Mediation, Moderation, and Conditional Process Analysis: A Second Course

Amanda Kay Montoya

Offered by

GLOBAL SCHOOL EMPIRICAL RESEARCH METHODS

June 2025

What You'll Need

- This course is hands-on. You'll want to have a computer in front of you with either SPSS (27 or later), SAS (9.2 or later, with PROC IML) or R (v3.6 or later). But if not, you'll still benefit.
- Various data, program, and PDF files available in the course materials, including a copy of these slides.
- Comfort with the fundamentals of mediation, moderation, and conditional process analysis as covered in *Introduction to mediation, moderation, and conditional process analysis* and/or the first course.
- Comfort with ordinary least squares regression analysis.
- Some prior experience working with PROCESS is helpful. A refresher or tutorial on setting up PROCESS in SPSS, SAS, and R is in one of the first modules of this course. Most of the computations can be and will be done with PROCESS.

Things will get complex quickly. You need to know the fundamentals already.

Workshop Topics

Review Content from Week I (1 day)

Core Topics:

- Multiple mediator models with a single moderator (.5 days)
    - Differential dominance (Parallel mediator model with a single moderator)
    - Moderation in serial mediation
- Multicategorical variables (1.5 days)
    - Mediation
    - Moderation
    - Moderated Mediation
- Models with Multiple Moderators (1.5 days)
    - Moderation
    - Moderated Mediation
- Creating and Editing models in PROCESS (.5 days)

Workshop Topics

Alternative topics:

- Causal mediation analysis (0.5 days)
- Simple repeated-measures designs (1 day)
    - Mediation
    - Moderation
    - Moderated Mediation
- Multilevel (1 day)
    - Mediation
    - Moderated Mediation
- Handling Measurement Error in PROCESS
- Regression Diagnostics with PROCESS
- Models with Dichotomous Outcomes

Review of fundamentals

A study of teams

Cole, M. S., Walker, F., & Bruch, H. (2008). Affective mechanisms linking dysfunctional team behavior to performance in work teams: A moderated mediation study. *Journal of Applied Psychology, 93*, 945-958.

Dysfunctional team behavior
How often members of the team do things to weaken the work of others or hinder change and innovation.

Negative affective tone of the work environment
How often team members report feeling negative emotions at work such as “angry”, “disgust”, and so forth.

Emotional expressiveness of the team
The supervisor’s perception as to how easy it is to tell how team members are feeling.

Team performance
Supervisor’s judgment as to the team’s efficiency, ability to get task done in a timely fashion with high quality.

The data: teams (SPSS and SAS)

teams.sav

teams.sav [DataSet1] - IBM SPSS Statistics Data Editor

	dysfunc	negtone	negexp	perform	division	d1	d2	d3
1	-.23	-.51	-.49	.12	1	1	0	0
2	-.13	.22	-.49	.52	1	1	0	0
3	.00	-.08	.84	-.08	1	1	0	0
4	-.33	-.11	.84	-.08	1	1	0	0
5	.39	-.48	.17	.12	1	1	0	0
6	1.02	.72	-.82	1.12	1	1	0	0
7	-.35	-.18	-.66	-.28	1	1	0	0
8	-.23	-.13	-.16	.32	2	0	1	0
9	.39	.52	-.16	-1.08	2	0	1	0
10	-.08	-.26	-.16	-.28	2	0	1	0
11	-.23	1.08	-.16	-1.08	2	0	1	0
12	.09	.53	.50	-.28	4	0	0	0
13	-.29	-.19	.84	-.28	4	0	0	0
14	-.06	.15	.50	-.88	3	0	0	1
15	.27	.73	-.16	-.08	3	0	0	1
16	.18	-.18	-.16	.32	3	0	0	1
17	.38	.22	-1.16	.52	3	0	0	1
18	.43	.52	.50	-.68	3	0	0	1
19	.03	.20	.60	-.23	1	1	0	0
20	-.50	-.68	.26	.17	1	1	0	0
21	.03	.31	-.40	.17	1	1	0	0
22	.41	.75	-1.06	-.23	1	1	0	0
23	-.08	-.48	-.06	-.23	1	1	0	0

The data file in SPSS is in SPSS data (.sav) form and ready to analyze.

teams.sas

teams

```
data teams;
input dysfunc negtone negexp perform division d1 d2 d3;
cards;
   -.23    -.51    -.49     .12    1    1    0    0
   -.13     .22    -.49     .52    1    1    0    0
    .00    -.08     .84    -.08    1    1    0    0
   -.33    -.11     .84    -.08    1    1    0    0
    .39    -.48     .17     .12    1    1    0    0
   1.02     .72    -.82    1.12    1    1    0    0
   -.35    -.18    -.66    -.28    1    1    0    0
   -.23    -.13    -.16     .32    2    0    1    0
    .39     .52    -.16   -1.08    2    0    1    0
   -.08    -.26    -.16    -.28    2    0    1    0
   -.23    1.08    -.16   -1.08    2    0    1    0
    .09     .53     .50    -.28    4    0    0    0
   -.29    -.19     .84    -.28    4    0    0    0
   -.06     .15     .50    -.88    3    0    0    1
    .27     .73    -.16    -.08    3    0    0    1
    .18    -.18    -.16     .32    3    0    0    1
    .38     .22   -1.16     .52    3    0    0    1
    .43     .52     .50    -.68    3    0    0    1
    .03     .20     .60    -.23    1    1    0    0
   -.50    -.68     .26     .17    1    1    0    0
    .03     .31    -.40     .17    1    1    0    0
    .41     .75   -1.06    -.23    1    1    0    0
   -.08    -.48    -.06    -.23    1    1    0    0
    .66    1.45     .60   -1.23    1    1    0    0
```

SAS data are provided to you in the form of a SAS program that creates a temporary SAS data file when executed.

The data: teams.csv

Data files for R users are available as comma-delimited text (CSV) files, with variable names at the top (the "header"). Here is an excerpt of the teams.csv data file:

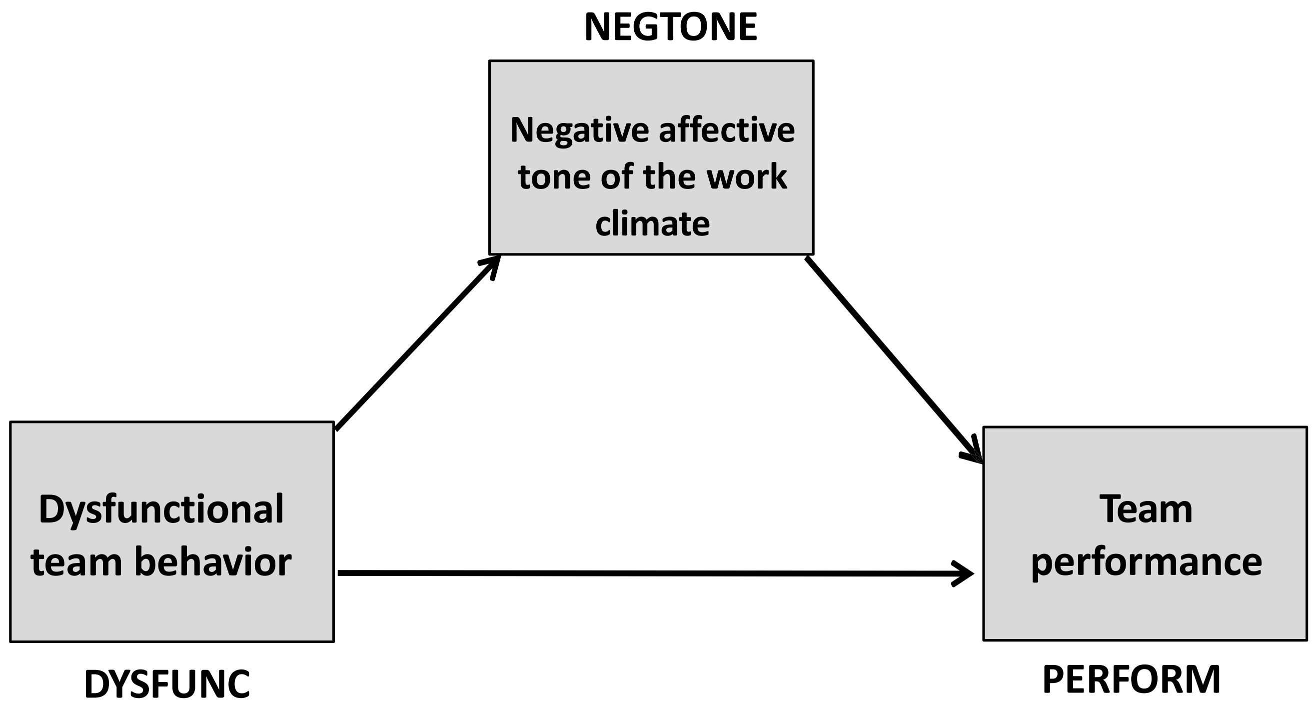
```
dysfunc,negtone,negexp,perform,division,d1,d2,d3
-.23,-.51,-.49,.12,1,1,0,0
-.13,.22,-.49,.52,1,1,0,0
0,-.08,.84,-.08,1,1,0,0
-.33,-.11,.84,-.08,1,1,0,0
```

The R code below reads the data and stores them in a data frame named "teams." Put the data file in a location on your computer and then change the path in the code to correspond to that location. Or set a new working directory (setwd) and leave off the path.

```
teams<-read.table("c:/mmcpa/teams.csv", sep=",",header=TRUE)
head(teams)
```

```
> teams<-read.table("c:/mmcpa/teams.csv", sep=",",header=TRUE)
> head(teams)
  dysfunc negtone negexp perform division d1 d2 d3
1   -0.23   -0.51  -0.49    0.12        1  1  0  0
2   -0.13    0.22  -0.49    0.52        1  1  0  0
3    0.00   -0.08   0.84   -0.08        1  1  0  0
4   -0.33   -0.11   0.84   -0.08        1  1  0  0
5    0.39   -0.48   0.17    0.12        1  1  0  0
```

Our mediation question

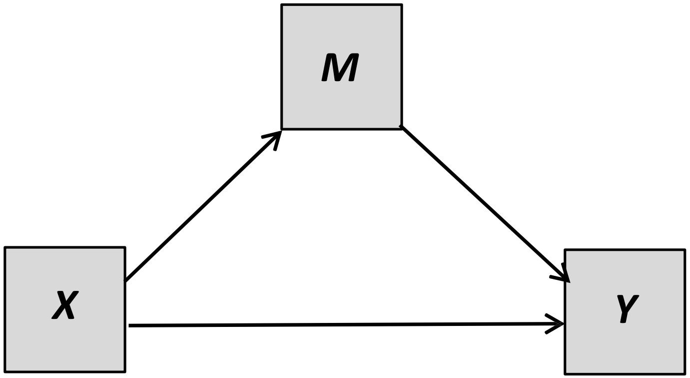


Does dysfunctional team behavior influence team performance by increasing the negative emotional tone of the work climate which in turn affects performance?

A tenet of modern mediation analysis: This question is not asked contingent on evidence of simple association between *X* (dysfunctional team behavior) and *Y* (performance)

The most basic intervening variable model

- For *M* to be an intermediary variable, it must be located *causally between X* and *Y*.
- *M* is sometimes called a "mediator", but it goes by other names as well.



Mediator models are causal models and carry with them the usual criteria for making causal claims. These are often difficult to establish unequivocally, statistically or otherwise. We do our best, recognizing that sometimes our data do not lend themselves to causal claims. That is, theory is sometimes the sole foundation upon which our causal claims rest. That's ok so long as we recognize this.

Understanding Cause & Effect

As scientists we're often looking to support a claim that "X causes Y." Many of us are familiar with the phrase "correlation is not causation." But what then is needed to support a claim of cause?

Often we rely on experimentation to help us support the claim of cause. But what happens when we cannot (ethically or practically) manipulate our causal variable?

Consider the claim "Smoking tobacco causes lung cancer." Is it unethical to randomize people to smoke or not smoke. How then do we know this claim is true?

John Stewart Mill - Necessary Conditions for Cause:

1. Temporal Ordering
2. Elimination of competing explanations
3. Covariation

Research on tobacco use easily found evidence for 1 & 3, and slowly over time accumulation of evidence supported 2.

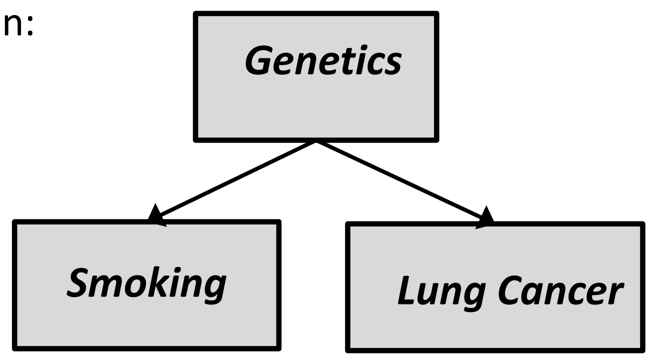
Understanding Cause & Effect

While **covariation** is a necessary condition of cause, we typically want assurance that our estimate of a causal effect is **unbiased**, for this we do not need to consider covariation.

An unbiased estimate of a causal effect means that if we had the *whole population* and estimated the causal effect using our statistical estimand, we could get the correct value.

Imagine we know that following is true in the population:

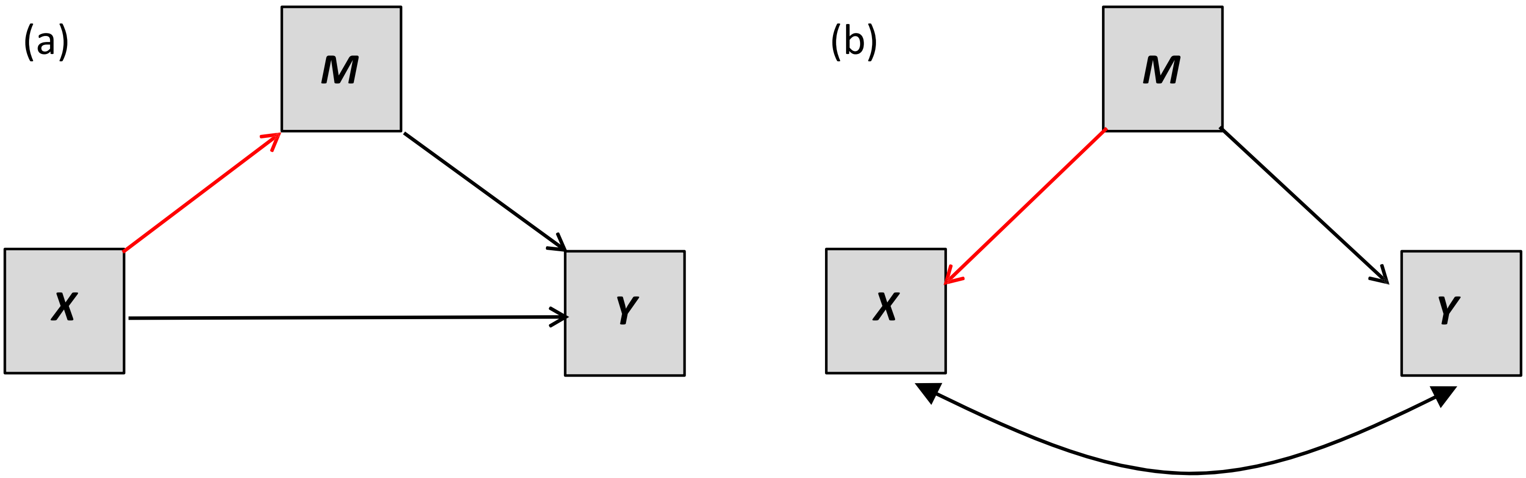
Smoking does not cause lung cancer
Genetics causes smoking and lung cancer



If I calculate the correlation between smoking and lung cancer, in the population, it will not be zero (the true population causal effect), but rather because there is an *alternative explanation* my estimate is biased.

