

Online supplement to Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67, 451-470. DOI: 10.1111/bmsp.12028

[This document contains corrections to a few typos that were found on the version available through the journal's web page and adds a note about new features in PROCESS that eliminate the need to run PROCESS twice]

This document contains instructions for the implementation of the method described in Hayes and Preacher (2014) using Mplus as well as using the PROCESS and MEDIANTE macros for SPSS and SAS. Following the code, various miscellaneous issues and extensions are addressed, including interpretation of model coefficients using sequential group coding, accounting for random measurement error, dealing with confounds statistically, and models with multiple mediators.

Mplus Code Corresponding to the Web Portal Customization Example

Any structural equation modeling program can produce estimates of the coefficients in a mediation model. Mplus offers features such as bootstrap confidence intervals for indirect effects and inferential tests for functions of parameters that make it a particularly good choice for the kind of analysis we describe in the manuscript. Importantly, the constraints of the freely available demonstration version of Mplus (available from <http://www.statmodel.com/>) do not preclude its use for estimation of mediation models with a single mediator and a categorical independent variable with as many as three levels. The code below implements the method described in the manuscript and can easily be adapted to mediation analysis with multiple mediators, latent variables, or an independent variable with more than three levels.

```
DATA:
  FILE IS c:\sri.txt;
VARIABLE:
  NAMES ARE cond custom attitude inter;
  USEVARIABLES ARE attitude inter d1 d2;

!indicator coding
DEFINE:
  IF (cond EQ 1) THEN d1 = 0;
  IF (cond EQ 1) THEN d2 = 0;
  IF (cond EQ 2) THEN d1 = 1;
  IF (cond EQ 2) THEN d2 = 0;
  IF (cond EQ 3) THEN d1 = 0;
  IF (cond EQ 3) THEN d2 = 1;

!model definition
MODEL:
  inter ON d1 (a1)
    d2 (a2);
  attitude ON inter (b)
    d1 (cp1)
    d2 (cp2);

!relative indirect effects;
MODEL INDIRECT:
  attitude IND inter d1;
  attitude IND inter d2;
MODEL CONSTRAINT:
  NEW (tot1 tot2);
  tot1=a1*b+cp1;
  tot2=a2*b+cp2;
```

The resulting output is below. This output was used to construct parts of Table 2 in the manuscript.

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
INTER ON				
D1	1.575	0.487	3.233	0.001
D2	2.250	0.487	4.619	0.000
ATTITUDE ON				
INTER	0.359	0.091	3.965	0.000
D1	1.105	0.370	2.985	0.003
D2	2.158	0.398	5.426	0.000
Intercepts				
ATTITUDE	2.810	0.454	6.187	0.000
INTER	4.250	0.344	12.338	0.000
Residual Variances				
ATTITUDE	1.166	0.213	5.477	0.000
INTER	2.373	0.433	5.477	0.000
New/Additional Parameters				
TOT1	1.670	0.384	4.353	0.000
TOT2	2.965	0.384	7.728	0.000

TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Effects from D1 to ATTITUDE				
Sum of indirect	0.565	0.226	2.506	0.012
Specific indirect				
ATTITUDE				
INTER				
D1	0.565	0.226	2.506	0.012
Effects from D2 to ATTITUDE				
Sum of indirect	0.807	0.268	3.008	0.003
Specific indirect				
ATTITUDE				
INTER				
D2	0.807	0.268	3.008	0.003

For contrast coding as described in the text, replace the DEFINE section above with

```
!orthogonal contrast coding
DEFINE:
  if (cond eq 1) then d1 = -0.667;
  if (cond eq 1) then d2 = 0;
  if (cond eq 2) then d1 = 0.333;
  if (cond eq 2) then d2 = -0.5;
  if (cond eq 3) then d1 = 0.333;
  if (cond eq 3) then d2 = 0.5;
```

For sequential coding as discussed later in this supplement, replace the DEFINE section of the core program with

```
!sequential coding
DEFINE:
  if (cond eq 1) then d1 = 0;
  if (cond eq 1) then d2 = 0;
  if (cond eq 2) then d1 = 1;
  if (cond eq 2) then d2 = 0;
  if (cond eq 3) then d1 = 1;
  if (cond eq 3) then d2 = 1;
```

