue that some overall test of both processes might be more appropriate. If so, then the obvious candidate here is to provide tests of whether the estimated terms on either side of the equality of Equation 7 differ significantly from zero. Thus one might be tempted to develop tests of either of the following null hypotheses and call these overall tests of mediated moderation or moderated mediation:

$$Ho\colon \beta_{43}\beta_{63} = 0.$$

$$Ho\colon \beta_{64}\beta_{53} + \beta_{65}\beta_{51} = 0.$$

This would provide overall tests analogous to those provided by MacKinnon et al. (2002) in the case of simple mediation. It would

[13] We are assuming here that both of these parameters have the same sign.

certainly seem that some of the tests examined by MacKinnon et al. could be extended to testing at least the first of the above null hypotheses.

Although such a test might be a desirable addition to our piecemeal approach, we think that such an overall test would not be sufficient in and of itself to demonstrate either mediated moderation or moderated mediation. Just as we believe that establishing simple mediation requires that the researcher examine a series of models and plausible inferences, so too uncovering the more complex processes we are dealing with involves a bit of detective work, examining whether the overall pattern of results, including confidence intervals for a number of different parameters in the models, support the notions of mediated moderation and moderated mediation. We also believe that it is of theoretical interest for researchers to differentiate the alternative ways in which mediated moderation and moderated mediation may be produced, and this would not occur if a single overall test was the only thing examined.

Tests of Simple Mediation

In the section where we presented analyses of hypothetical data, we calculated simple effects to help interpretation. One may want to actually test these simple effects and, by doing so, to test simple mediation at various levels of a moderator. In order to do so, one would need to deviate the moderator not from its mean (as in centering), but from whatever values are of interest. So for instance, with a dichotomous moderator, one may want to test simple mediation at each level of the moderator. Here, instead of using a contrast-coded moderator, one would use codes that give a value of zero to first one, and then the other, of the two levels of the moderator (see Judd & McClelland, 1989, for tests of simple effects in the context of multiple regressions). This would be equivalent to the use of dummy coding conventions. Consider first Equations 5 and 6 when we use Mo^a to code the moderator, with values of 0 for level A and + 1 for level B. (The superscript indicates the level of the moderator receiving the 0 value). The resulting equations are:

$$Me = \beta^a_{50} + \beta^a_{51}X + \beta^a_{52}Mo^a + \beta^a_{53}XMo^a + \varepsilon_5 \qquad (5a)$$

$$Y = \beta^a_{60} + \beta^a_{61}X + \beta^a_{62}Mo^a + \beta^a_{63}XMo^a + \beta^a_{64}Me + \beta^a_{65}MeMo^a + \varepsilon_6 \qquad (6a)$$

In these equations, β^a_{51} now equals the simple effect of the treatment on the mediator for individuals in level A of the Moderator. Likewise, β^a_{64} equals the simple effect of the mediator on the outcome for level A. And β^a_{61} equals the simple residual effect of the treatment on the outcome, over and above mediation, again for level A of the moderator. The significance tests associated with the estimates of these coefficients can be used to test whether those simple effects differ reliably from zero.

Alternatively we can redefine the moderator as Mo^b (i.e., Level B = 0; Level A = 1):

$$Me = \beta^b_{50} + \beta^b_{51}X + \beta^b_{52}Mo^b + \beta^b_{53}XMo^b + \varepsilon_5 \qquad (5b)$$

$$Y = \beta^b_{60} + \beta^b_{61}X + \beta^b_{62}Mo^b + \beta^b_{63}XMo^b + \beta^b_{64}Me + \beta^b_{65}MeMo^b + \varepsilon_6 \qquad (6b)$$

And now β^b_{51}, β^b_{64}, and β^b_{61} represent the simple effects for level B of the moderator and one can test whether their estimates differ from zero reliably. In the more general case of a continuous moderator, one would use the same strategy but would deviate the moderator from values of interest, for instance one standard deviation above and below its mean.

Although these alternative specifications of the moderator allow one to estimate and test the simple mediation effects at different levels of the moderator, it has to be remembered that to demonstrate moderated mediation it is necessary to show that these simple mediation effects significantly depend on the moderator. In other words, it is not sufficient simply to show significant simple mediation at one level of the moderator but not at another. One must show that the process of mediation is significantly different at different levels of the moderator and this is done by testing the various interaction terms in Equations 5 and 6, as previously described.

Extensions to Purely Correlational Data

Throughout our presentation we have made the strong assumption that the treatment is an experimentally manipulated one, permitting causal inference about the overall treatment effect and its effect on the mediator. This has also permitted us to assume that the treatment is independent of the moderator, an important condition necessary for the derivation we gave in the Appendix for the equality underlying both mediated moderation and moderated mediation (Equation 7).