Mediation and spuriousness

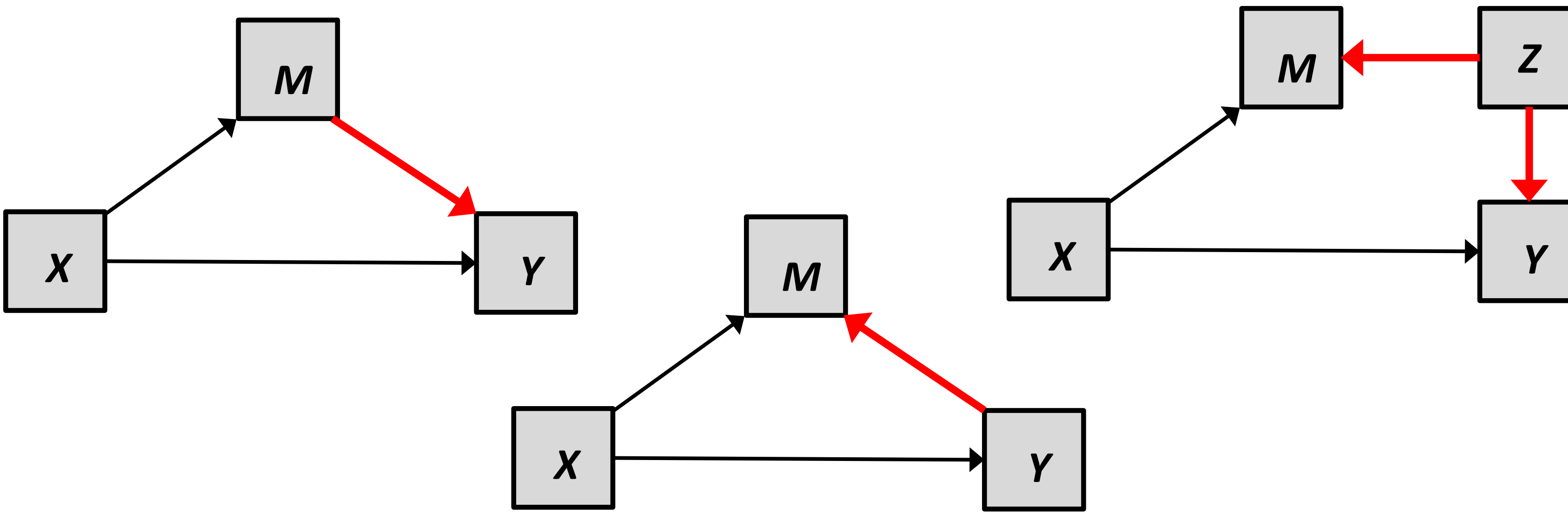
Statistical mediation analysis cannot distinguish between mediation and spuriousness. If (b) can be deemed plausible, that weakens the case for (a) regardless of what the data analysis tells you.



So the inferences one can make are, as always, design-bound. Mediation is a causal process, but causal claims are only justified if the data allow such claims, regardless of what the statistics say.

Causal order

Manipulation of and random assignment to *X* affords causal inference for the effect of *X* on *M* and *Y*, but not the effect of *M* on *Y*. We cannot establish causal order for the *M*-*Y* path using the methods we will be talking about. Theory is important. Multiple studies can help, one of which involves manipulation of *M*.



When *X* is not experimentally manipulated, all paths are subject to potential alternative causal orders. But that doesn't mean we cannot or should not plunge ahead, despite these limitations in interpretation.

Total, Direct, and Indirect Effects

Let a, b, c, and c' be quantifications of causal effects, such as regression coefficients in an OLS model (or maximum likelihood path estimates in a structural equation model)

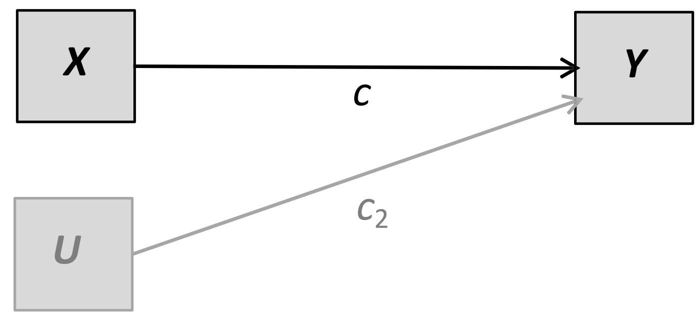
A simple mediation model with an (optional) adjustment for a potential confounding variable (U)

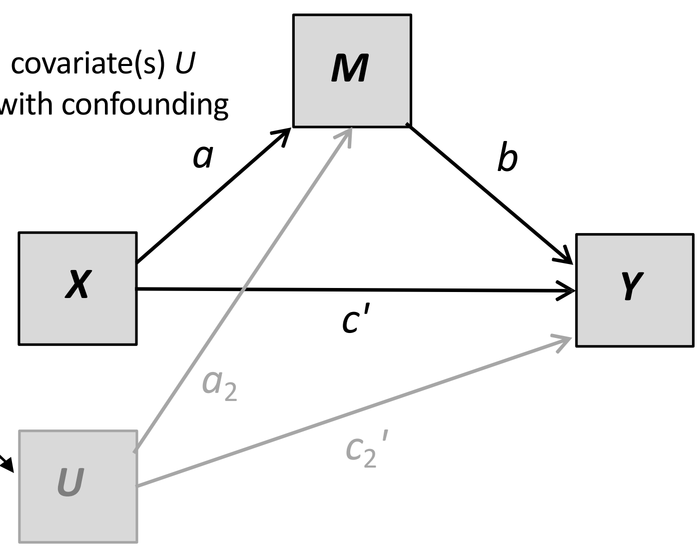
$$Y_i = c_0 + cX_i + c_2U_i$$

$$M_i = a_0 + aX_i + a_2U_i$$

$$Y_i = c'_0 + c'X_i + bM_i + c'_2U_i$$

Optional covariate(s) U to deal with confounding





total effect = direct effect + indirect effect

$c = c' + (a \times b)$

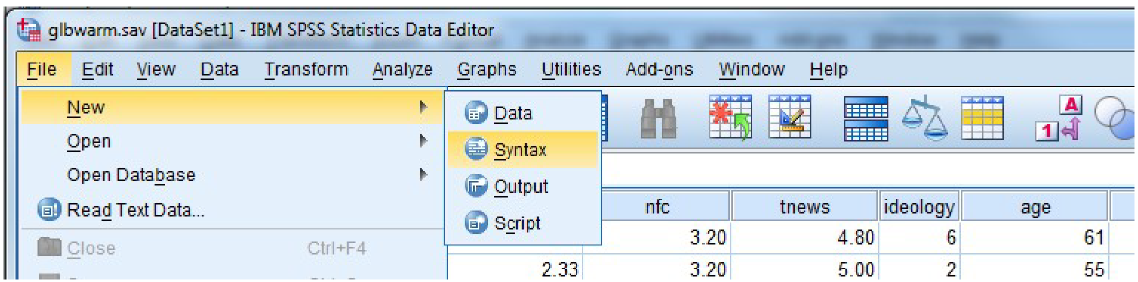
indirect effect = total effect – direct effect

$(a \times b) = c - c'$

Using SPSS syntax

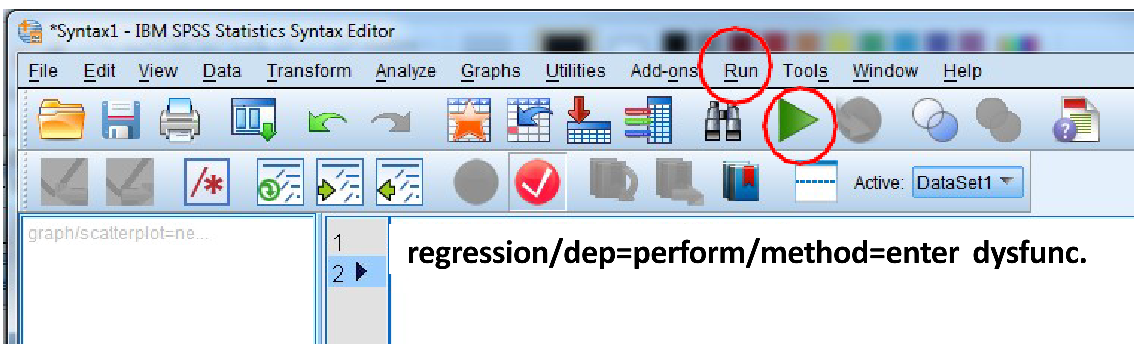
We will use syntax to instruct SPSS what to do in this class. There are many benefits of learning how to write SPSS syntax. See my “Little Syntax Guide” available at afhayes.com

(1) Open a new syntax window (File > New > Syntax)



(3) Click and drag to highlight code you want to execute and press the button or select various options under “Run” in the syntax window menu.

(2) Type your command(s) into the blank window that opens



The total effect of dysfunctional behavior

The total effect of *X* is the simple association between *X* (dysfunctional team behavior) and *Y* (performance).

SPSS:
```
regression/dep=perform/method=enter dysfunc.
```

SAS:
```
proc reg data=teams;model perform=dysfunc;run;
```

R:
```
summary(lm(perform~dysfunc,data=teams))
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.03549    0.06812  -0.521    0.604
dysfunc      0.11025    0.18402   0.599    0.551

Residual standard error: 0.5253 on 58 degrees of freedom
Multiple R-squared:  0.00615,   Adjusted R-squared:  -0.01098
F-statistic: 0.3589 on 1 and 58 DF,  p-value: 0.5514
```

c = 0.110, $t(58)$=0.599, p = .551

It might seem that the evidence does not support a claim that dysfunctional behavior influences performance. But lack of correlation does not disconfirm causation. The total effect can be a very misleading characterization of the effect of *X* on *Y* because it aggregates across pathways of influence---the direct and indirect effects.

Regression Coefficient Interpretations

Elements of interpretations:

- One unit difference on X corresponds to *b* unit difference on Y
- Control for all other predictors in the model
- Difference is **estimated**
- Avoid implying causality unless it is valid based on design

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.03549    0.06812  -0.521    0.604
dysfunc      0.11025    0.18402   0.599    0.551

Residual standard error: 0.5253 on 58 degrees of freedom
Multiple R-squared:  0.00615,   Adjusted R-squared:  -0.01098
F-statistic: 0.3589 on 1 and 58 DF,  p-value: 0.5514
```

TRY WRITING AN INTERPRETATION FOR C

21st century mediation analysis

- No longer based on the causal steps, “criteria for mediation” approach popularized by Baron and Kenny (1986) and that still is waning in use.
- Focus placed on the direct and indirect effect(s) rather than the individual paths and tests of significance on those paths.
- No need to condition the hunt for an indirect effect(s) on a statistically significant total effect.
- “Complete” and “partial” mediation are empty and dying concepts.
- Use of tests of indirect effect(s) that respect the nonnormality of the sampling distribution of the product of regression coefficients, such as bootstrap confidence intervals.

The bootstrap confidence interval

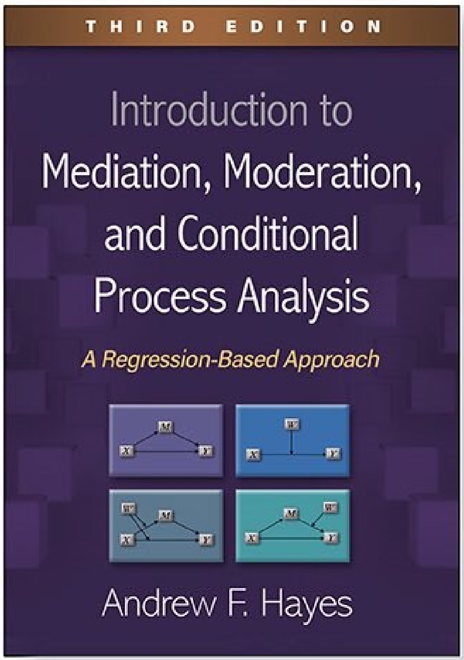
Bootstrapping empirically simulates the sampling distribution of the indirect effect and uses that simulation to generate a confidence interval for estimation and hypothesis testing. **It has become the preferred inferential method for inference about indirect effects in mediation analysis.**

- Take a random sample of size *n* from the sample *with replacement.*
- Estimate the indirect effect in this "bootstrap sample."
- Repeat (1) and (2) a total of *k* times, where *k* is at least a few thousand. I recommend 5,000 to 10,000 or so, as well as seeding the random number generator.
- Use the distribution of the indirect effect over the *k* bootstrap samples as an empirical approximation of the sampling distribution of the indirect effect.
- The lower and upper bounds of a 95% bootstrap confidence interval using the *percentile method* are the 2.5th and 97.5th percentiles of the *k* bootstrap estimates of the indirect effect.

This approach makes no assumptions about the shape of the sampling distribution of *ab*.

PROCESS

- An observed variable OLS regression-based modeling tool for moderation, mediation, and conditional process analysis.

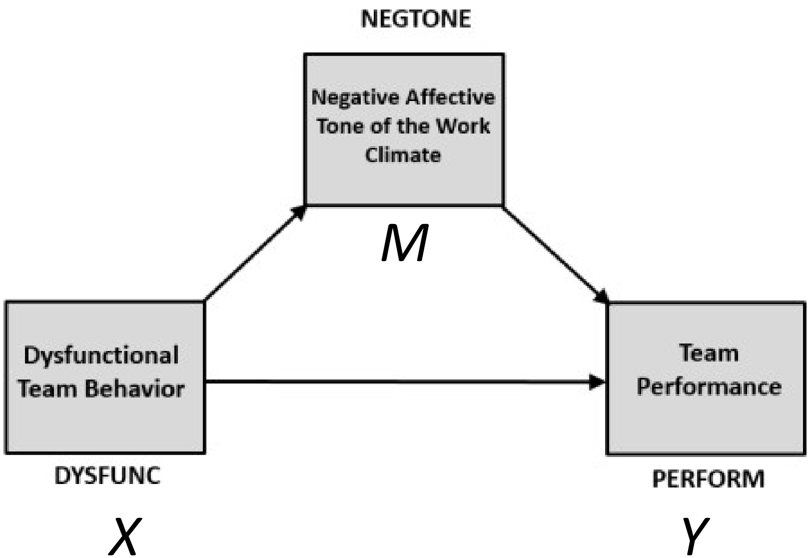


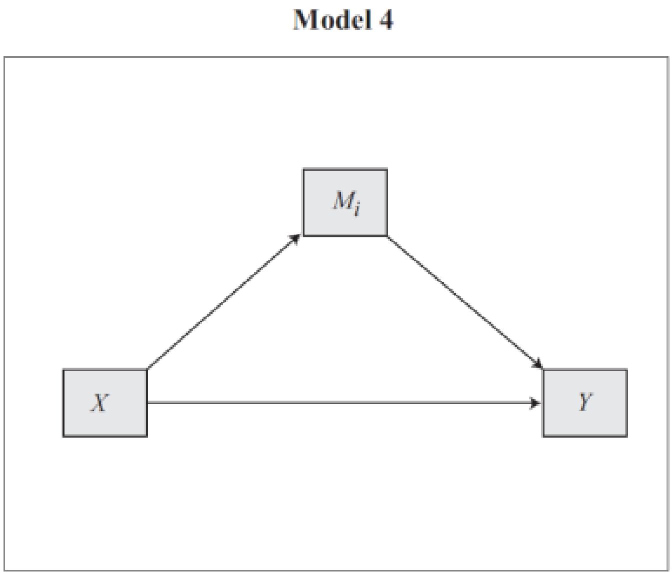
- First released in March of 2012, and re-released (version 3) in 2018 with many new features. PROCESS is documented in *Introduction to Mediation, Moderation, and Conditional Process Analysis* (2nd edition) www.guilford.com
- Available for SPSS (in macro and "custom dialog" form), SAS, and R.
- PROCESS is free and can be downloaded from www.processmacro.org. You've been given the most current version with your course files.

- It has become widely used in many disciplines. It makes difficult, tedious things easy.

Estimation of the model using PROCESS

PROCESS Model 4






```
process y=perform/x=dysfunc/m=negtone/model=4/total=1/boot=10000/seed=3341.
```

```
%process (data=teams,y=perform,x=dysfunc,m=negtone,model=4,total=1,boot=10000,
   seed=3341)
```

```
process(data=teams,y="perform",x="dysfunc",m="negtone",model=4,total=1,
   boot=10000,seed=3341)
```

PROCESS output