To generate 95% and 99% bias corrected bootstrap confidence intervals for relative indirect effects (as well as all other parameter estimates), add the lines below to the program. For percentile confidence intervals, change “bcbootstrap” below to “bootstrap”.

```
ANALYSIS:
bootstrap = 10000;

OUTPUT:
cinterval (bcbootstrap);
```

Estimation using PROCESS for SPSS and SAS

NOTE: The text in this section is what was provided to the journal when the article was published. Since this paper was published, a feature was added to PROCESS that allows for the specification of X as a multicategorical variable in model 4. This eliminates the need to run PROCESS twice using the procedure described below. For instructions, see the addendum to the documentation for PROCESS. PROCESS can be downloaded from www.processmacro.org

PROCESS is a freely-available regression-based path analysis macro for both SPSS and SAS that estimates the model coefficients in mediation and moderation models of various forms while also providing modern inferential methods for inference about indirect effects including bootstrap confidence intervals. Its use in mediation analysis is described in Hayes (2013) along with documentation of its many features, and can be downloaded from [web address withheld for peer review]

One documented limitation of PROCESS is that only a single *X* variable can be specified in a mediation model, and it must be either dichotomous or continuous. However, with the strategic use of covariates, manual construction of the indicator codes prior to execution, and multiple executions of the macro, PROCESS can estimate a model as in Figure 2 of the manuscript. The results generated by PROCESS will be identical to what Mplus generates, with the exception of standard errors which will tend to be slightly smaller than OLS standard errors in smaller samples. These differences in standard errors dissipate rapidly as sample size increases.

The example SPSS PROCESS code and output below corresponds to the analysis of the web portal customization study using indicator coding of customization condition. Variables named ATTITUDE and

INTER contain measurements of attitudes toward the web portal and perceived interactivity, respectively, and variable COND codes experimental condition (1 = control, 2 = moderate customization, 3 = high customization).

Because PROCESS allows only a single independent variable that must be either dichotomous or continuous, it must be tricked into estimating a model with a multicategorical independent variable. This is done by running PROCESS $k - 1$ times, where k is the number of levels of the independent variable, and using $k - 1$ group codes constructed prior the execution of PROCESS. At each run, one of the group codes is used as X and the others as covariate(s), with the code serving as X being swapped with a covariate at subsequent PROCESS runs. So that the same bootstrap samples are used in consecutive executions, the random number generator should be seeded using the **seed** = command, with the same seed used time. This seed can be chosen arbitrarily.

This code first constructs two dummy variables coding experimental condition with the control condition (cond = 1) as the reference category. The following PROCESS command then executes a mediation model with the first dummy variable as X and the other as a covariate. This generates estimates of a_1 , a_2 , b , c_1 , c_2 , c'_1 , and c'_2 corresponding to the values in Table 2 of the manuscript, as well as a bias-corrected bootstrap confidence interval for a_1b based on 10,000 bootstrap samples. The summary table at the end includes the three effects of X , which in this case are the relative total, direct, and indirect effects for moderate customization relative to the control condition (c_1 , and c'_1 , and a_1b), in that order.

```
compute d1=(cond=2).
compute d2=(cond=3).
process vars=attitude inter d1 d2/y=attitude/m=inter/x=d1/total=1/
model=4/boot=10000/seed=3423.
```

```
Model = 4
Y = attitude
X = d1
M = inter
```

```
Statistical Controls:
CONTROL= d2
```

```
Sample size
60
```

```
*****
Outcome: inter
```

Model Summary

R	R-sq	F	df1	df2	p
.5220	.2725	10.6734	2.0000	57.0000	.0001

Model

	coeff	se	t	p	LLCI	ULCI
constant	4.2500	.3534	12.0256	.0000	3.5423	4.9577
d1	1.5750	.4998	3.1512	.0026	.5742	2.5758
d2	2.2500	.4998	4.5018	.0000	1.2492	3.2508

```
*****
Outcome: attitude
```

Model Summary

R	R-sq	F	df1	df2	p
.7771	.6039	28.4646	3.0000	56.0000	.0000

Model						
	coeff	se	t	p	LLCI	ULCI
constant	2.8100	.4701	5.9772	.0000	1.8682	3.7517
inter	.3588	.0937	3.8302	.0003	.1712	.5465
d1	1.1048	.3831	2.8842	.0056	.3375	1.8722
d2	2.1576	.4116	5.2423	.0000	1.3331	2.9821

