STATISTICAL HORIZONS

Mediation, Moderation, and Conditional Process Analysis

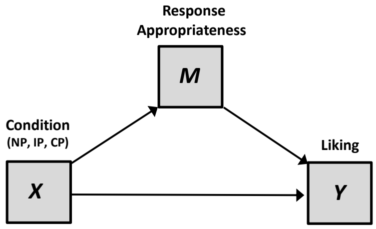
Andrew Hayes, Ph.D.

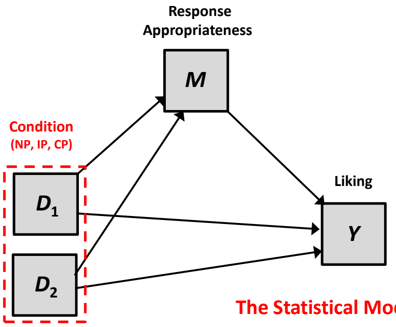
Upcoming Seminar:

July 10-14, 2017, Chicago, Illinois

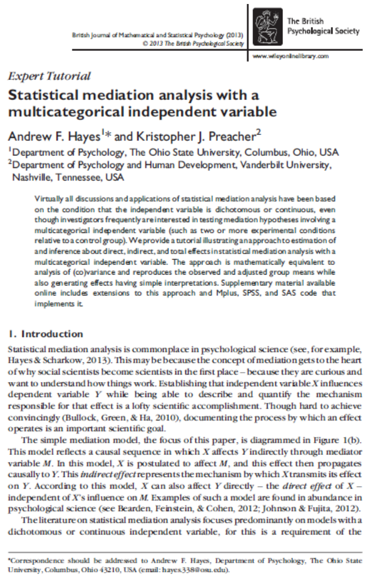
Mediation analysis with a multicategorical independent variable

The Conceptual Model





The Statistical Model



British Journal of Mathematical and Statistical Psychology (2013)
© 2013 The British Psychological Society
The British Psychological Society
www.wileyonlinelibrary.com

Expert Tutorial

Statistical mediation analysis with a multicategorical independent variable

Andrew F. Hayes[1]* and Kristopher J. Preacher[2]
[1]Department of Psychology, The Ohio State University, Columbus, Ohio, USA
[2]Department of Psychology and Human Development, Vanderbilt University, Nashville, Tennessee, USA

Virtually all discussions and applications of statistical mediation analysis have been based on the condition that the independent variable is dichotomous or continuous, even though investigators frequently are interested in testing mediation hypotheses involving a multicategorical independent variable (such as two or more experimental conditions relative to a control group). We provide a tutorial illustrating an approach to estimation of and inference about direct, indirect, and total effects in statistical mediation analysis with a multicategorical independent variable. The approach is mathematically equivalent to analysis of (co)variance and reproduces the observed and adjusted group means while also generating effects having simple interpretations. Supplementary material available online includes extensions to this approach and Mplus, SPSS, and SAS code that implements it.

1. Introduction

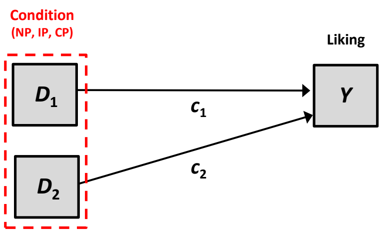
Statistical mediation analysis is commonplace in psychological science (see, for example, Hayes & Scharkow, 2013). This may be because the concept of mediation gets to the heart of why social scientists become scientists in the first place – because they are curious and want to understand how things work. Establishing that independent variable *X* influences dependent variable *Y* while being able to describe and quantify the mechanism responsible for that effect is a lofty scientific accomplishment. Though hard to achieve convincingly (Bullock, Green, & Ha, 2010), documenting the process by which an effect operates is an important scientific goal.

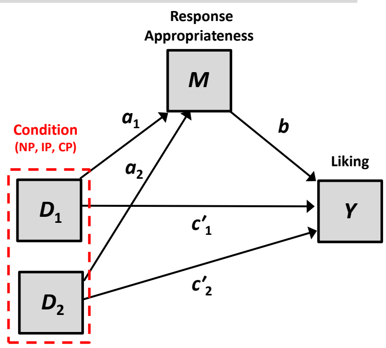
The simple mediation model, the focus of this paper, is diagrammed in Figure 1(b). This model reflects a causal sequence in which *X* affects *Y* indirectly through mediator variable *M*. In this model, *X* is postulated to affect *M*, and this effect then propagates causally to *Y*. This *indirect effect* represents the mechanism by which *X* transmits its effect on *Y*. According to this model, *X* can also affect *Y* directly – the *direct effect* of *X* – independent of *X*'s influence on *M*. Examples of such a model are found in abundance in psychological science (see Bearden, Feinstein, & Cohen, 2012; Johnson & Fujita, 2012).