```
Model  : 4
    Y  : perform
    X  : dysfunc
    M  : negtone

Sample
Size:  60

Custom
Seed:     3341

**************************************************************************
OUTCOME VARIABLE:
 negtone
```

$$\widehat{M}_i = 0.026 + 0.620X_i$$

Model Summary

R	R-sq	MSE	F	df1	df2	p
.4384	.1922	.2268	13.7999	1.0000	58.0000	.0005

Model

path *a*	coeff	se	t	p	LLCI	ULCI
constant	.0257	.0618	.4159	.6791	-.0979	.1493
dysfunc	.6198	.1668	3.7148	.0005	.2858	.9537

TRY WRITING AN INTERPRETATION FOR a-path

PROCESS output

OUTCOME VARIABLE:
perform

$$\widehat{Y}_i = -0.0218 + 0.4414X_i - 0.5344M_i$$

Model Summary

R	R-sq	MSE	F	df1	df2	p
.4893	.2394	.2149	8.9725	2.0000	57.0000	.0004

TRY WRITING AN INTERPRETATION FOR b-path

Model

	coeff	se	t	p	LLCI	ULCI	
constant	-.0218	.0602	-.3615	.7190	-.1423	.0988	
dysfunc	.4414	.1807	2.4434	.0177	.0797	.8032	path *c'*
negtone	-.5344	.1278	-4.1814	.0001	-.7903	-.2785	path *b*

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

OUTCOME VARIABLE:
perform

$$\widehat{Y}_i = -0.036 + 0.110X_i$$

Optional total effect model generated with the use of the **total=1** option)

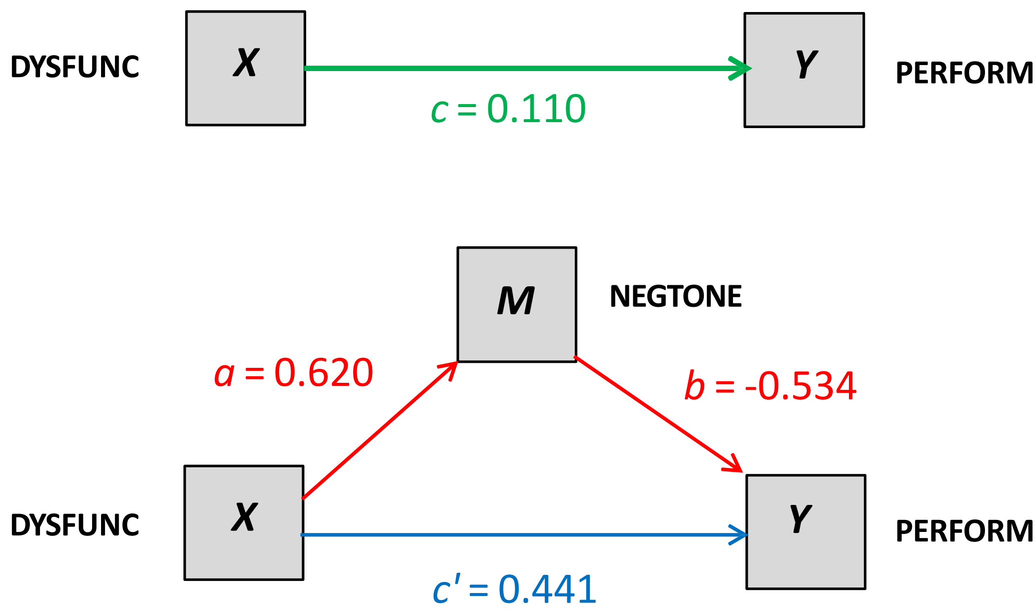
Model Summary

R	R-sq	MSE	F	df1	df2	p
.0784	.0062	.2760	.3589	1.0000	58.0000	.5514

Model

	coeff	se	t	p	LLCI	ULCI	
constant	-.0355	.0681	-.5210	.6044	-.1718	.1009	
dysfunc	.1102	.1840	.5991	.5514	-.2581	.4786	path *c*

Total, direct, and indirect effects of *X*



Direct effect of X on $Y = c' = 0.441$

Indirect effect of X on Y via $M = ab = 0.620(-0.534) = -0.331$

Total effect of X on $Y = c' + ab = 0.441 + -0.331 = 0.110 = c$

PROCESS output

```
************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Total effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .1102      .1840      .5991      .5514     -.2581      .4786

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .4414      .1807     2.4434      .0177      .0797      .8032

Indirect effect(s) of X on Y:
           Effect     BootSE   BootLLCI   BootULCI
negtone    -.3312      .1456     -.6453     -.0852
```

path *c*

path *c′*

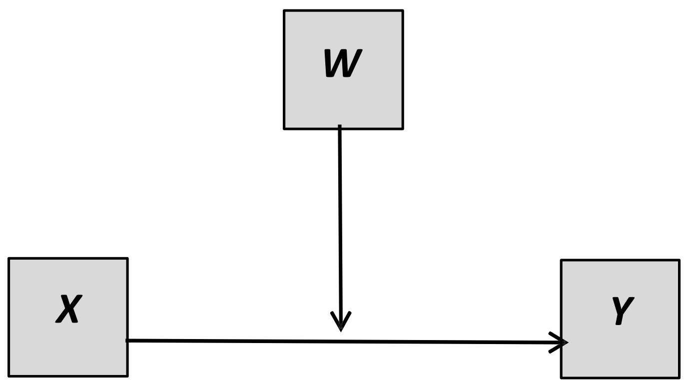
ab with 95% bootstrap confidence interval

TRY WRITING AN INTERPRETATION FOR INDIRECT EFFECT

Teams one unit higher in dysfunctional behavior are expected to be 0.620 units higher in negative affective tone of the work climate (a = 0.620), and more negative tone is related to lower performance (b = -0.534). So dysfunctional behavior negatively affects performance by increasing the negativity of the work climate which in turn lowers performance (point estimate: -0.331, 95% bootstrap CI = -0.645 to -0.085). Independent of this mechanism, dysfunctional teams perform significantly better (direct effect = 0.441, p = 0.018, 95% CI = 0.079 to 0.803). **ALL THIS WITHOUT A SIGNIFICANT TOTAL EFFECT.**

Moderation

Moderation. The effect of X on Y can be said to be *moderated* if its size or direction is dependent on some third variable W. It tells us about the conditions that facilitate, enhance, or inhibit the effect, or for whom or what the effect is large vs. small, present versus absent, and so forth.



In this figure, W is depicted to *moderate* the size of the effect of X on Y, meaning that the size of the effect of X on Y depends on W. In such a case, we say W is the *moderator* of the $X \rightarrow Y$ relationship, or that X and W *interact* in their influence on Y. X is sometimes called the focal predictor, and W the moderator.

"Moderation" and "Conditional Effect"

Moderation is said to exist when antecedent variable *X*'s effect on consequent *Y* is contingent on some third variable *W*.

"Simple linear moderation" is typically estimated by allowing *X*'s effect on *Y* to depend linearly on *W* (although other forms of moderation or "interaction" are possible):

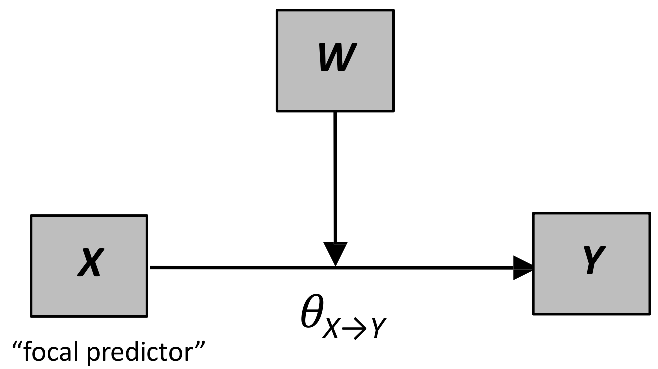
$$\hat{Y}_i = b_0 + (b_1 + b_3 W_i)X_i + b_2 W_i$$

$$= b_0 + b_1 X_i + b_2 W_i + b_3 X_i W_i$$

In this model, the *conditional effect* of *X* on *Y*, $\theta_{X \to Y}$, is a linear function of *W*: $\theta_{X \to Y} = b_1 + b_3 W$. That is,

$$\hat{Y}_i = b_0 + \theta_{X \to Y} X_i + b_2 W_i$$

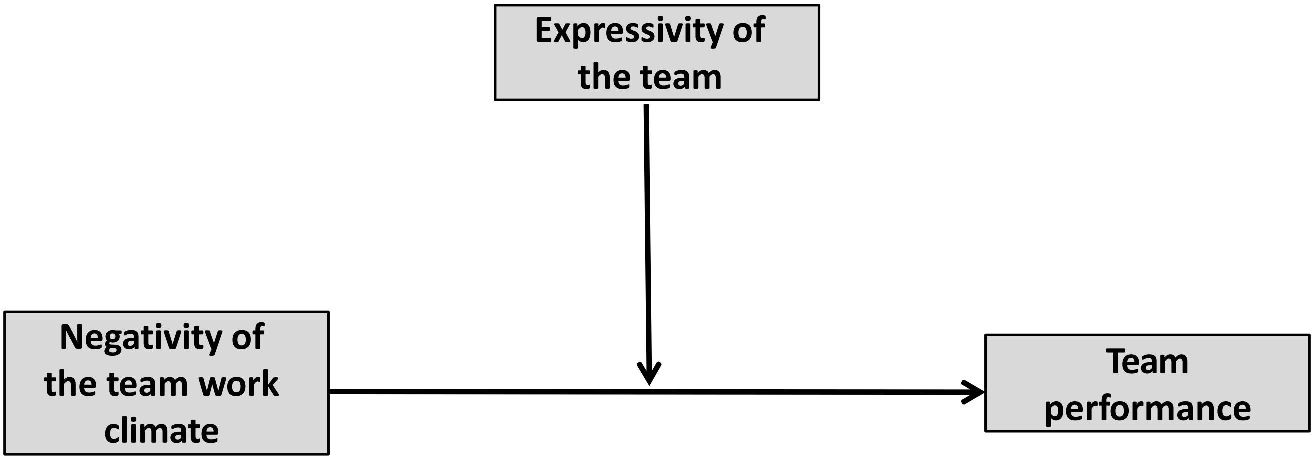
where $\theta_{X \to Y} = b_1 + b_3 W$



Two cases who differ by one unit on *X* are estimated to differ by $b_1 + b_3 W$ units on *Y*.

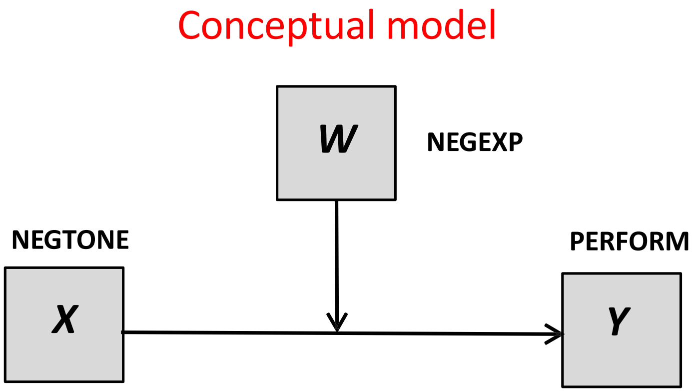
There is no effect of *X* on *Y* that one can reduce to a single estimate, for the effect of *X* on *Y* depends on *W* unless b_3 is zero. An inference about the coefficient for the product of *X* and *W* is a test of moderation. The effect of *X* is a *conditional effect*. The answer to the question "What is the size of *X*'s effect?" therefore should have some kind of condition placed on it. "It depends on *W*."

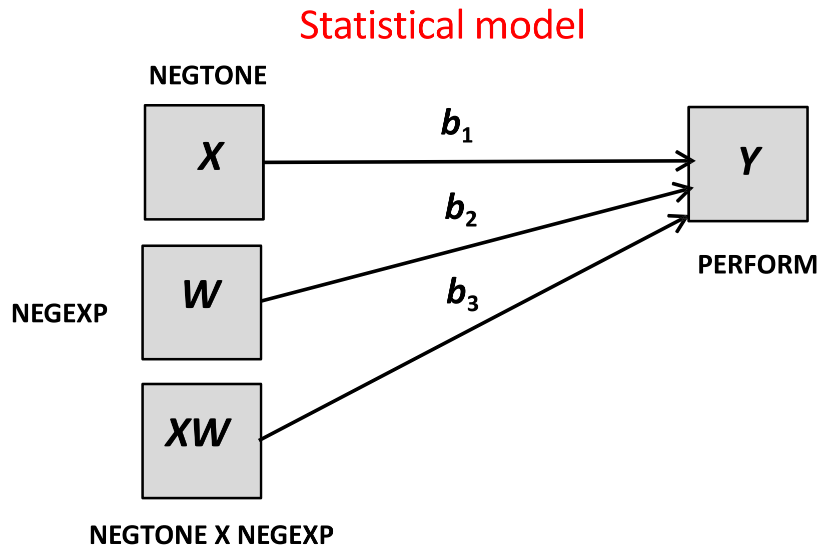
Example



Does the relationship between negativity of the team work climate and performance of the team depend on how openly members of the team express their feelings?

Example





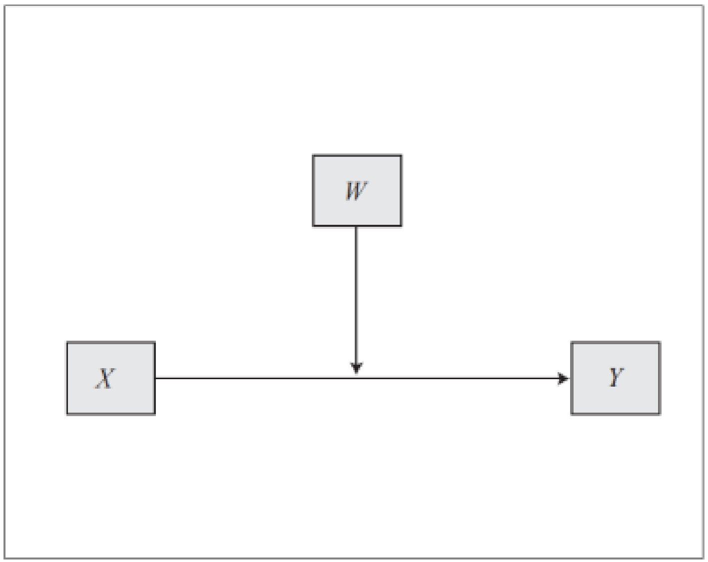
$$\hat{Y}_i = b_0 + (b_1 + b_3 W_i)X_i + b_2 W_i$$
$$= b_0 + b_1 X_i + b_2 W_i + b_3 X_i W_i$$

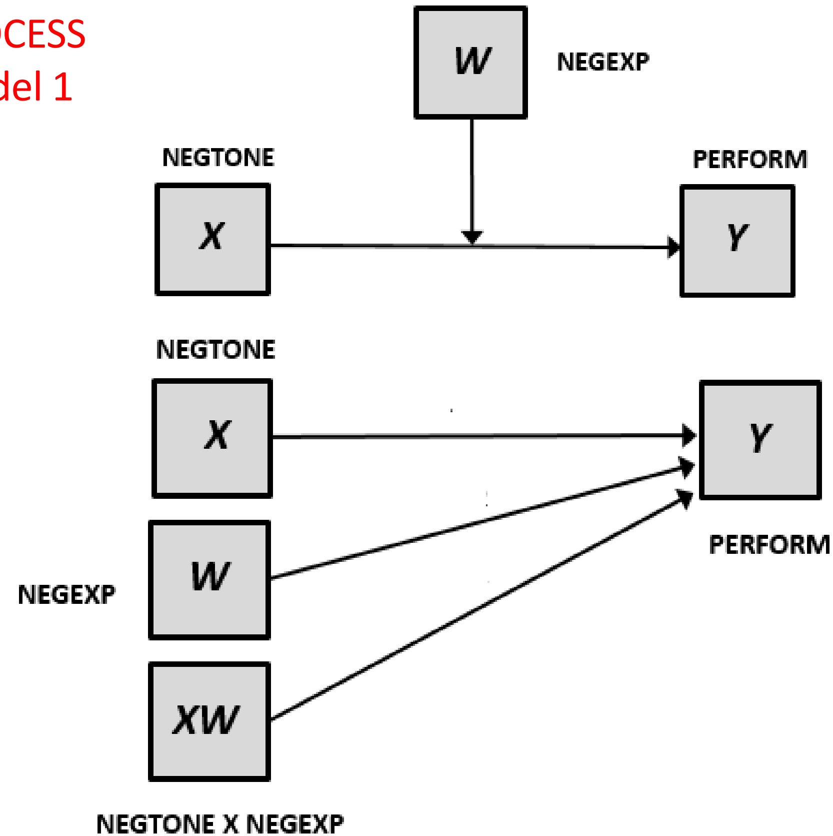
- Centering of *X* and *W* is not required
- Standardization of *X* and *W* is not required
- Building model in steps (first *X* and *W*, then *XW*) is not necessary.
- Doing any of these has no effect on the test of interaction or the conditional effects
- The centering, standardization, and hierarchical entry myths are widespread. I debunked these in the first course. Also see IMMCPA, 2nd edition.

Estimation Using PROCESS

PROCESS
Model 1

Model 1






```
process y=perform/x=negtone/w=negexp/plot=1/jn=1/model=1.
```

```
%process (data=teams,y=perform,x=negtone,w=negexp,plot=1,jn=1,model=1);
```

```
process(data=teams,y="perform",x="negtone",w="negexp",plot=1,jn=1,model=1)
```

PROCESS output

```
Model  : 1
    Y  : perform
    X  : negtone
    W  : negexp

Sample
Size:  60
```

$$\hat{Y}_i = b_0 + b_1 X_i + b_2 W_i + b_3 X_i W_i$$

$$\hat{Y}_i = -0.0030 - 0.3105X_i - 0.0135W_i - 0.6027X_i W_i$$