***** TOTAL EFFECT MODEL *****

Outcome: attitude

Model Summary

R	R-sq	F	df1	df2	p
.7072	.5002	28.5213	2.0000	57.0000	.0000

Model						
	coeff	se	t	p	LLCI	ULCI
constant	4.3350	.2783	15.5749	.0000	3.7776	4.8924
d1	1.6700	.3936	4.2426	.0001	.8818	2.4582
d2	2.9650	.3936	7.5326	.0000	2.1768	3.7532

***** TOTAL, DIRECT, AND INDIRECT EFFECTS *****

Total effect of X on Y

Effect	SE	t	p	LLCI	ULCI
1.6700	.3936	4.2426	.0001	.8818	2.4582

Direct effect of X on Y

Effect	SE	t	p	LLCI	ULCI
1.1048	.3831	2.8842	.0056	.3375	1.8722

Indirect effect of X on Y

	Effect	Boot SE	BootLLCI	BootULCI
inter	.5652	.2724	.1643	1.2665

***** ANALYSIS NOTES AND WARNINGS *****

Number of bootstrap samples for bias corrected bootstrap confidence intervals:
10000

Level of confidence for all confidence intervals in output:
95.00

Missing from the output above is the relative indirect effect for high customization relative to none (a_2b) along with a bootstrap confidence interval for inference. The code below generates this relative indirect effect by switching d1 and d2 in the **x=** specification. Most of the output is identical to the code generated by the command above, so that output is suppressed by using the **detail=0** option. Using the same random number seed as in the prior run of PROCESS produces a bootstrap confidence interval based on the same set of bootstrap samples. The effects for X in this summary table are the relative total, direct, and indirect effects for high customization relative to the control condition (c_2 , c'_2 , and a_2b), in that order.

```
process vars=attitude inter d1 d2/y=attitude/m=inter/x=d2/total=1/
model=4/boot=10000/seed=3423/detail=0.
```

```
Model = 4
Y = attitude
X = d2
M = inter
```

Statistical Controls:
CONTROL= d1

Sample size
60

***** TOTAL, DIRECT, AND INDIRECT EFFECTS *****

Total effect of X on Y

Effect	SE	t	p	LLCI	ULCI
2.9650	.3936	7.5326	.0000	2.1768	3.7532

Direct effect of X on Y

Effect	SE	t	p	LLCI	ULCI
2.1576	.4116	5.2423	.0000	1.3331	2.9821

Indirect effect of X on Y

	Effect	Boot SE	BootLLCI	BootULCI
inter	.8074	.3273	.3217	1.6442

***** ANALYSIS NOTES AND WARNINGS *****

Number of bootstrap samples for bias corrected bootstrap confidence intervals:
10000

Level of confidence for all confidence intervals in output:
95.00

The SPSS compute commands above generate indicator codes with the control group as the reference group. The commands to generate the contrast codes used in the example analysis would be

```
if (cond=1) d1 = -0.667.  
if (cond=1) d2 = 0.  
if (cond=2) d1 = 0.333.  
if (cond=2) d2 = -0.5.  
if (cond=3) d1 = 0.333.  
if (cond=3) d2 = 0.5.
```

For the sequential coding example described below, the following SPSS commands construct the sequential codes:

```
compute d1 = (cond > 1).  
compute d2 = (cond > 2).
```

The PROCESS macro is available for SAS but requires PROC IML. The command structure is very similar to the SPSS version, but the construction of group codes requires commands that are different than those used in SPSS. The SAS code below conducts the example analysis using indicator coding of groups, assuming the data reside in a SAS data file named "web":

```
data web;set web;d1=(cond=2);d2=(cond=3);run;  
%process (data=web,vars=attitude inter d1 d2,y=attitude,m=inter,x=d1,  
total=1,model=4,boot=10000,seed=3423);  
%process (data=web,vars=attitude inter d1 d2,y=attitude,m=inter,x=d2,  
total=1,model=4,boot=10000,seed=3423,detail=0);
```

For the contrast codes corresponding to the example analysis in this paper, change the DATA line to read:

```
data web;set web;
if (cond=1) then do;d1=-0.667;d2=0;end;
if (cond=2) then do;d1=0.333;d2=-0.5;end;
if (cond=3) then do;d1=0.333;d2=0.5;end;
run;
```

For the sequential codes described in the example below, the DATA line should read

```
data web;set web;d1=(cond>1);d2=(cond>2);run;
```