The literature on statistical mediation analysis focuses predominantly on models with a dichotomous or continuous independent variable, for this is a requirement of the

*Correspondence should be addressed to Andrew F. Hayes, Department of Psychology, The Ohio State University, Columbus, Ohio 43210, USA (email: hayes.338@osu.edu).

(Relative) total, direct, and indirect effects





c_1 and c_2: *Relative* total effects of experimental condition on liking
c'_1 and c'_2: *Relative* direct effects of experimental condition on liking
a_1b and a_2b: *Relative* indirect effects of condition on liking through perceived response appropriateness.

$$c_1 = c'_1 + a_1b;\ \text{therefore, } a_1b = c_1 - c'_1$$
$$c_2 = c'_2 + a_2b;\ \text{therefore, } a_2b = c_2 - c'_2$$

The relative total effects partition perfectly into relative direct and relative indirect effects. The relative indirect effects are the relative total effects minus the relative direct effects.

Coding the groups

We'll use dummy codes setting the no protest condition to the reference group. Condition (variable name COND) is coded 0 (no protest condition), 1 (individual protest condition), and 2 (collective protest condition).

Condition	D_1	D_2
No protest	0	0
Individual	1	0
Collective	0	1

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

```
data protest;set protest;
d1 = (cond=1);
d2 = (cond=2);
if (cond=.) then d1=.;
if (cond=.) then d2=.;
run;
```

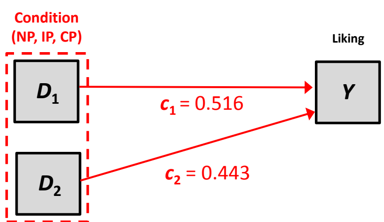
So effects for D_1 will compare individual protest to no protest, and effects for D_2 will compare collective protest to no protest.

The total effect of experimental condition on liking (*c* paths)

```
regression/dep = liking/method = enter d1 d2.
```

```
proc reg data=protest;model liking=d1 d2;run;
```

We did this already!



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.215[a]	.046	.031	1.03324

a. Predictors: (Constant), d2, d1

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.523	2	3.262	3.055	.051[b]
	Residual	134.515	126	1.068		
	Total	141.039	128			

a. Dependent Variable: LIKING: liking of the target

b. Predictors: (Constant), d2, d1

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	5.310	.161		32.908	.000
	d1	.516	.226	.233	2.287	.024
	d2	.443	.223	.202	1.986	.049

a. Dependent Variable: LIKING: liking of the target

Relative total effects

Relative to those told she did not protest, those told she individually protested liked her more on average (c_1 = 0.516, p = .024). Relative to those told she did not protest, those told she collectively protested also liked her more on average (c_2 = 0.443, p = .049).

The effect of experimental condition on perceived response appropriateness (a paths)

```
regression/dep = respappr/method = enter d1 d2.
```

```
proc reg data=protest;model respappr=d1 d2;run;
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.511[a]	.261	.249	1.16829

a. Predictors: (Constant), d2, d1

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	60.653	2	30.327	22.219	.000[b]
	Residual	171.977	126	1.365		
	Total	232.631	128			

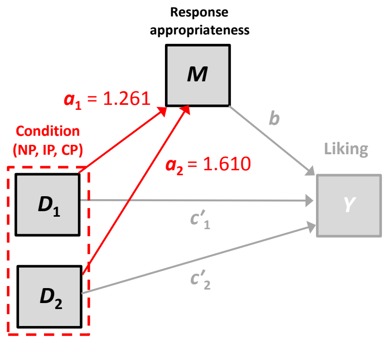
a. Dependent Variable: RESPAPPR: appropriateness of response

b. Predictors: (Constant), d2, d1

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.884	.182		21.288	.000
	d1	1.261	.255	.443	4.946	.000
	d2	1.610	.252	.572	6.384	.000

a. Dependent Variable: RESPAPPR: appropriateness of response



Relative to those told she did not protest, those told she individually protested felt her response was more appropriate on average ($a_1 = 1.261$ $p < .001$). Relative to those told she did not protest, those told she collectively protested felt her response was more appropriate on average ($a_2 = 1.610$, $p < .001$).

The direct effect of condition on liking (c' paths)

along with the effect of response appropriateness on liking (b path)