```
**************************************************************************
OUTCOME VARIABLE:
 perform

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .5089      .2590      .2131     6.5237     3.0000    56.0000      .0007

Model
              coeff         se          t          p       LLCI       ULCI
constant     -.0030      .0600     -.0507      .9598     -.1233      .1172
negtone      -.3105      .1188    -2.6137      .0115     -.5485     -.0725
negexp       -.0135      .1207     -.1120      .9112     -.2554      .2283
Int_1        -.6027      .2441    -2.4696      .0166    -1.0916     -.1138

Product terms key:
 Int_1    :        negtone  x        negexp

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*W      .0807     6.0992     1.0000    56.0000      .0166
```

PROCESS generates the product for you.

Improvement in fit when the product is added to the model

b_1 = -0.311

b_3 = -0.603

Generating a visual depiction using the plot option

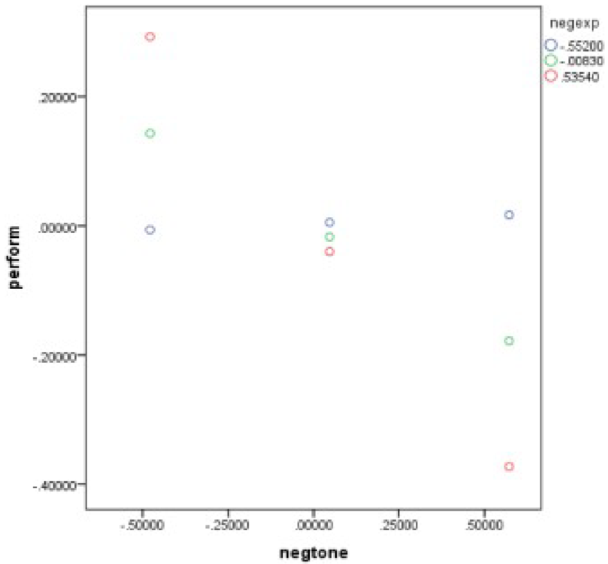
```
DATA LIST FREE/
   negtone    negexp     perform     .
  BEGIN DATA.
        -.4500    -.5308       -.0001
        -.0350    -.5308        .0038
         .5224    -.5308        .0091
        -.4500    -.0600        .1212
        -.0350    -.0600        .0074
         .5224    -.0600       -.1456
        -.4500     .6000        .2913
        -.0350     .6000        .0124
         .5224     .6000       -.3623
END DATA.
GRAPH/SCATTERPLOT=
 negtone  WITH     perform  BY       negexp   .
```

PROCESS for SPSS writes the code for you. Cut and paste as syntax into SPSS and run to see the interaction.

```
data;
input negtone negexp perform;
datalines;
     -.4500      -.5308      -.001
     -.0350      -.5308      .0038
      .5224      -.5308      .0091
     -.4500      -.0600      .1212
     -.0350      -.0600      .0074
      .5224      -.0600     -.1456
     -.4500       .6000      .2913
     -.0350       .6000      .0124
      .5224       .6000     -.3623
run;
proc
     sgplot;reg  x=negtone   y=perform/group=negexp;run;
```

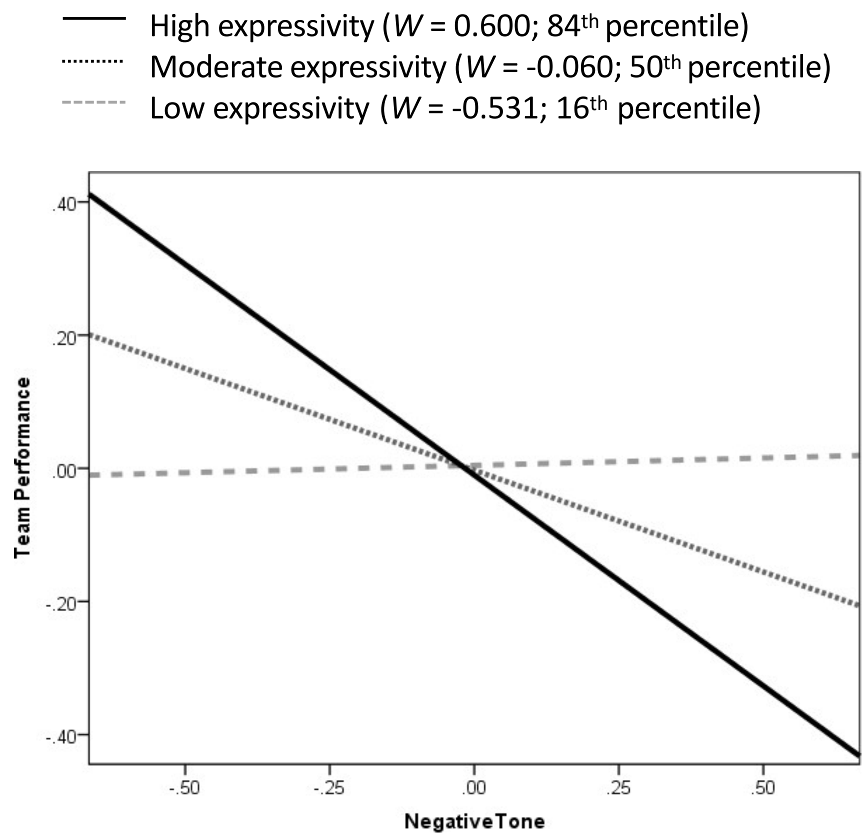
Process for SAS only gives you this. The rest of the code you have to enter yourself.

Generating a visual depiction using the plot option



Before editing

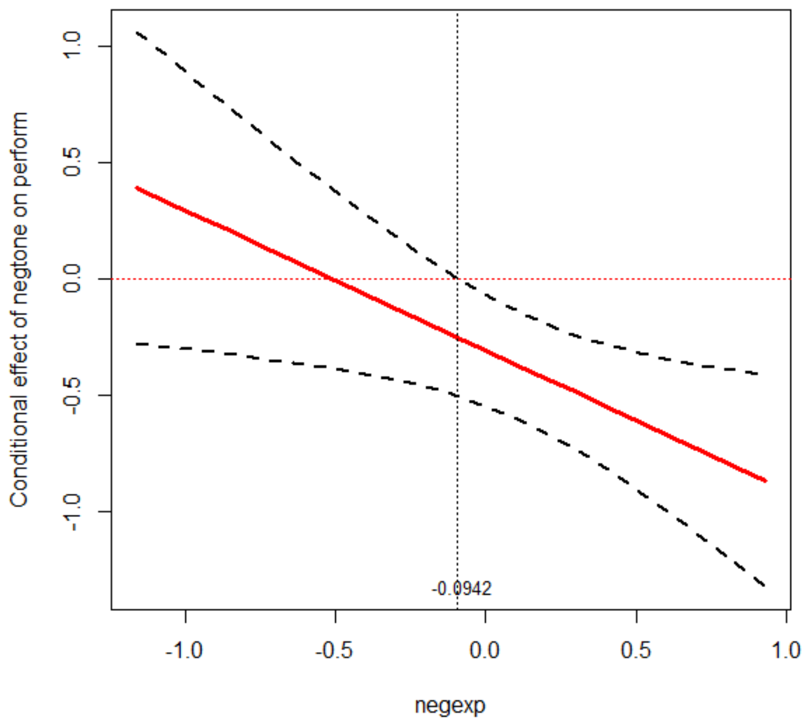
After editing

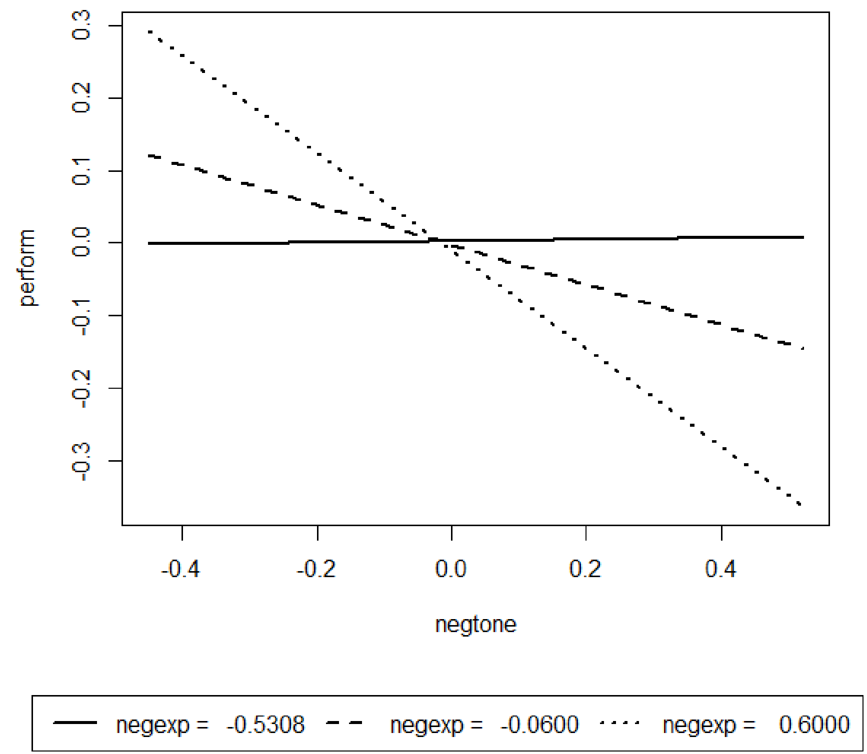


Substantive interpretation: A more negative tone of the work climate reduces performance moreso among teams who are more expressive in their emotions.

Plots in R

PROCESS will generate plots by default in R, but these cannot be editted





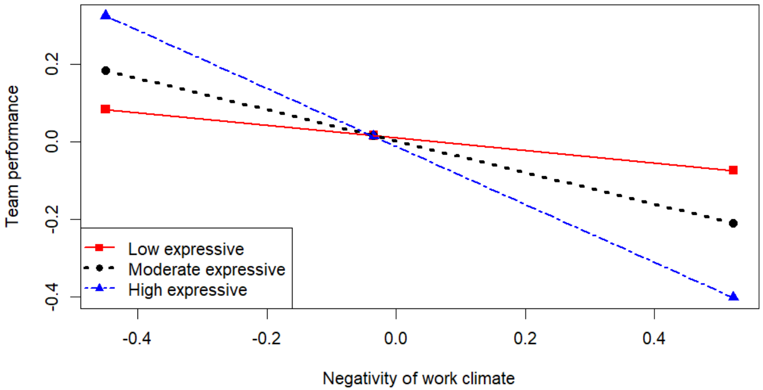
Creating Plots with Save Option (R)

You can also use the save = 2 option to save the model output as a matrix object

```
modsave <- process(data=teams,y="perform",x="negtone",w="negexp",plot=1,jn=1,
                   model=1, save = 2)
```

```
> modsave[34:42, 1:3]
        [,1]    [,2]          [,3]
 [1,] -0.4500 -0.5308 -0.0001016393
 [2,] -0.0350 -0.5308  0.0038076810
 [3,]  0.5224 -0.5308  0.0090584163
 [4,] -0.4500 -0.0600  0.1212258404
 [5,] -0.0350 -0.0600  0.0073712349
 [6,]  0.5224 -0.0600 -0.1455505896
 [7,] -0.4500  0.6000  0.2913110923
 [8,] -0.0350  0.6000  0.0123668711
 [9,]  0.5224  0.6000 -0.3622921866
```

Location of plot data will differ based on process code and other requested options




```
x <- modsave[34:42,1]
w <- modsave[34:42, 2]
y <- modsave[34:42, 3]
wmarker<-c(15,15,15,16,16,16,17,17,17)
plot(y=y,x=x,cex=1.2,pch=wmarker,  xlab="Negativity of
work climate", ylab="Team performance",
col=c("red","red","red","black","black","black","blue",
"blue","blue"))
legend.txt<-c("Low expressive","Moderate expressive",
"High expressive")
legend("bottomleft", legend = legend.txt,cex=1,
lty=c(1,3,6),lwd=c(2,3,2),pch=c(15,16,17),
col=c("red","black","blue"))
lines(x[w==-.5308],y[w==-.5308],lwd=2,col="red")
lines(x[w==-.060],y[w==-.060],lwd=3,lty=3,col="black")
lines(x[w==.600],y[w==.600],lwd=2,lty=6,col="blue")
```

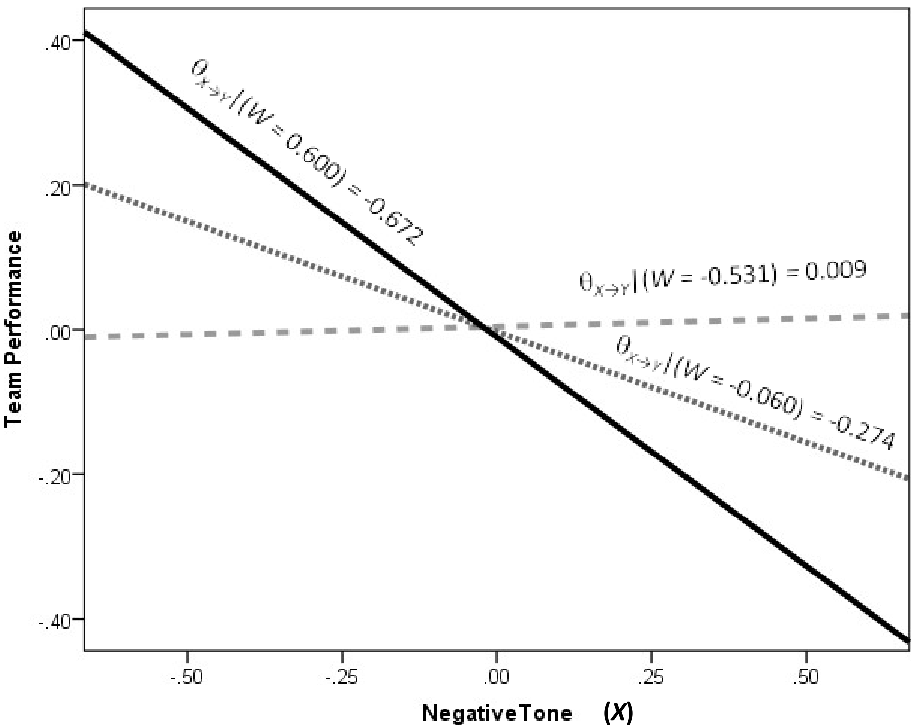
The conditional effect of *X* on *Y*

$$\widehat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3X_iW_i$$

—— High expressivity (W = 0.600)
········ Moderate expressivity (W = -0.060)
- - - - Low expressivity (W = -0.531)

$$\widehat{Y}_i = -0.003 - 0.311X_i - 0.014W_i - 0.603X_iW_i$$

$$\widehat{Y}_i = -0.003 + (-0.311 - 0.603W_i)X_i - 0.014W_i$$



$$\widehat{Y}_i = -0.003 + \theta_{X\to Y}X_i - 0.014W_i$$

where

$$\boldsymbol{\theta_{X\to Y} = -0.311 - 0.603W}$$

This is the equation for the slope of the line relating *X* to *Y*.

	W	$\theta_{X\to Y}$
16th	-0.531	0.009
50th	-0.060	-0.274
84th	0.600	-0.672

Probing Interactions

- With evidence of moderation (typically, as revealed by a statistically significant coefficient for a relevant product term), "conditional effects", a.k.a. "simple slopes", are typically estimated and interpreted. Sometimes called "spotlight" analysis
    - Dichotomous moderator: Estimate effect of focal predictor on *Y* at the two values of the moderator
    - Quantitative/continuous moderator: Pick-a-point approach—Estimate the effect of focal predictor for "representative values" of moderator
- Johnson-Neyman technique avoids the arbitrariness of the pick-a-point approach. Sometimes called a "floodlight" analysis.
    - Mathematical derivation of points of transition on the moderator continuum between a statistically significant and nonsignificant effect of the focal predictor: "Regions of significance"

PROCESS output: Probing the interaction

PROCESS sees that the moderator is quantitative (because it has more than 2 values) and implements the pick-a-point procedure with moderator values equal to the 16th, 50th, and 84th percentiles of the distribution of the moderator.