*Estimation using **MEDIATE** for SPSS*

NOTE: The text in this section is what was provided to the journal when the article was published. Since this paper was published, a feature was added to PROCESS that allows for the specification of X as a multicategorical variable in model 4. The resulting PROCESS output looks very similar to what **MEDIATE produces.**

MEDIATE is a freely available SPSS macro (downloadable from [web address blinded for review]) that facilitates the estimation of mediation models with multicategorical independent variables along with the ability to generate bootstrap confidence intervals for indirect effects. It is very limited in its features relative to PROCESS, but it does have one handy option that automates the construction of codes for a categorical independent variable. The code and output below corresponds to the analysis of the web portal customization study using indicator coding of customization condition. Variables named ATTITUDE and INTER contain measurements of attitudes toward the web portal and perceived interactivity, respectively, and variable COND codes experimental condition (1 = control, 2 = moderate customization, 3 = high customization). The **catx=1** option specifies indicator coding and sets the control condition as the reference group. See the documentation for additional information about the MEDIATE macro and its options.

```
mediate y=attitude/x=cond/m=inter/samples=10000/total=1/catx=1.
```

Run MATRIX procedure:

VARIABLES IN THE FULL MODEL:

```
Y = attitude
M1 = inter
X = cond
```

CODING OF CATEGORICAL X FOR ANALYSIS:

cond	D1	D2
1.0000	.0000	.0000
2.0000	1.0000	.0000
3.0000	.0000	1.0000

OUTCOME VARIABLE:

attitude

MODEL SUMMARY (TOTAL EFFECTS MODEL)

R	R-sq	Adj R-sq	F	df1	df2	p
.7072	.5002	.4826	28.5213	2.0000	57.0000	.0000

MODEL COEFFICIENTS (TOTAL EFFECTS MODEL)

	Coeff.	s.e.	t	p
Constant	4.3350	.2783	15.5749	.0000
D1	1.6700	.3936	4.2426	.0001
D2	2.9650	.3936	7.5326	.0000

OUTCOME VARIABLE:

inter

MODEL SUMMARY

R	R-sq	Adj R-sq	F	df1	df2	p
.5220	.2725	.2469	10.6734	2.0000	57.0000	.0001

MODEL COEFFICIENTS

	Coeff.	s.e.	t	p
Constant	4.2500	.3534	12.0256	.0000
D1	1.5750	.4998	3.1512	.0026
D2	2.2500	.4998	4.5018	.0000

OUTCOME VARIABLE:

attitude

MODEL SUMMARY

R	R-sq	adj R-sq	F	df1	df2	p
.7771	.6039	.5827	28.4646	3.0000	56.0000	.0000

MODEL COEFFICIENTS

	Coeff.	s.e.	t	p
Constant	2.8100	.4701	5.9772	.0000
inter	.3588	.0937	3.8302	.0003
D1	1.1048	.3831	2.8842	.0056
D2	2.1576	.4116	5.2423	.0000

TEST OF HOMOGENEITY OF REGRESSION (X*M INTERACTION)

	R-sq	F	df1	df2	p
inter	.0043	.2969	2.0000	54.0000	.7443

INDIRECT EFFECT(S) THROUGH:

inter

	Effect	SE(boot)	LLCI	ULCI
D1	.5652	.2694	.1693	1.2548
D2	.8074	.3252	.3338	1.6587

***** ANALYSIS NOTES AND WARNINGS *****

NOTE: Indicator coding is used for categorical X

Number of samples used for indirect effect confidence intervals: 10000

Level of confidence for confidence intervals: 95.0000

Bias corrected bootstrap confidence intervals for indirect effects are printed in output

Sequential Coding of Groups

In the web portal customization study, the three levels of the manipulation can be rank ordered with respect to degree of customization (none, moderate, or high). When the categories of a multicategorical predictor can be so ordered, sequential coding can be useful. With sequential codes, the relative direct and indirect effects can be interpreted as the effects of membership in one group relative to the group one step sequentially lower in the ordered system. Darlington (1990, pp. 236-237) describes sequential coding for a categorical variable with any number of ordered categories. With only three groups, the coding is relatively simple. For the control condition (the lowest level of customization), D_1 and D_2 are set to 0, for the moderately customized condition (the next highest level of customization), $D_1 = 1$, $D_2 = 0$, and for the highest level of customization, $D_1 = D_2 = 1$.