```
regression/dep = liking/method = enter respappr d1 d2.
```

```
proc reg data=protest;model liking=respappr d1 d2;run;
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.503[a]	.253	.235	.91798

a. Predictors: (Constant), d2, RESPAPPR: appropriateness of response, d1

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	35.703	3	11.901	14.123	.000[b]
	Residual	105.336	125	.843		
	Total	141.039	128			

a. Dependent Variable: LIKING: liking of the target

b. Predictors: (Constant), d2, RESPAPPR: appropriateness of response, d1

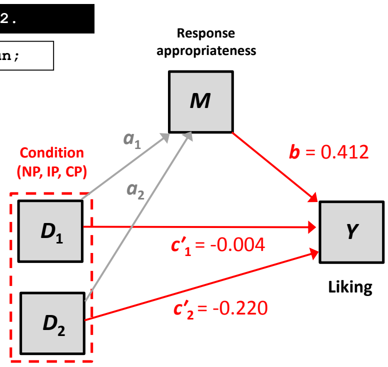
Coefficients[a]

b path

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.710	.307		12.071	.000
	RESPAPPR: appropriateness of response	.412	.070	.529	5.884	.000
	d1	-.004	.219	-.002	-.017	.987
	d2	-.220	.228	-.100	-.966	.336

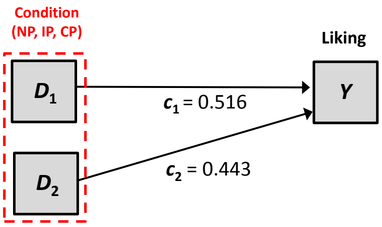
a. Dependent Variable: LIKING: liking of the target

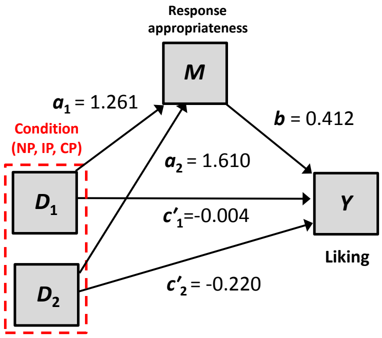
Relative direct effects



Controlling for perceived responses appropriateness, those told she individually protested did not like her any more, on average, than those told she did not protest ($c'_1 = -0.004$, $p = .987$). And those told she collectively protested did not like her any more, on average, than those told she did not protest ($c'_2 = -0.220$, $p = .336$). Holding condition constant, those who perceived her behavior as relatively more appropriate likely her relatively more ($b = 0.412$).

(Relative) total, direct, and indirect effects



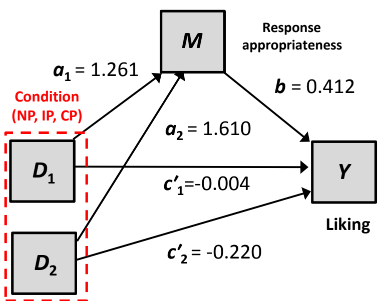


c_1 and c_2: *Relative* total effects of condition on liking (c_1 = 0.516, c_2 = 0.443).
c'_1 and c'_2: *Relative* direct effects of condition on liking (c'_1 = -0.004, c'_2 = -0.220).
a_1b and a_2b: *Relative* indirect effects of condition on liking through perceived response appropriateness
a_1b = 1.261(0.412) = 0.520, a_2b = 1.610(0.412) = 0.663

$c_1 = c'_1 + a_1b$: 0.516 = -0.004 + 1.261(0.412) = -0.004 + 0.520
$c_2 = c'_2 + a_2b$: 0.443 = -0.220 + 1.610(0.412) = -0.220 + 0.663

The relative total effects partition perfectly into relative direct and relative indirect effects. The relative indirect effects are the relative total effects minus the relative direct effects.