```
Focal predict: negtone (X)
      Mod var: negexp   (W)

Conditional effects of the focal predictor at values of the moderator(s):
```

W

	negexp	Effect	se	t	p	LLCI	ULCI
16th	-.5308	.0094	.1970	.0478	.9620	-.3853	.4041
50th	-.0600	-.2743	.1234	-2.2235	.0302	-.5215	-.0272
84th	.6000	-.6722	.1631	-4.1210	.0001	-.9989	-.3454

$$\theta_{X \to Y} = b_1 + b_3 W = -0.311 - 0.603W$$

Negative affective tone of the work climate is significantly and negatively related to performance among moderately ($\theta_{X \to Y | W=-0.060}$ = -0.274, $p < .05$) and highly expressive teams ($\theta_{X \to Y | W=0.600}$ = -0.672, $p < .001$) but not significantly related to performance among less expressive teams ($\theta_{X \to Y | W=-0.531}$ = 0.009, $p = .962$).

Additional probing options

Using the **moments** option, PROCESS will probe at values of the moderator corresponding to the mean, a standard deviation below the mean, and a standard deviation above the mean. Or use the **wmodval** option to request a specific value or values of the moderator at which you'd like the conditional effect(s)

```
process y=perform/x=negtone/w=negexp/moments=1/model=1.
```

```
%process (data=teams,y=perform,x=negtone,w=negexp,moments=1,model=1);
```

```
process(data=teams,y="perform",x="negtone",w="negexp",moments=1,model=1)
```

```
Conditional effects of the focal predictor at values of the moderator(s):

     negexp     Effect         se          t          p       LLCI       ULCI
     -.5520      .0222      .2013      .1104      .9125     -.3809      .4254
     -.0083     -.3055      .1193    -2.5598      .0132     -.5446     -.0664
      .5354     -.6332      .1523    -4.1574      .0001     -.9383     -.3281
```

```
process y=perform/x=negtone/w=negexp/model=1/wmodval=-0.5,0.5.
```

```
%process (data=teams,y=perform,x=negtone,w=negexp,model=1,wmodval=-0.5 .50);
```

```
process(data=teams,y="perform‟,x="negtone",w="negexp",model=1,wmodval=c(-0.5,.50))
```

```
Conditional effects of the focal predictor at values of the moderator(s):

     negexp     Effect         se          t          p       LLCI       ULCI
     -.5000     -.0091      .1910     -.0479      .9620     -.3917      .3734
      .5000     -.6119      .1468    -4.1687      .0001     -.9059     -.3178
```

Johnson-Neyman Technique

The JN option finds regions of significance, if they exist, using the Johnson-Neyman method. This method is implemented in PROCESS using the **jn=1** option.

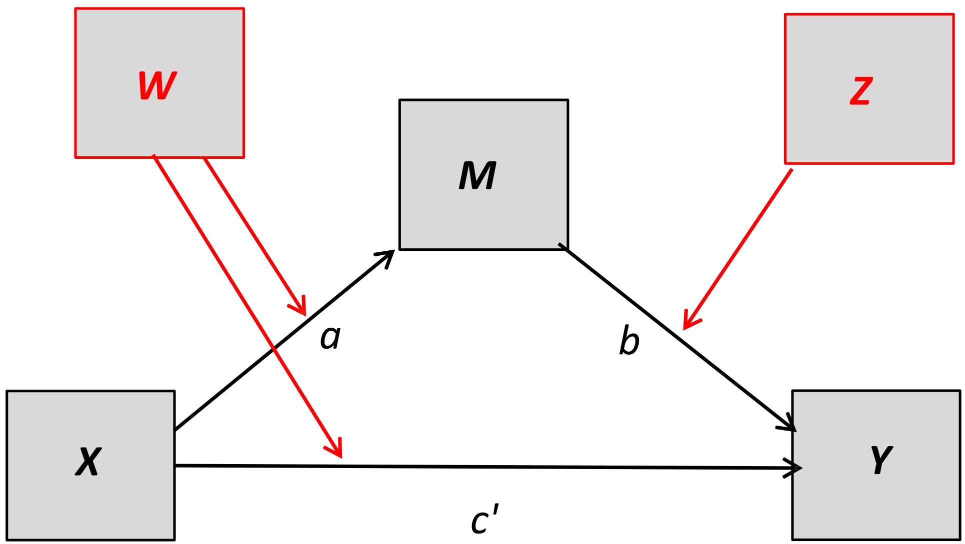
```
Moderator value(s) defining Johnson-Neyman significance region(s):
      Value      % below      % above
     -.0941      45.0000      55.0000
```

$\theta_{X \to Y} = b_1 + b_3 W = -0.311 - 0.603W$

Conditional effect of focal predictor at values of the moderator:

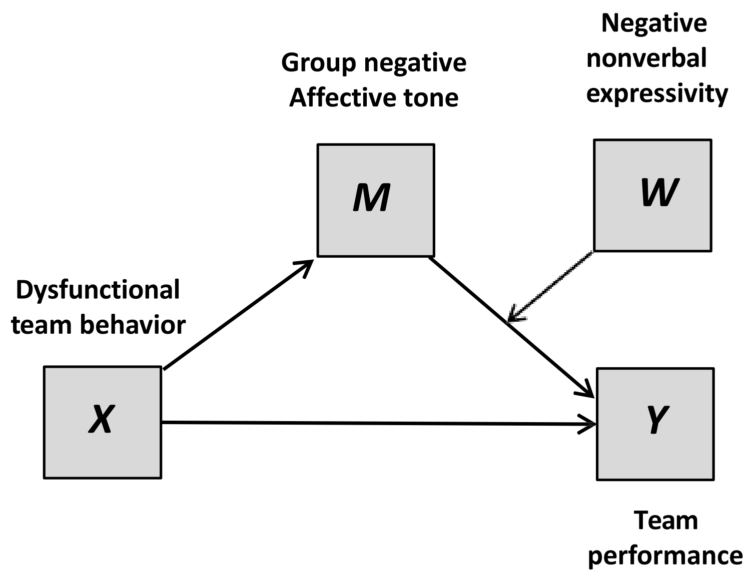
negexp	Effect	se	t	p	LLCI	ULCI
-1.1600	.3887	.3340	1.1636	.2495	-.2805	1.0578
-1.0555	.3257	.3102	1.0499	.2983	-.2957	.9471
-.9510	.2627	.2867	.9164	.3634	-.3116	.8369
-.8465	.1997	.2635	.7579	.4517	-.3281	.7275
-.7420	.1367	.2408	.5678	.5724	-.3456	.6191
-.6375	.0737	.2187	.3371	.7373	-.3644	.5119
-.5330	.0107	.1975	.0544	.9568	-.3848	.4063
-.4285	-.0522	.1773	-.2946	.7694	-.4075	.3030
-.3240	-.1152	.1588	-.7258	.4710	-.4332	.2028
-.2195	-.1782	.1423	-1.2521	.2158	-.4633	.1069
-.1150	-.2412	.1289	-1.8711	.0666	-.4994	.0170
-.0941	-.2538	.1267	-2.0033	.0500	-.5075	.0000
-.0105	-.3042	.1195	-2.5457	.0137	-.5435	-.0648
.0940	-.3672	.1151	-3.1913	.0023	-.5977	-.1367
.1985	-.4302	.1162	-3.7023	.0005	-.6629	-.1974
.3030	-.4931	.1227	-4.0183	.0002	-.7390	-.2473
.4075	-.5561	.1339	-4.1539	.0001	-.8243	-.2879
.5120	-.6191	.1486	-4.1657	.0001	-.9168	-.3214
.6165	-.6821	.1660	-4.1093	.0001	-1.0146	-.3496
.7210	-.7451	.1853	-4.0220	.0002	-1.1162	-.3740
.8255	-.8081	.2059	-3.9250	.0002	-1.2205	-.3956
.9300	-.8711	.2275	-3.8289	.0003	-1.3268	-.4153

Conditional Process Analysis



- The indirect effect of *X* on *Y* through *M* is estimated as the product of the *a* and *b* paths
- But what if size of *a* or *b* (or both) depends on another variable (i.e., is moderated)?
- If so, then the magnitude of the indirect effect therefore depends on the moderator, meaning that "mediation is moderated".
- When *a* or *b* is moderated, it is sensible then to estimate "conditional indirect effects"—values of indirect effect conditioned on values of the moderator variable that moderates *a* and/or *b*.
- Direct effects can also be conditional. For instance, in the above, *W* moderates *X*'s direct effect on *Y*.

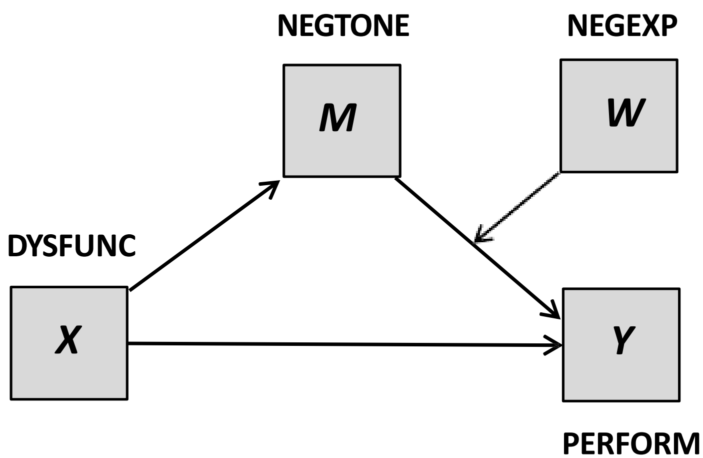
Conditional process analysis: Example



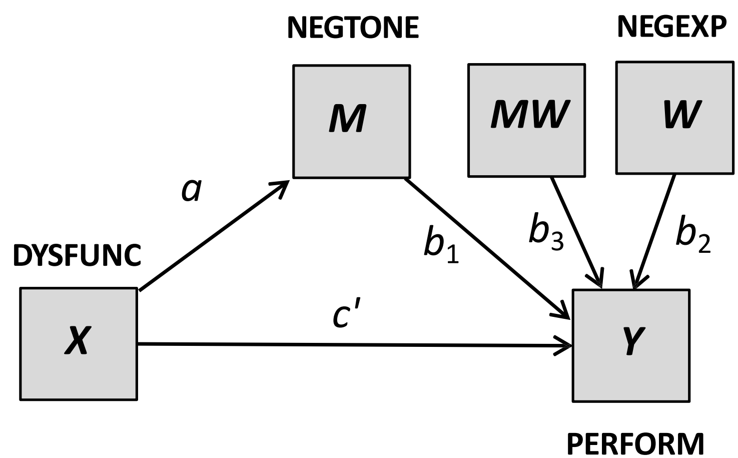
This is a model of negative affective tone as the mechanism by which dysfunctional team behavior influences performance, with that mechanism being contingent on the extent to which team members hide their negative feelings from the team. This "nonverbal expressivity" is postulated as moderating the effect of negative tone on performance. This is a "second stage" moderated mediation model because the moderation of the mechanism operates in $M \rightarrow Y$ stage of the process.

Conceptual and statistical models

Conceptual Model



Statistical Model



$$\widehat{M_i} = a_0 + aX_i$$

$$\widehat{Y_i} = c'_0 + c'X_i + b_1 M_i + b_2 W_i + b_3 M_i W_i$$

or, equivalently,

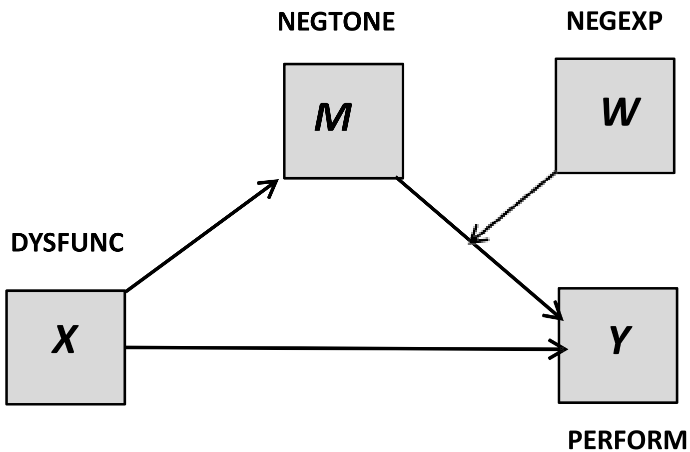
$$\widehat{Y_i} = c'_0 + c'X_i + (b_1 + b_3 W_i) M_i + b_2 W_i$$

or, equivalently,

$$\widehat{Y_i} = c'_0 + c'X_i + \theta_{M \to Y} M_i + b_2 W_i \quad \text{where} \quad \theta_{M \to Y} = b_1 + b_3 W$$

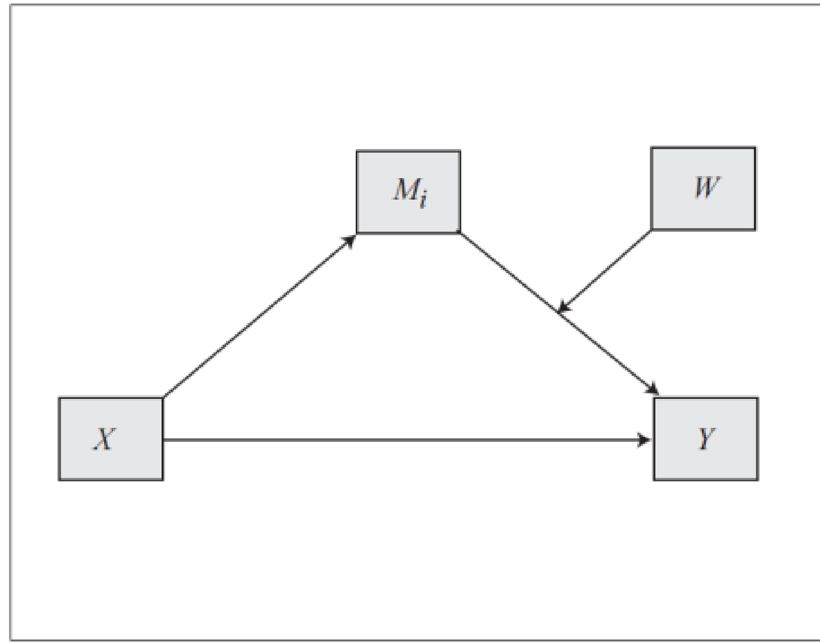
Estimation of the model in PROCESS

PROCESS takes all of the computational burden off your shoulders.



This is PROCESS model 14

Model 14



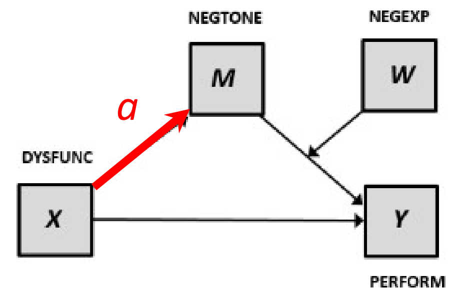

```
process x=dysfunc/m=negtone/y=perform/w=negexp/plot=1/model=14/seed=61326.
```

```
%process (data=teams,x=dysfunc,m=negtone,y=perform,w=negexp,plot=1,model=14,
   seed=61326);
```

```
process (data=teams,x="dysfunc",m="negtone",y="perform",w="negexp",plot=1,
   model=14,seed=61326)
```

PROCESS output

```
***************** PROCESS Procedure for SPSS Release 3.50 ****************

          Written by Andrew F. Hayes, Ph.D.       www.afhayes.com
    Documentation available in Hayes (2018). www.guilford.com/p/hayes3

**************************************************************************
Model : 14
    Y   : perform
    X   : dysfunc
    M  : negtone
    W  : negexp

Sample
Size:  60

Custom
Seed:  61326
```



$$\widehat{M_i} = 0.026 + 0.620X_i$$

```
**************************************************************************
OUTCOME VARIABLE:
 negtone