Estimating the coefficients in Equations 6, 7, and 8 in the manuscript yields the following results: $i_1 = 4.250$, $i_2 = 2.810$, $i_3 = 4.335$, $a_1 = 1.575$, $p = 0.001$; $a_2 = 0.675$, $p = 0.166$; $b = 0.359$, $p < 0.001$; $c'_1 = 1.105$, $p = 0.003$; $c'_2 = 1.053$, $p = .002$; $c_1 = 1.670$, $p < .001$; $c_2 = 1.295$, $p < 0.002$. As with the other two methods of coding groups described in the manuscript, the resulting models reproduce the group means on M as well as Y (adjusted and unadjusted).

The relative indirect effects are still estimated as products of coefficients. The a_1 coefficients quantify the mean differences in perceived interactivity between the moderate customization and control condition (a_1) and between the high and moderate customization conditions (a_2). That is,

$$a_1 = \bar{M}_{\text{moderate}} - \bar{M}_{\text{control}} = 5.825 - 4.250 = 1.575$$

and

$$a_2 = \bar{M}_{\text{high}} - \bar{M}_{\text{moderate}} = 6.500 - 5.825 = 0.675.$$

When a_1 and a_2 are multiplied by the effect of interactivity on attitudes, holding customization condition constant ($b = 0.359$), the result is the relative indirect effects of customization on attitudes through perceived interactivity:

$$a_1b = 1.575(0.359) = 0.565$$

and

$$a_2b = 0.675(0.359) = 0.242.$$

The relative indirect effect a_1b estimates the indirect effect of moderate customization relative to none through perceived interactivity on attitudes. Those who browsed using a moderately customized portal had attitudes that were 0.565 units more favorable on average (with a 95% bias-corrected bootstrap confidence interval from 0.169 to 1.255) than those assigned to the noncustomized portal condition as a result of this indirect mechanism linking customization to attitudes through perceived interactivity. The relative indirect effect a_2b estimates the indirect effect of high relative to moderate customization through perceived interactivity. Browsing with a highly customized portal resulted in attitudes that were 0.242 units more favorable on average than browsing using a moderately customized portal as a result of this indirect mechanism linking customization to attitudes through perceived interactivity. Zero cannot be rejected as a plausible value for this indirect effect, as a 95% bias-corrected bootstrap confidence interval straddled zero (-0.043 to 0.710).

Using this coding system, the relative direct effect c'_1 corresponds to the effect of moderate customization on attitudes relative to none, independent of perceived interactivity, and the relative direct effect

c'_2 is the effect of high customization relative to moderate customization. This corresponds to differences between the adjusted means:

$$\begin{aligned} c'_1 &= \bar{Y}_{moderate}^* - \bar{Y}_{control}^* = 5.897 - 4.792 = 1.105 \\ c'_2 &= \bar{Y}_{high}^* - \bar{Y}_{moderate}^* = 6.950 - 5.897 = 1.053. \end{aligned}$$

As when other coding systems are used, the relative total effects can be estimated using Equation 8 in the manuscript or by adding the relative direct and indirect effects. With sequential coding, c_1 estimates the mean difference in attitude between the moderately customized and control groups, and c_2 estimates the mean difference in attitude between the highly customized and moderately customized groups. That is,

$$\begin{aligned} c_1 &= \bar{Y}_{moderate} - \bar{Y}_{control} = 6.005 - 4.335 = 1.670 \\ c_2 &= \bar{Y}_{high} - \bar{Y}_{moderate} = 7.300 - 6.005 = 1.295. \end{aligned}$$

As both effects are positive, this suggests attitudes increase in favorability as customization increases. Finally, notice that as with indicator or contrast coding, the relative total effects partition cleanly into the relative direct and relative indirect effects: $c_1 = c'_1 + a_1b = 1.105 + 0.565 = 1.670$ and $c_2 = c'_2 + a_2b = 1.053 + 0.242 = 1.295$.

Random Measurement Error

The example analyses in the manuscript and this supplement ignore the potential influence of random measurement error in X , M , or Y . In experiments, and even when X is an observed categorical variable, measurement error in X is often negligible to nonexistent unless the categories were constructed through some kind of artificial categorization of a continuum or there is some ambiguity or subjectivity in the decision as to which category a particular case in the data belongs. But M and/or Y may and often do contain some random measurement error, such as when they are sum scores from a psychological test, personality inventory, or attitude scale. If M , Y , or both is measured with error, the result is bias in the estimation of the effects of X , reduced statistical power, or both (see, e.g., Darlington, 1990, pp. 201-204; Ledgerwood & Shrout, 2012).