Relative indirect effects



a_1b and a_2b: *Relative* indirect effects of condition on liking through perceived response appropriateness, a_1b = 1.261(0.412) = 0.520, a_2b = 1.610(0.412) = 0.663

The relative indirect effects quantify group differences in *Y* that result from the effect of *X* on *M* which in turn affects *Y*. Inference is best based on a bootstrap confidence interval. More on this soon.

Relative total, direct, and indirect effects

	Perceived Response Appropriateness (M)		Liking (Y)		
	$\bar{M}$	SD	$\bar{Y}$	SD	$\bar{Y}^*$
No protest (n = 41)	3.884	1.457	5.310	1.302	5.715
Individual protest (n = 43)	5.145	1.075	5.826	0.819	5.711
Collective protest (n = 45)	5.494	0.936	5.753	0.936	5.495
All groups combined	4.866	1.348	5.637	1.050	5.637

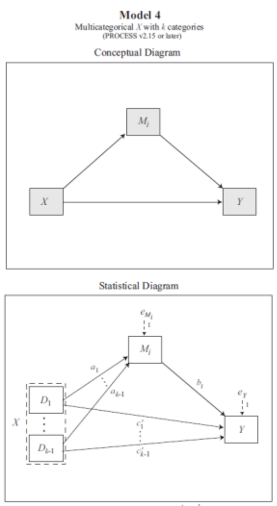
$a_1 = 1.261$ $a_2 = 1.610$ $c_1 = 0.516$ $c_2 = 0.443$ $c'_1 = -0.004$ $c'_2 = -0.220$

$\bar{Y}^*$ = adjusted mean, adjusted to the sample mean of perceived response appropriateness.

$c_1 = c'_1 + a_1 b$: $0.516 = -0.004 + 1.261(0.412) = -0.004 + 0.520$

$c_2 = c'_2 + a_2 b$: $0.443 = -0.220 + 1.610(0.412) = -0.220 + 0.663$

Estimation using PROCESS



New to version 2.15, PROCESS has an option in model 4 for specifying X as a multicategorical variable with up to 9 categories. Four options are available for coding the groups.

"MCX=1" tells PROCESS that the focal predictor X is a multicategorical variable and to use dummy coding to represent the groups. Other coding options are available. *See the documentation addendum in your course files.*

MCX	Coding system
1	Simple dummy coding
2	Sequential ("adjacent categories") coding
3	Helmert coding
4	Effect coding

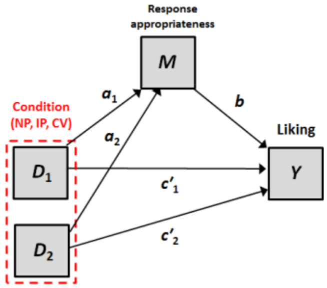
```
process vars=liking respappr cond/y=liking/m=respappr/x=cond/model=4/mcx=1/total=1/boot=10000.
```

```
%process (data=protest,vars=liking respappr cond,y=liking,m=respappr,x=cond,model=4,mcx=1,total=1,
    boot=10000);
```

PROCESS output

```
Model = 4
    Y = liking
    X = cond
    M = respappr

Sample size
        129

Coding of categorical X variable for analysis:
  cond     D1     D2
   .00    .00    .00
  1.00   1.00    .00
  2.00    .00   1.00
```

D1 codes individual protest, D2 codes collective protest. No protest Is the reference group. (The group with the numerically smallest value on the categorical variable is always the reference)




```
*************************************************************************
Outcome: respappr

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .5106      .2607     1.3649    22.2190     2.0000   126.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
Constant     3.8841      .1825    21.2881      .0000     3.5231     4.2452
D1           1.2612      .2550     4.9456      .0000      .7565     1.7659
D2           1.6103      .2522     6.3842      .0000     1.1111     2.1095
```

$$\hat{M} = 3.884 + 1.261 D_1 + 1.610 D_2$$

a_1 path

a_2 path

```
*************************************************************************
Outcome: liking

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .5031      .2531      .8427    14.1225     3.0000   125.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
Constant     3.7103      .3074    12.0711      .0000     3.1020     4.3187
respappr      .4119      .0700     5.8844      .0000      .2734      .5504
D1           -.0037      .2190     -.0169      .9865     -.4371      .4297
D2           -.2202      .2280     -.9658      .3360     -.6715      .2310
```