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .4384      .1922      .2268    13.7999     1.0000    58.0000      .0005
```

Model

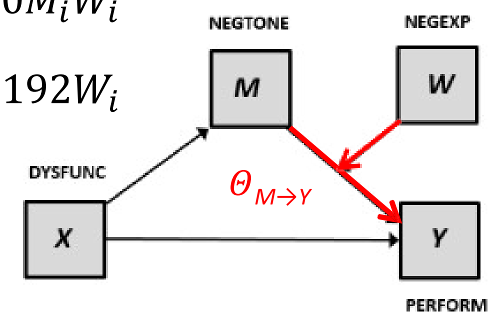
a = 0.620

	coeff	se	t	p	LLCI	ULCI
constant	.0257	.0618	.4159	.6791	-.0979	.1493
dysfunc	.6198	.1668	3.7148	.0005	.2858	.9537

PROCESS output

$$\widehat{Y}_i = -0.119 + 0.3661X_i - 0.4357M_i - 0.0192W_i - 0.5170M_iW_i$$

$$= -0.119 + 0.3661X_i + (-0.4357 - 0.5170W_i)M_i - 0.0192W_i$$



Model Summary

R	R-sq	MSE	F	df1	df2	p
.5586	.3120	.2015	6.2350	4.0000	55.0000	.0003

Model

	coeff	se	t	p	LLCI	ULCI
constant	-.0119	.0585	-.2029	.8399	-.1292	.1054
dysfunc	.3661	.1778	2.0585	.0443	.0097	.7224
negtone	-.4357	.1306	-3.3377	.0015	-.6974	-.1741
negexp	-.0192	.1174	-.1634	.8708	-.2545	.2161
Int_1	-.5170	.2409	-2.1458	.0363	-.9998	-.0341

b_1 = -0.436

b_3 = -0.517

Product terms key:

Int_1 : negtone x negexp

Test(s) of highest order unconditional interaction(s):

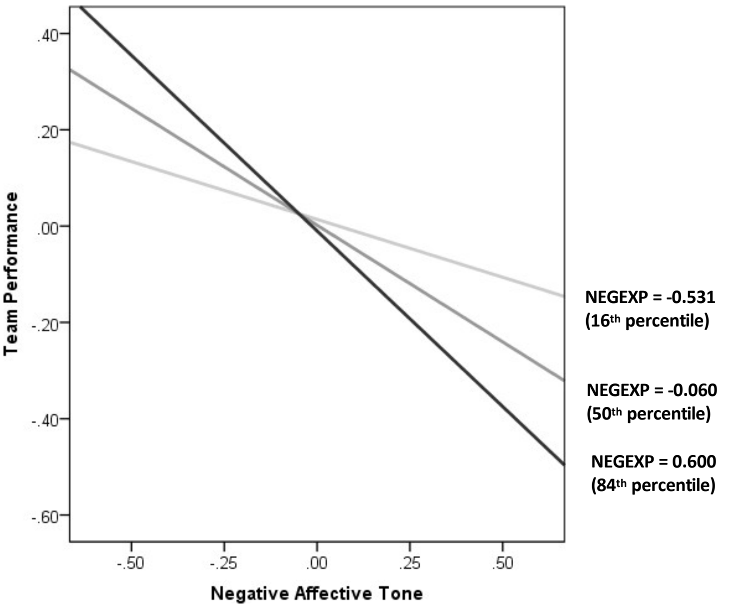
	R2-chng	F	df1	df2	p
M*W	.0576	4.6043	1.0000	55.0000	.0363

$$\theta_{M \to Y} = b_1 + b_3W = -0.436 - 0.517W$$

PROCESS generates what you need to visualize this

```
DATA LIST FREE/
  negtone       negexp        perform   .
BEGIN DATA.
       -.4500      -.5308        .0836
       -.0350      -.5308        .0166
        .5224      -.5308       -.0733
       -.4500      -.0600        .1841
       -.0350      -.0600        .0161
        .5224      -.0600       -.2095
       -.4500       .6000        .3250
       -.0350       .6000        .0154
        .5224       .6000       -.4004

END DATA.
GRAPH/SCATTERPLOT=
 negtone  WITH      perform  BY         negexp    .
```

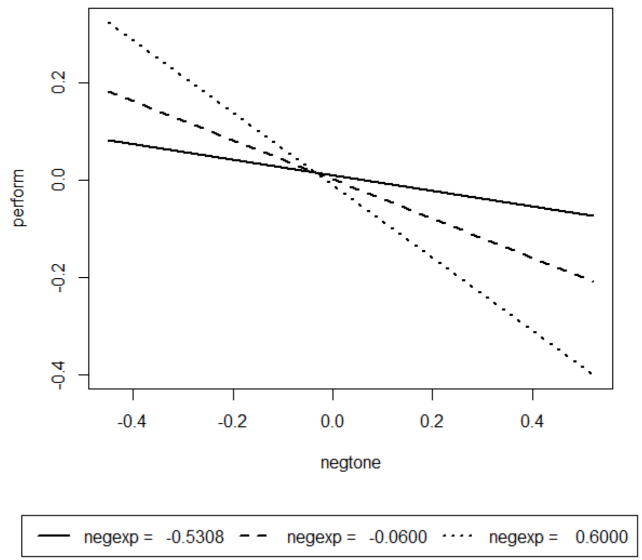
```
data;
input negtone negexp perform;
cards;
     -.4500      -.5308       .0836
     -.0350      -.5308       .0166
      .5224      -.5308      -.0733
     -.4500      -.0600       .1841
     -.0350      -.0600       .0161
      .5224      -.0600      -.2095
     -.4500       .6000       .3250
     -.0350       .6000       .0154
      .5224       .6000      -.4004

run;
proc sgplot;reg x=negtone y=perform/group=negexp;run;
```



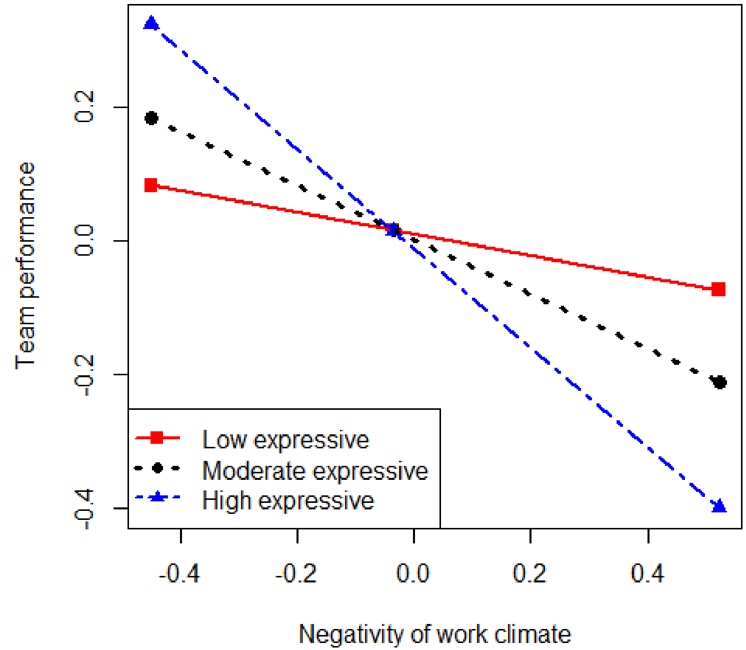
After some editing in SPSS

Generating a picture using R

From PROCESS output

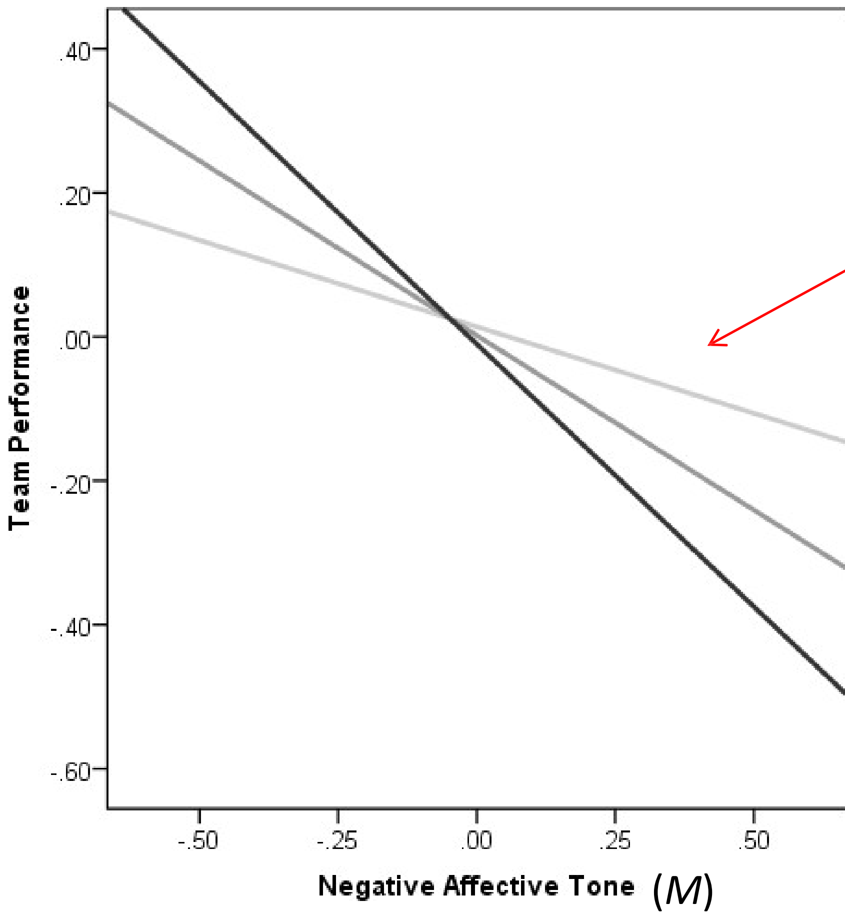


Code Below




```
x <- modsave[14:22,1]
w <- modsave[14:22, 2]
y <- modsave[14:22, 3]
wmarker<-c(15,15,15,16,16,16,17,17,17)
plot(y=y,x=x,cex=1.2,pch=wmarker,  xlab="Negativity of work climate", ylab="Team performance",
col=c("red","red","red","black","black","black","blue","blue","blue"))
legend.txt<-c("Low expressive","Moderate expressive", "High expressive")
legend("bottomleft", legend = legend.txt,cex=1,  lty=c(1,3,6),lwd=c(2,3,2),pch=c(15,16,17),
col=c("red","black","blue"))
lines(x[w==-.5308],y[w==-.5308],lwd=2,col="red")
lines(x[w==-.060],y[w==-.060],lwd=3,lty=3,col="black")
lines(x[w==.600],y[w==.600],lwd=2,lty=6,col="blue")
```

The conditional effect of *M* on *Y*



NEGEXP = -0.531 (16th percentile)

NEGEXP = -0.060 (median)

NEGEXP = 0.600 (84th percentile)

$\theta_{M \to Y} = b_1 + b_3 W =$
$= -0.436 - 0.517W$
= slopes of these lines

$$\theta_{M \to Y} = -0.436 - 0.517W$$

W	$\boldsymbol{\theta_{M \to Y}}$
-0.531	-0.161
-0.060	-0.405
0.600	-0.746

The conditional effect of *M* on *Y*

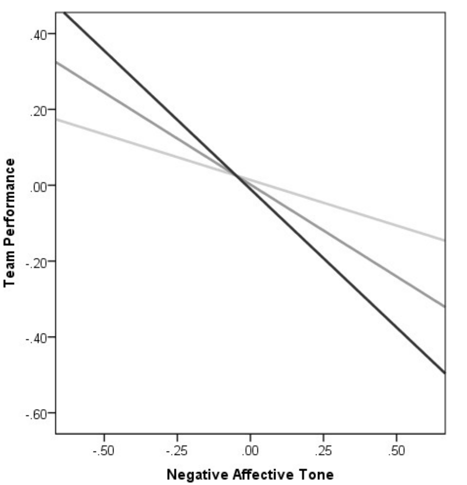
PROCESS provides output for examining $\theta_{M\rightarrow Y}$. It sees that the moderator is quantitative (i.e., it has more than 2 values) so it generates $\theta_{M\rightarrow Y}$ at values of *W* equal to the 16th, 50th, and 84th percentiles of the distribution.

$$\theta_{M\rightarrow Y} = b_1 + b_3 W = -0.436 - 0.517W$$

```
Focal predict: negtone (M)
      Mod var: negexp   (W)

Conditional effects of the focal predictor at values of the moderator(s):
```

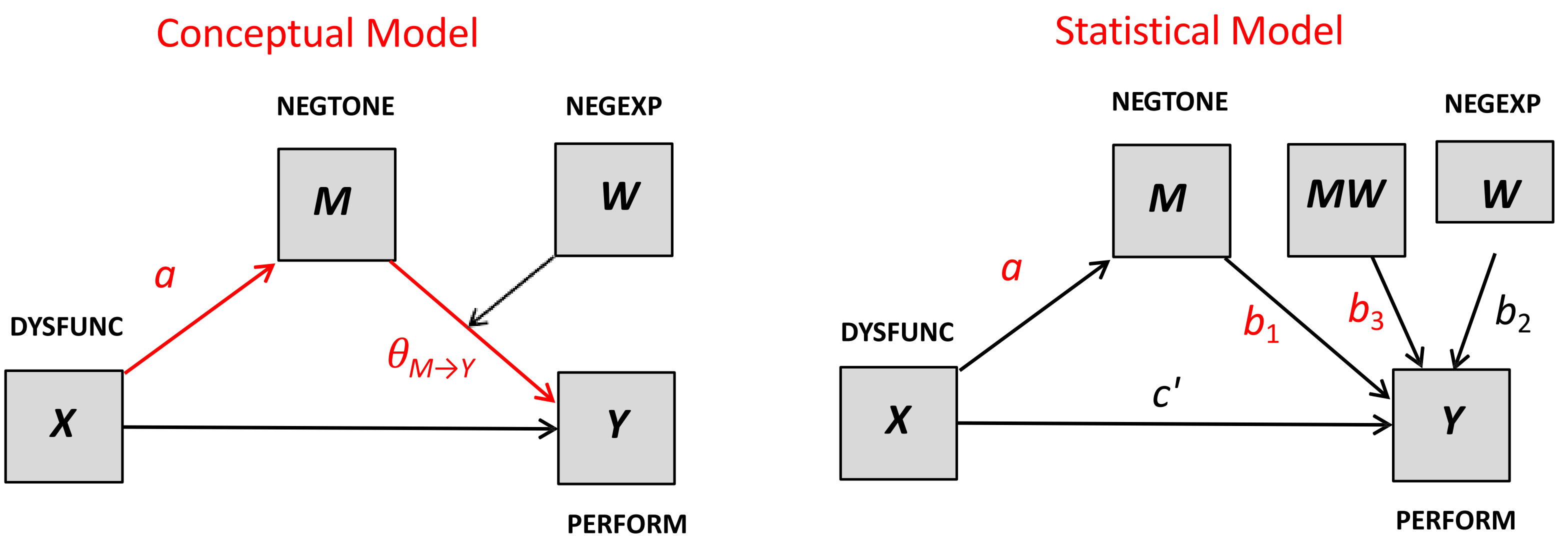
	W negexp	Effect	se	t	p	LLCI	ULCI
16th	-.5308	-.1613	.2088	-.7729	.4429	-.5797	.2570
50th	-.0600	-.4047	.1357	-2.9834	.0042	-.6766	-.1329
84th	.6000	-.7459	.1626	-4.5879	.0000	-1.0718	-.4201

These are the slopes of these lines



Use the **moments** or **wmodval** options if you'd rather condition the effect of M at different values of W.

The conditional indirect effect of *X*

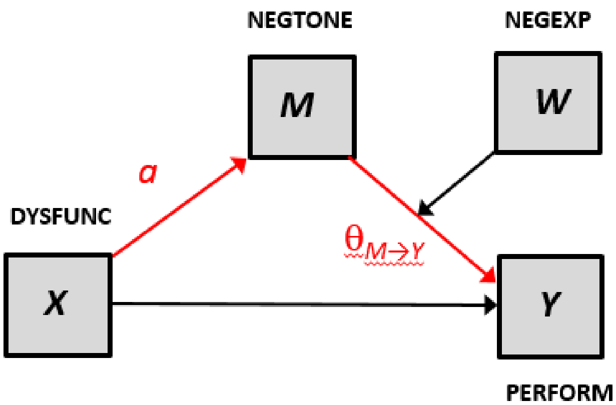


The conditional indirect effect of *X* on *Y* through *M* is the product of the effect of *X* on *M* (*a*) and the conditional effect of *M* on *Y* given *W* ($\theta_{M\rightarrow Y} = b_1 + b_3W$):

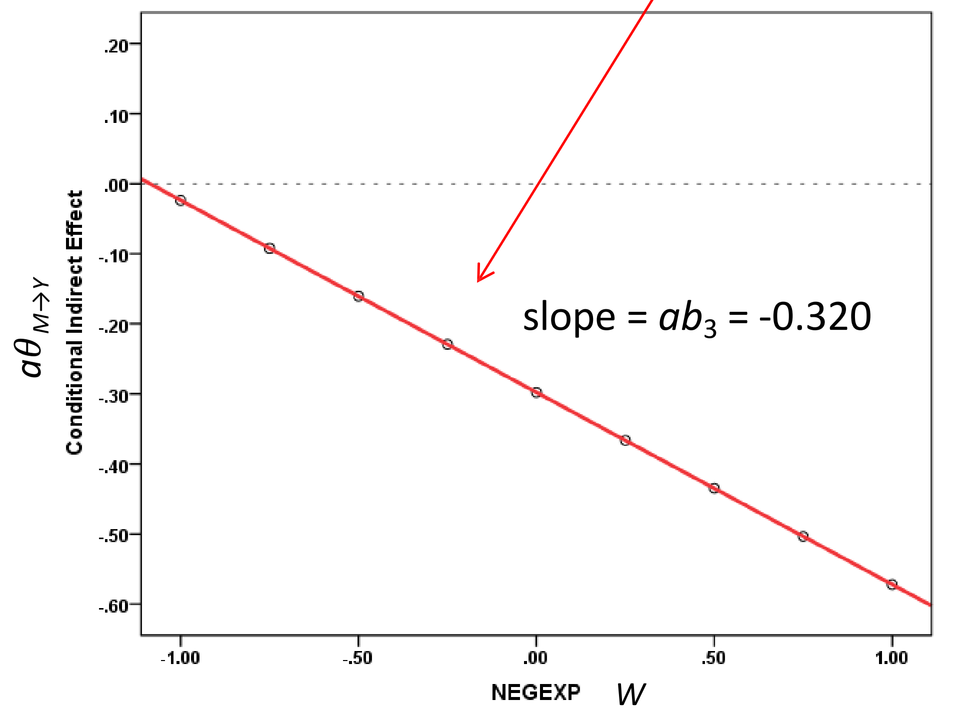
$$a\theta_{M\rightarrow Y} = a(b_1 + b_3W) = 0.620(-0.436 - 0.517W)$$

The indirect effect of dysfunctional team behavior on team performance through negative tone is a function of negative nonverbal expressivity.

A visual representation and test of moderated mediation



$$a\theta_{M\rightarrow Y} = a\,(b_1 + b_3W) = 0.620(-0.436 - 0.517W)$$
$$= ab_1 + ab_3W = -0.270 - 0.320W$$

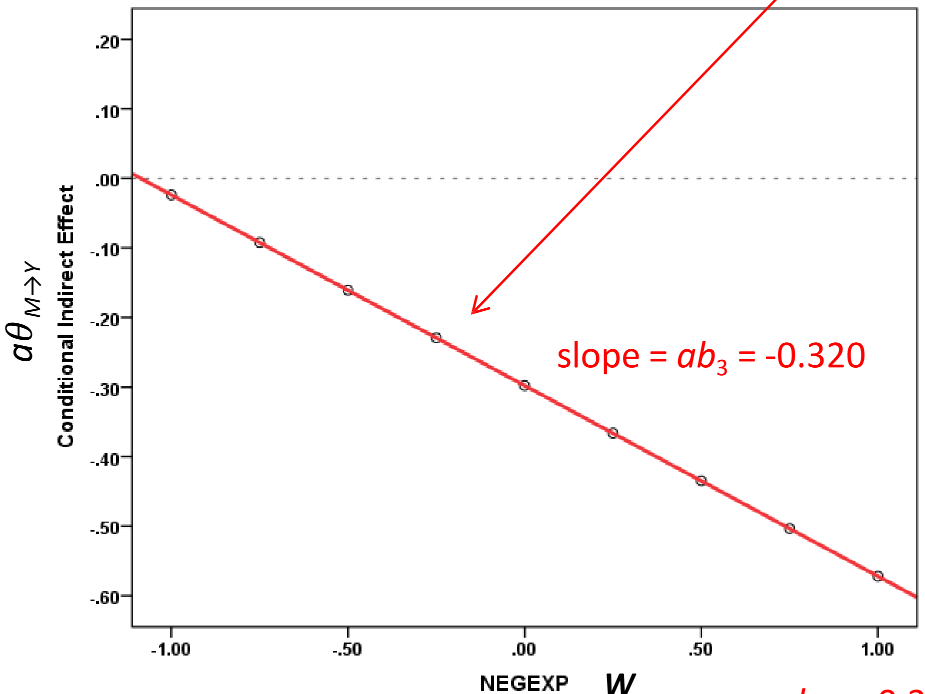


The indirect effect is increasingly negative as negative nonverbal expressivity increases. The "**index of moderated mediation**" is ab_3 = -0.320. It quantifies the relationship between the moderator and the indirect effect in this model. **An inference about the slope of this is an inference about whether the indirect effect is moderated:**

As ab_3 is a product of regression coefficients, an inferential test should respect the nonnormality of the sampling distribution of the product. A bootstrap confidence interval is a good choice.

This test is provided automatically by PROCESS

$$a\theta_{M\to Y} = a(b_1 + b_3W) = 0.620(-0.436 - 0.517W)$$
$$= ab_1 + ab_3W = -0.270 - 0.320W$$



The indirect effect is a function of W (negative nonverbal expressivity) in our model. This function is a line.

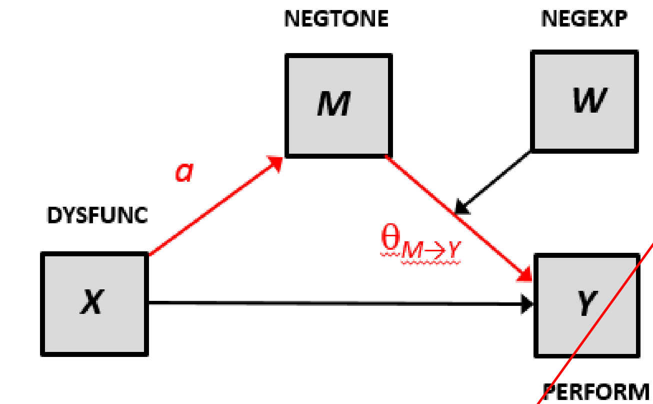
$$a\theta_{M\to Y} = a(b_1 + b_3W)$$
$$= ab_1 + ab_3W$$
$$= -0.270 - 0.320W$$

ab_3 = -0.320, 95% bootstrap CI = -0.775 to -0.042