In the literature on mediation, we have seen (all too often for our tastes) researchers take three measured variables, claim a mediational model, and believe, because a "test of mediation" (e.g., one of those recommended by MacKinnon et al., 2002) comes out, that they have established mediation. But in this case, there is no way to decide which variable is the treatment whose effect is mediated, which is the mediator, and which is the outcome. It is not our goal to encourage these kinds of practices, even though we recognize that some will use the models that we develop in this article to claim either mediated moderation or moderated mediation in the presence of a purely correlational research design. *If* the researchers' causal model is the correct one, then it stands to reason that the models we have outlined in this article can in fact be used to examine issues of mediated moderation and moderated mediation with purely correlational data. Although the fundamental relationship of Equation 7 would not be expected to hold, the resulting parameter estimates would be interpretable in a manner consistent with what we have presented.

Practical Issues

A potentially important issue that we have not previously discussed concerns the biasing effects of measurement error in the moderator, mediator, and outcome variables. Unfortunately the direction of bias due to measurement error is difficult to know, given the complexities of the models we are dealing with. The obvious solution is to use multiple indicators of these variables, weighting them appropriately, as in a structural equation latent variable approach (SEM). With a dichotomous moderator, this approach can be implemented by a multiple Group SEM estimation (but this assumes that the moderator is measured perfectly).

With a continuously measured moderator, there exists no straightforward solution for estimating SEM models with latent variable interactions (see Jaccard & Wan, 1995; Kenny & Judd, 1984; Li et al., 1998).

One obvious alternative solution is to measure all three variables (mediator, moderator, and outcome) with multiple indicators which are then combined or aggregated into index scores, assuming they manifest high internal consistency. Ultimately, for the sorts of interactive models that underlie the analyses we have outlined, we suspect that this is the most efficient way to deal with the measurement error issue. Of course, this approach deals only with random, rather than systematic, measurement error.

Another practical issue to discuss in estimating these processes is that of sufficient statistical power. The literature on simple mediation suggests that some tests of mediation may be relatively unpowerful (MacKinnon et al., 2002). This may also be true in the cases we have been considering. The obvious advice is that the researcher who is interested in examining moderated mediation and mediated moderation should do whatever he or she can to maximize statistical power.

We have confined our discussion to cases with a single mediator and a single moderator. In the case of moderated mediation, it seems likely that there may exist multiple indirect paths from the treatment to the outcome, with one mediated pathway being more potent for some individuals and another being more important for others. In this case, with potentially collinear mediators (and moderators), power issues obviously would become particularly important.

Conclusion

When we first set out to write this article, our goals were to clearly define both mediated moderation and moderated mediation and to elaborate in greater detail than had been done previously the analytic models underlying both. Earlier discussions of these processes (Baron & Kenny, 1986; James & Brett, 1984; Wegener & Fabrigar, 2000) are certainly informative, but it seemed to us that both processes deserved a more detailed treatment. Thus, we envisioned ourselves writing a largely didactic piece, laying out the analytic roadmaps for these alternative analyses.

Although we hope that we have accomplished this, it became apparent to us, as our ideas (and this article) evolved, that we had much more to say at a theoretical level about the inherent similarities between the two processes. They both rely on the same analytic models, they both imply moderated indirect effects of the treatment variable on the outcome, and they both imply that the overall moderation of the treatment effect is altered once the (moderated) mediating process is controlled. At the same time, we would not want to claim that moderated mediation and mediated moderation are one and the same. By defining the prototypic moderated mediation case, we can clearly differentiate between them. Additionally, mediated moderation clearly does imply that the overall moderation of a treatment effect is reduced once the (moderated) mediating process is controlled.

That said, however, we also have argued that if one allows for the possibility of moderated mediation in cases where there is overall moderation of the treatment effect, then the distinction between the two processes can become more a matter of theoretical preference than anything else. Is the emphasis on the reduction of the moderation of the treatment effect once the (moderated) mediating process is controlled? Or is the emphasis more on the fact that there is a moderated mediating process?

The analytic models are clear cut and the two processes can be defined in the prototypic cases. But these, as we have already said, represent probably two sides of the same coin. In talking about that coin, we can either concentrate on describing each side in turn, or we can recognize that they both define the common coin.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Thousand Oaks, 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.

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin, 117,* 348–357.

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 evaluation. *Evaluation Review, 5,* 602–619.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model comparison approach.* San Diego, CA: Harcourt, Brace, Jovanovich.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96,* 201–210.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 233–265). Boston, MA: McGraw-Hill.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59,* 877–884.

Li, F., Harmer, P., Duncan, T. E., Duncan, S. C., Acock, A., & Boles, S. (1998). Approaches to testing interaction effects using structural equation modeling methodology. *Multivariate Behavioral Research, 33,* 1–39.

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., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30,* 41–62.