$$\hat{Y} = 3.710 - 0.004 D_1 - 0.220 D_2 + 0.412 M$$

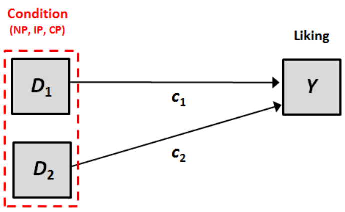
b path

c'_1 path

c'_2 path

Output P

PROCESS output




```
************************** TOTAL EFFECT MODEL ****************************
Outcome: liking

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .2151      .0463     1.0676     3.0552     2.0000   126.0000      .0506

Model
              coeff         se          t          p       LLCI       ULCI
Constant     5.3102      .1614    32.9083      .0000     4.9909     5.6296
D1            .5158      .2255     2.2870      .0239      .0695      .9621
D2            .4431      .2231     1.9863      .0492      .0016      .8845
```

c_1 path

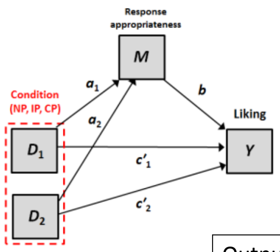
c_2 path

$$\hat{Y} = 5.310 + 0.516 D_1 + 0.443 D_2$$

Output P

PROCESS output

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

Relative total effects of X of Y
        coeff        se         t          p       LLCI      ULCI
D1      .5158     .2255    2.2870      .0239     .0695     .9621
D2      .4431     .2231    1.9863      .0492     .0016     .8845

Omnibus test of total effect of X on Y
       R-sq          F        df1        df2          p
      .0463     3.0552     2.0000   126.0000      .0506
=====

Relative direct effects of X on Y
        coeff        se         t          p       LLCI      ULCI
D1     -.0037     .2190    -.0169      .9865    -.4371     .4297
D2     -.2202     .2280    -.9658      .3360    -.6715     .2310

Omnibus test of direct effect of X on Y
       R-sq          F        df1        df2          p
      .0087      .7286     2.0000   125.0000      .4846
=====
Relative indirect effect(s) of X on Y through:
 respappr

            Effect   SE(boot)      LLCI      ULCI
D1           .5195      .1490     .2728     .8564
D2           .6633      .1633     .3792    1.0271
Omnibus      .1026      .0349     .0472     .1820

----------
```



Output P

Indirect effect a_1b with bootstrap confidence interval

Indirect effect a_2b with bootstrap confidence interval

Those told she individually protested liked her more than those told she did not protest because protesting was perceived as more appropriate than not, which in turn enhanced liking (point estimate = 0.520, 95% CI: 0.273 to 0.856). There is no direct effect of individually protesting on liking. Those told she collectively protested liked her more than those told she did not protest because protesting was perceived as more appropriate than not, which in turn enhanced liking (point estimate= 0.663, 95% CI: 0.379 to 1.027). There is no direct effect of collectively protesting on liking.

Omnibus inference

PROCESS gives us tests of the *k*-1 relative total effects. It also provides a test of equality of the *k* group means on Y---the "omnibus" total effect. This is equivalent to a single-factor ANOVA.

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

Relative total effects of X of Y
        coeff        se         t          p       LLCI      ULCI
D1      .5158     .2255    2.2870      .0239     .0695     .9621
D2      .4431     .2231    1.9863      .0492     .0016     .8845

Omnibus test of total effect of X on Y
       R-sq          F        df1        df2          p
      .0463     3.0552     2.0000   126.0000      .0506
=====

Relative direct effects of X on Y
        coeff        se         t          p       LLCI      ULCI
D1     -.0037     .2190    -.0169      .9865    -.4371     .4297
D2     -.2202     .2280    -.9658      .3360    -.6715     .2310

Omnibus test of direct effect of X on Y
       R-sq          F        df1        df2          p
      .0087      .7286     2.0000   125.0000      .4846
=====
Relative indirect effect(s) of X on Y through:
 respappr

            Effect   SE(boot)      LLCI      ULCI
D1           .5195      .1490     .2728     .8564
D2           .6633      .1633     .3792    1.0271
Omnibus      .1026      .0349     .0472     .1820

----------
```

Output P

Test of the "omnibus" total effect.

The three conditions differ on average in liking of the attorney , $F(2,126) = 3.055$, $p = .051$.