```
Index of moderated mediation:
              Index      BootSE    BootLLCI    BootULCI
negexp        -.3204      .1928     -.7725      -.0415
```

This slope is statistically different from zero. The indirect effect depends on negative nonverbal expressivity…. The mediation is moderated.

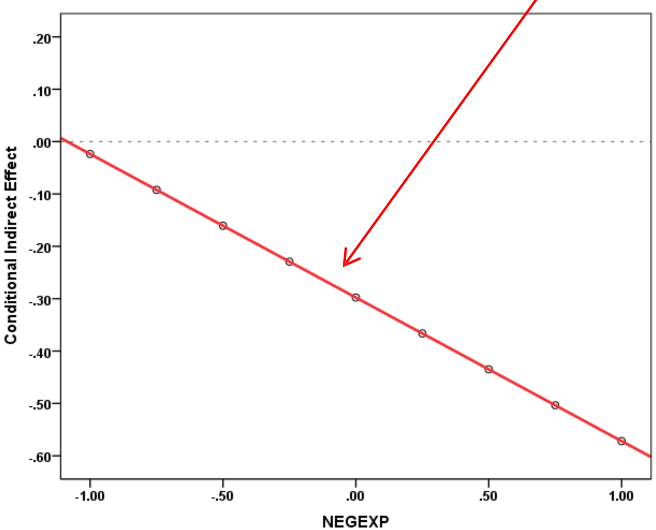
With evidence of moderation of the indirect effect, we can probe



$$a\theta_{M \to Y} = a\,(b_1 + b_3W) = 0.620(-0.436 - 0.517W)$$
$$= ab_1 + ab_3W = -0.270 - 0.320W$$

The indirect effect is increasingly negative as negative nonverbal expressivity increases.

With evidence of moderation of the indirect effect, we can now probe this moderation of mediation through an analogue of the pick-a-point approach used in moderation analysis.



	NEGEXP (*W*)	*a*	$\theta_{M \to Y}$	$a\theta_{M \to Y}$
16th	-0.531	0.620	-0.161	-0.100
50th	-0.060	0.620	-0.405	-0.251
84th	0.600	0.620	-0.746	-0.462

Conditional indirect effects

We need an inferential test for these conditional indirect effects. Bootstrap confidence intervals are perfect for the job. PROCESS does this for us automatically.

With evidence of moderation of the indirect effect, now we probe

Conditional indirect effect of trauma on depression through PTSD

PROCESS sees that the moderator is continuous so prints the conditional indirect effect at the 16th, 50th, and 84th percentiles of the distribution of the moderator and bootstrap confidence intervals for inference.

```
****************** DIRECT AND INDIRECT EFFECTS OF X ON Y *****************

Conditional indirect effects of X on Y:

INDIRECT EFFECT:
 dysfunc     ->     negtone     ->     perform
```

$a\theta_{M \to Y} = a\,(b_1 + b_3W) = 0.620(-0.436 - 0.517W)$
$= ab_1 + ab_3W = -0.270 - 0.320W$

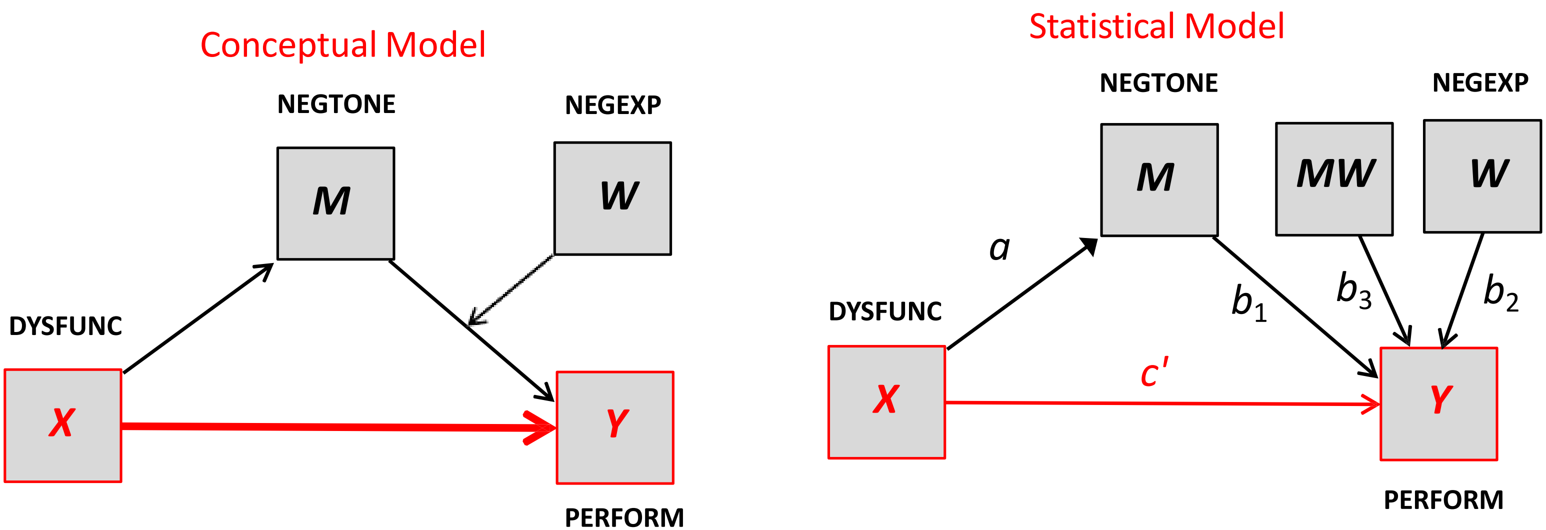
	negexp	Effect	BootSE	BootLLCI	BootULCI
16th	-.5308	-.1000	.1513	-.3708	.2564
50th	-.0600	-.2508	.1189	-.4988	-.0357
84th	.6000	-.4623	.1720	-.8144	-.1443

Conditional indirect effects with 95% bootstrap CIs based on 5,000 bootstrap samples.

The indirect effect of dysfunctional behavior on performance through negative tone is negative among teams relatively moderate (point estimate: -0.251, 95% CI from -0.499 to -0.036) and relatively high (point estimate: -0.462, 95% CI from -0.814 to -0.144) in expressivity but not different from zero among those low in expressivity (point estimate: -0.100, 95% CI from -0.371 to 0.256).

The direct effect of *X*

The direct effect of *X* is the effect of *X* of *Y* that does not operate through *M.*



$$\hat{Y}_i = c'_0 + \boxed{c'}X_i + b_1 M_i + b_2 W_i + b_3 M_i W_i$$

```
****************** DIRECT AND INDIRECT EFFECTS OF X ON Y *****************

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .3661      .1778     2.0585      .0443      .0097      .7224
```

Holding negative tone of the work climate and expressivity constant, teams that engage in more dysfunctional behavior perform *better.*

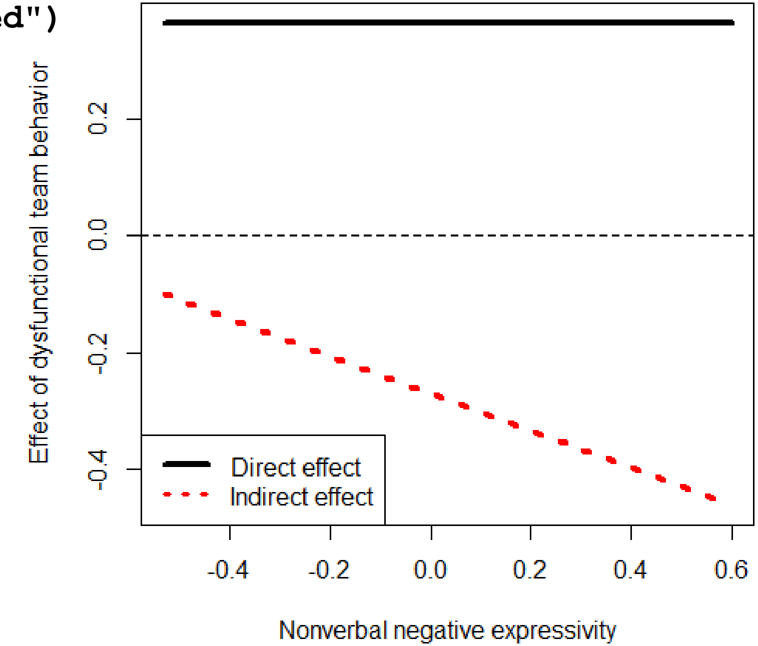
A visual representation of the results in R