The method described in the manuscript can easily be extended using Mplus or another SEM program using single indicator latent variables with reliability-weighted errors (see e.g., Kline, 2005) or latent variable model with a measurement model component that links the latent variable causally to its indicators. Both approaches potentially reduce at least some of the deleterious effects of random measurement error. As with any measurement model, the researcher should ascertain whether the measurement model for the latent variable(s) satisfies various criteria for claiming “good fit,” for direct and indirect effects linking latent variables that are not modeled well have little substantive meaning. For discussions of latent variable mediation analysis, see Cheung and Lau (2008), Coffman and MacCallum (2005), Lau and Cheung (2012), and MacKinnon (2008).

Multiple Mediators

The approach we have illustrated for estimating relative indirect and direct effects can be extended to models with any number (m) of mediators operating in parallel. Figure S1 depicts a model with m proposed mediators and a multicategorical X with k categories. The relative total effects, c_i , can be estimated if desired using Equation 8 in the manuscript, whereas the relative indirect and direct effects are pieced together from parameter estimates from $m + 1$ linear models, one for each of the m mediators and one for Y :

$$\begin{aligned} M_j &= i_{1j} + a_{1j}D_1 + a_{2j}D_2 + \dots + a_{(k-1)j}D_{k-1} + e_{Mj} & (S1) \\ Y &= i_2 + c'_1D_1 + c'_2D_2 + \dots + c'_{k-1}D_{k-1} + b_1M_1 + b_2M_2 + \dots + b_mM_m + e_Y & (S2) \end{aligned}$$

The same relationships among relative total, indirect, and direct effects exist in multiple-mediator models as in single-mediator models. The relative total effect for D_i can be partitioned into the relative direct effect for D_i plus the sum of the *relative specific indirect effects* for D_i , $a_{i1}b_1 + a_{i2}b_2 + \dots + a_{im}b_m$. That is,

$$c_i = c'_i + \sum_{j=1}^m a_{ij}b_j \quad (S3)$$

This last term in Equation S3 is the *relative total indirect effect* of D_i . Each relative specific indirect effect quantifies the component of the relative total indirect effect that is carried uniquely through that mediator. Inferential tests of relative specific indirect effects can be undertaken just as described in the manuscript, and these would typically be the focus of a mediation analysis. The Mplus code above can be modified without difficulty to include multiple mediators, and the PROCESS and MEDiate procedures for SPSS and SAS allow for multiple mediators operating in parallel in this fashion. See the documentation.

Covariates and Confounds

In a mediation model, the interpretation of an indirect effect as a causal one assumes that the mediator M is causally located between X and Y . That is, it is assumed that X causes M and M causes Y . When X is experimentally manipulated and sound experimental procedures are followed, a causal association between X and M and between X and Y is established by showing that the k groups differ on M and Y on average. Of course, as many others have emphasized before us (e.g., Bullock et al., 2010; Hayes, 2013; Mathieu, DeShon, & Bergh, 2008; Stone-Romero & Rosopa, 2010), this does not establish that M causes Y . It could be that Y causes M or that M and Y are spuriously associated (both are caused by some variable W) or epiphenomenally associated (M is correlated with the “true” intermediary variable W). If X is not experimentally manipulated, such threats to causal inference also exist in the interpretation of the association between X and M as well.

Spuriousness and epiphenomenality, as alternative explanations at least with respect to a given competing variable W , can be accounted for in a mediation model by including W as an additional predictor or “covariate” in the models of M and Y . For example, Equations 1, 2, and 3 in the manuscript with the inclusion of W as a covariate would be

$$M = i_1 + aX + d_1W + e_M \quad (S4)$$

$$Y = i_2 + c'X + bM + d_2W + e_Y \quad (S5)$$

$$Y = i_3 + cX + d_3W + e_Y \quad (S6)$$

The addition of covariates is simple in any OLS regression program; covariates can be added to each of the ON statements in the Mplus code above, and the PROCESS and MEDiate macros also accept covariates.

References

- Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what is the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology*, 98, 550-558.
- Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, 11, 296-325.
- Coffman, D. L., & MacCallum, R. C. (2005). Using parcels to convert path models into latent variable models. *Multivariate Behavioral Research*, 40, 235-259.
- Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.
- Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis*. New York: The Guilford Press.

- Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67, 451-470.
- Kline, R. B. (2005). *Principles and practices of structural equation modeling* (2nd Ed.). New York: Guilford Press.
- Lau, R. S., & Cheung, G. W. (2012). Estimating and comparing specific mediation effects in complex latent variable models. *Organizational Research Methods*, 15, 3-16.
- 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.
- MacKinnon, D. P. (2008). *An introduction to statistical mediation analysis*. New York: Routledge.
- Mathieu, J. E., DeShon, R. P., & Bergh, D. D. (2008). Mediation inferences in organizational research. *Organizational Research Methods*, 11, 203-223.
- 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.
- Stone-Romero, E. F., & Rosopa, P. J. (2010). Research design options for testing mediation models and their implications for facets of validity. *Journal of Managerial Psychology*, 25, 697-712.

Copyright 2014-2016

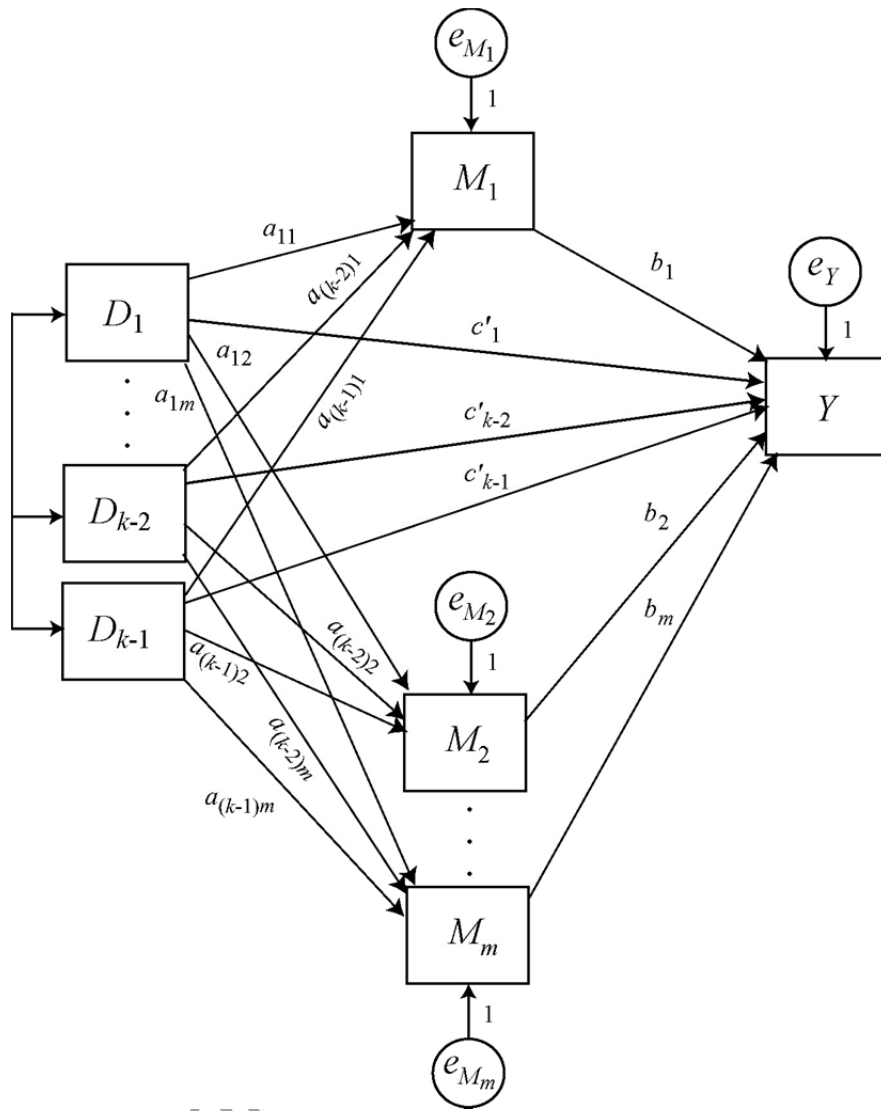


Figure S1. A multiple mediation model in path diagram form corresponding to a model with an independent variable X with k categories and m mediators operating in parallel. When estimating using a structural equation modeling program, it is recommended that the covariance between mediator errors be freely estimated (see e.g., Preacher and Hayes, 2008).