Petty, R. E., Schumann, D. W., Richman, S. A., & Strathman, A. J. (1993). Positive mood and persuasion: Different roles for affect under high- and low-elaboration conditions. *Journal of Personality and Social Psychology, 64,* 5–20.

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

Smeesters, D., Warlop, L., Van Avermaet, E., Corneille, O., & Yzerbyt, V. Y. (2003). Do not prime hawks with doves: The interplay of construct activation and consistency of social value orientation on cooperative behavior. *Journal of Personality and Social Psychology, 84,* 972–987.

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

Appendix

In the case of both mediated moderation and moderated mediation, the analysis proceeds by estimating three models:

$$Y = \beta_{40} + \beta_{41}X + \beta_{42}Mo + \beta_{43}XMo + \varepsilon_4 \tag{4}$$

$$Me = \beta_{50} + \beta_{51}X + \beta_{52}Mo + \beta_{53}XMo + \varepsilon_5 \tag{5}$$

$$Y = \beta_{60} + \beta_{61}X + \beta_{62}Mo + \beta_{63}XMo + \beta_{64}Me + \beta_{65}MeMo + \varepsilon_6 \tag{6}$$

Let us assume that Equations 5 and 6 represent the theoretical models that are responsible for generating the variance in both the mediator and the outcome. Accordingly, the model of Equation 4 is a misspecified model. It then becomes possible to derive the values of the parameters in this misspecified model in terms of the parameters of Equations 5 and 6.

First, we combine Equations 5 and 6 by substituting for *Me* in Equation 6 according to Equation 5. This leads to the following result:

$$\begin{aligned} Y = {} & \beta_{60} + \beta_{61}X + \beta_{62}Mo + \beta_{63}XMo \\ & + \beta_{64}(\beta_{50} + \beta_{51}X + \beta_{52}Mo + \beta_{53}XMo + \varepsilon_5) \\ & + \beta_{65}(\beta_{50} + \beta_{51}X + \beta_{52}Mo + \beta_{53}XMo + \varepsilon_5)Mo + \varepsilon_6 \end{aligned} \tag{6$'$}$$

which is given equivalently as:

$$\begin{aligned} Y = {} & (\beta_{60} + \beta_{64}\beta_{50}) + (\beta_{61} + \beta_{64}\beta_{51})X \\ & + (\beta_{62} + \beta_{64}\beta_{52} + \beta_{65}\beta_{50} + \beta_{65}\varepsilon_5)Mo + (\beta_{63} + \beta_{64}\beta_{53} + \beta_{65}\beta_{51})XMo \\ & + \beta_{65}\beta_{52}Mo^2 + \beta_{65}\beta_{53}XMo^2 + \beta_{65}\varepsilon_5 + \varepsilon_6 \end{aligned} \tag{6$'$}$$

Assuming that X is contrast coded, with an expected value of zero, and that *Mo* is normally distributed with a mean of zero and is independent of X, then it can be shown (Aiken & West, 1991; pp. 177–182; Kenny & Judd, 1984) that the expected covariances of *XMo* with both Mo^2 and XMo^2 equal zero. Accordingly, the parameter associated with *XMo* in the respecified Equation 6′ ($\beta_{63} + \beta_{64}\beta_{53} + \beta_{65}\beta_{51}$) must equal its parameter in the misspecified Equation 4. Accordingly, in the population:

$$\beta_{43} = \beta_{63} + \beta_{64}\beta_{53} + \beta_{65}\beta_{51}$$

equivalently:

$$\beta_{43} - \beta_{63} = \beta_{64}\beta_{53} + \beta_{65}\beta_{51}$$

Of course, in any sample of data, the estimated covariances of *XMo* with both Mo^2 and XMo^2 will not exactly equal zero and, accordingly, the above equality will only be approximate.

When *Mo* is a contrast coded dichotomous variable, then Mo^2 is constant for all cases (i.e., its variance equals zero). In this case, Equation 6′ reduces to:

$$\begin{aligned} Y = {} & (\beta_{60} + \beta_{64}\beta_{50} + \beta_{65}\beta_{52}) + (\beta_{61} + \beta_{64}\beta_{51} + \beta_{65}\beta_{53})X \\ & + (\beta_{62} + \beta_{64}\beta_{52} + \beta_{65}\beta_{50} + \beta_{65}\varepsilon_5)Mo \\ & + (\beta_{63} + \beta_{64}\beta_{53} + \beta_{65}\beta_{51})XMo + \beta_{64}\varepsilon_5 + \varepsilon_6 \end{aligned} \tag{6$'$}$$

Accordingly, in this case, the equality holds both in terms of expected values and in terms of estimated coefficients. In other words, with a contrast-coded moderator, the following is true of sample estimates:

$$b_{43} - b_{63} = b_{64}b_{53} + b_{65}b_{51}$$

Received October 11, 2004
Revision received June 20, 2005
Accepted June 20, 2005 ■