```
x<-c(0,1,0,1,0,1)
w<-c(-0.531,-0.531,-0.060,-0.060,0.600,0.600)
y<-c(0.366,-0.100,0.366,-0.251,0.366,-0.462)
plot(y=y,x=w,pch=15,col="white",
xlab="Nonverbal negative expressivity",
ylab="Effect of dysfunctional team behavior")
legend.txt<-c("Direct effect","Indirect effect")
legend("bottomleft",legend=legend.txt,lty=c(1,3),lwd=c(4,3),
col=c("black","red"))
lines(w[x==0],y[x==0],lwd=4,lty=1,col="black")
lines(w[x==1],y[x==1],lwd=4,lty=3,col="red")
abline(0,0,lwd=0.5,lty=2)
```

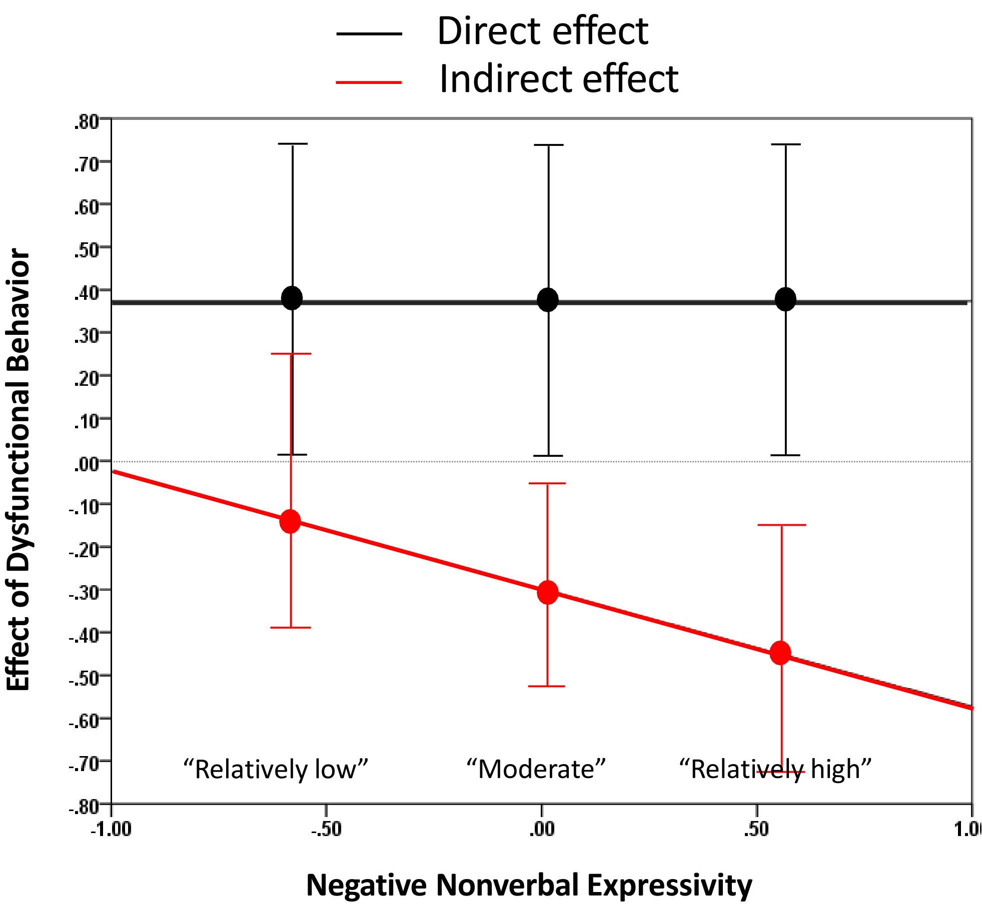
From PROCESS output

	negexp	effect
0	-0.531	0.366
1	-0.531	-0.100
0	-0.060	0.366
1	-0.060	-0.251
0	0.600	0.366
1	0.600	-0.462

0 = direct, 1 = indirect



Putting it all together



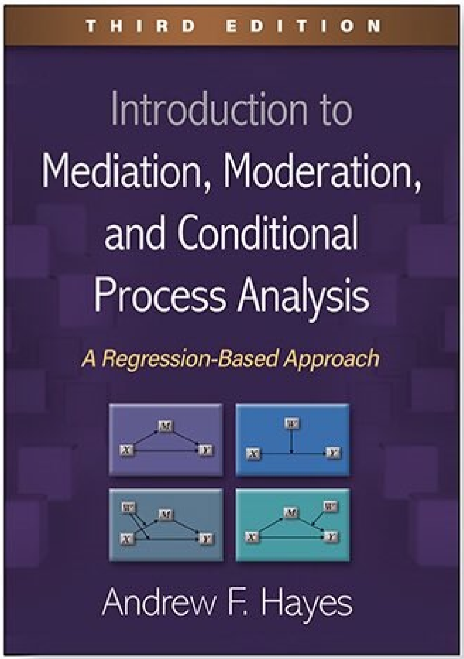
Point estimates with 95% confidence intervals (OLS confidence intervals for direct effects, percentile bootstrap confidence intervals for indirect effects).

More dysfunctional behavior tends to lead to a more negative work climate, yet this negative climate seems to lower performance only among teams that are more demonstrative of their negative feelings. Such a process does not operate among teams that hide their feelings. Independent of differences between teams in the negative affective tone of the work environment, teams that exhibit more dysfunctional behavior otherwise perform *better*.

Setting up PROCESS

PROCESS

- An observed variable OLS regression-based modeling tool for moderation, mediation, and conditional process analysis.



- First released in March of 2012, and re-released (version 3) in 2017 with many new features. V4 was released in 2022, adding in R. PROCESS is documented in *Introduction to Mediation, Moderation, and Conditional Process Analysis.*
- Available for SPSS (in macro and “custom dialog” form), SAS, and R.
- PROCESS is free and can be downloaded from www.processmacro.org. You’ve been given the most current version with your course files.

- It has become widely used in many disciplines. It makes difficult, tedious things easy.

Read the documentation (eventually)

The PROCESS documentation describes how to use PROCESS, as well as its various options, capabilities, and limitations. It is available as Appendix A in Hayes (2021). *Introduction to Mediation, Moderation, and Conditional Process Analysis*. I'll tell you all you need to know for the purposes of this class. PROCESS has 55 models preprogrammed. Diagrams of these models are available in the documentation.

Appendix A
Using PROCESS

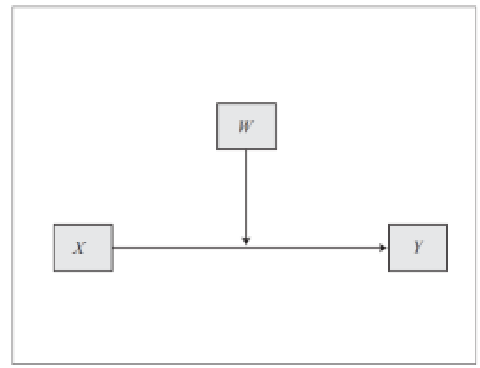
This appendix describes how to install and execute PROCESS, how to set up a PROCESS command, and it documents its many features, some of which are not described elsewhere in this book. As PROCESS is modified and features are added, supplementary documentation will be released at *www.afhayes.com*. Check this web page regularly for updates. Also available at this page is a complete set of model templates identifying each model that PROCESS can estimate.

This documentation focuses on the SPSS version of PROCESS. All features and functions described below are available in the SAS version as well and work as described here, with minor modifications to the syntax. At the end of this documentation (see page 438), a special section devoted to SAS describes some of the differences in syntax structure for the SAS version compared to what is described below.

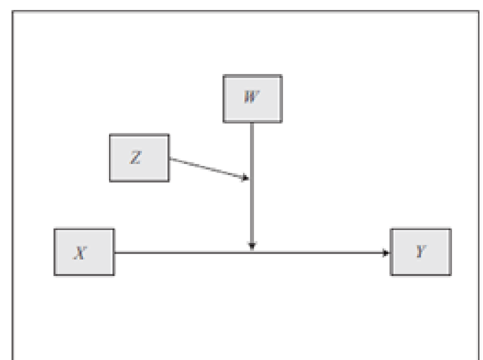
Overview

PROCESS is a computational tool for path analysis–based moderation and mediation analysis as well as their integration in the form of a conditional process model. In addition to estimating unstandardized model coefficients, standard errors, *t* and *p*-values, and confidence intervals using either OLS regression (for continuous outcomes) or maximum likelihood logistic regression (for dichotomous outcomes), PROCESS generates direct and indirect effects in mediation models, conditional effects (i.e., "simple slopes") in moderation models, and conditional indirect effects in conditional process models with a single or multiple mediators. PROCESS offers various methods for probing two- and three-way interactions and can construct percentile bootstrap, bias-corrected bootstrap, and Monte Carlo confidence intervals for indirect effects. In mediation models, multiple mediator variables can be specified to operate in parallel or in serial. Heteroscedasticity-consistent standard errors are available for inference about model coeffi-

Model 1

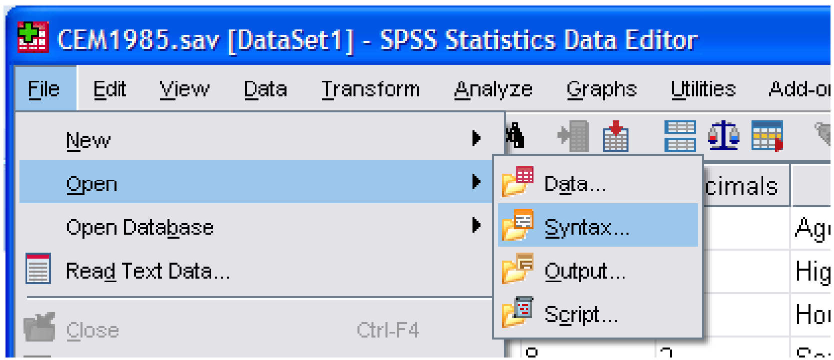


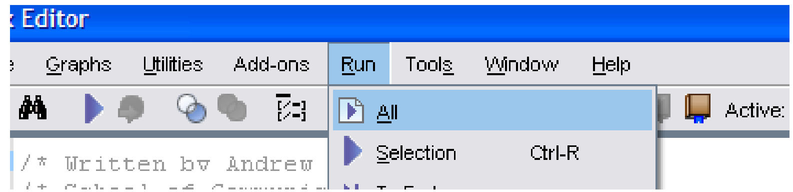
Model 3



PROCESS as a syntax-driven macro in SPSS

Open process (.sps, a “SPSS Syntax file”) and run the entire program **exactly as is.** This produces a new SPSS command called PROCESS.





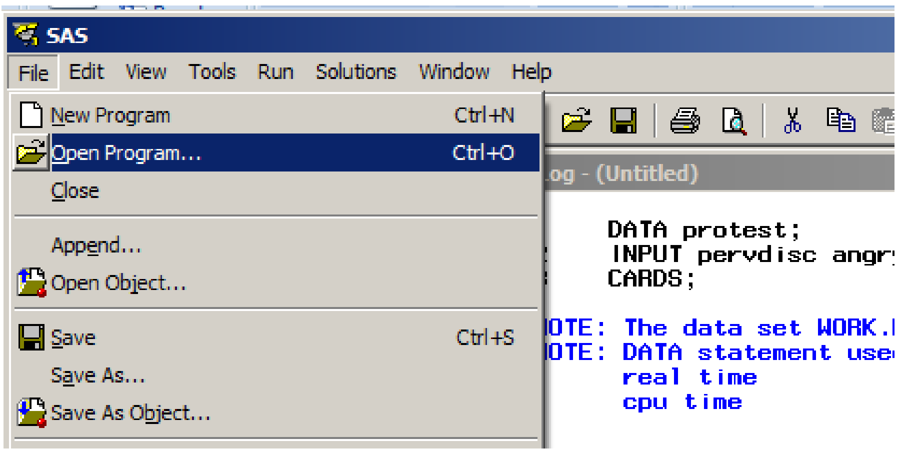
Once PROCESS is defined, you can run a properly-formatted PROCESS command in a new syntax window. Such a command might look something like below:

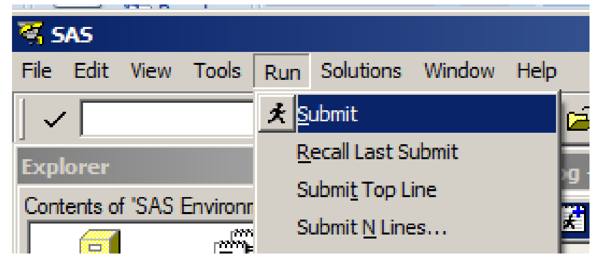
```
process y=react/x=cond/w=pmi/plot=1/model=1.
```

PROCESS goes away when you close SPSS. You have to run process.sps again to reactivate it.

PROCESS for SAS

In SAS, open process.sas and submit the entire program **exactly as is.** This produces a new SAS command called %PROCESS.





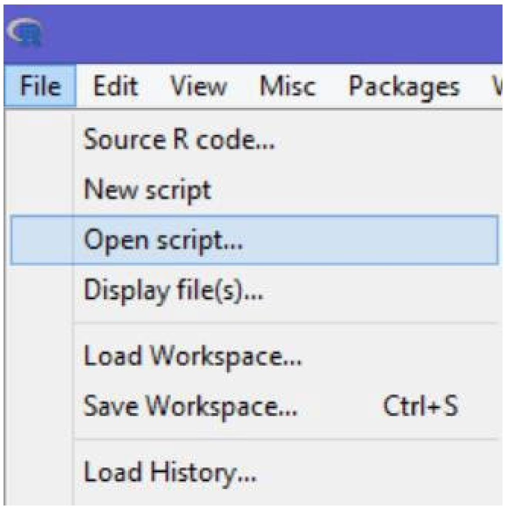
Once PROCESS is defined, you can run a properly-formatted PROCESS command in a new syntax window. Such a command might look something like below:

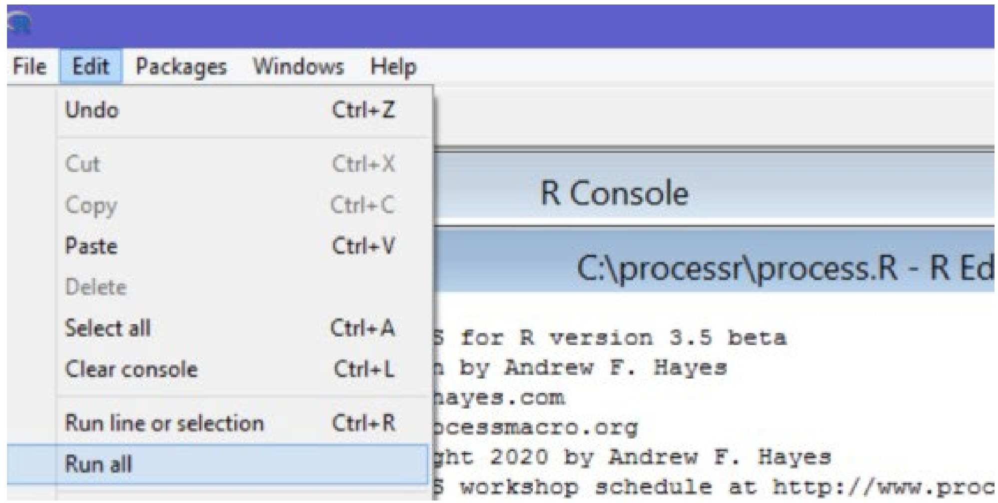
```
%process (data=pmi,y=react,x=cond,m=pmi,plot=1,model=1)
```

PROCESS goes away when you close SAS. You have to run process.sas again to reactivate it.

PROCESS for R

PROCESS for R is a script, not a package. Execute the script by opening **process.r** and running, as below. This can be done with or without R Studio. R Studio may be slower.





The script will scroll past you on the screen as it runs. It will take a few minutes before PROCESS will be ready to use. Save the workspace to avoid having to running the script again next time you open R. If you save the workspace, you can use R like a package next time.
In R studio, you can turn it into a package if you want and know how. Consult many sources available online for instructions on creating a package.

PROCESS is **not** available on CRAN. PROCESSR, which is available on CRAN, is not the same as PROCESS. I didn't create PROCESSR and I can't attest to its accuracy or quality.

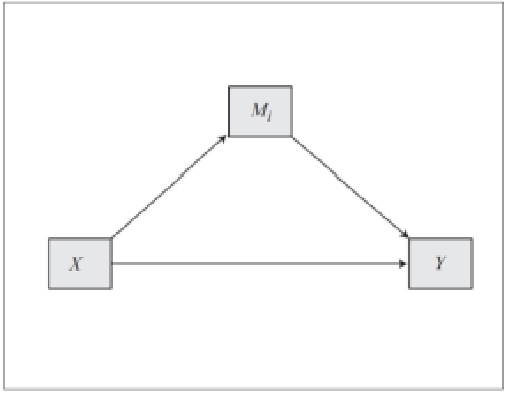
Model template system

PROCESS currently has 55 model templates, referred to by number in PROCESS, which specify the form of the model (moderation, mediation, conditional process) and which variables play what roles.

Example #1:

Model 4 is a simple or parallel multiple mediator model that estimates the direct and indirect effect(s) of *X* on *Y* through one or more mediators (*M*) (up to 10 mediators at once)

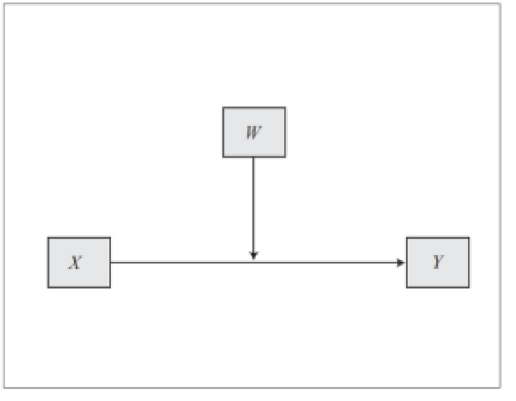
Model 4



Example #2:

Model 1 is a simple moderation model with *W* moderating the effect of *X* on *Y*.

Model 1



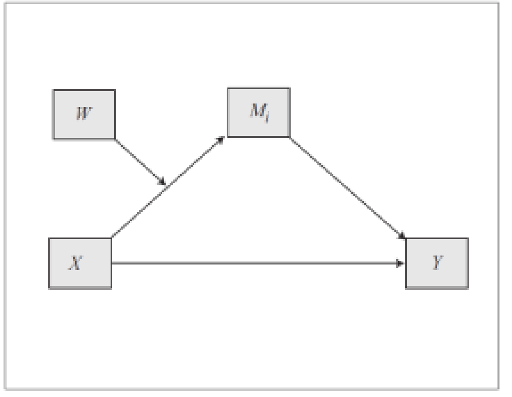
Model template system

PROCESS currently has 55 model templates, referred to by number in PROCESS, which specify the form of the model (moderation, mediation, conditional process) and which variables play what roles.

Example #3:

Model 7 is a "first stage" conditional process model that allows the indirect effect of *X* on *Y* through *M* to depend on *W* as a result of the moderation of the effect of *X* on *M*.

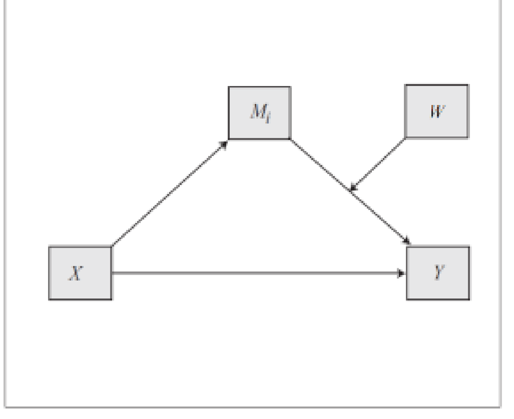
Model 7



Example #4:

Model 14 is a "second stage" conditional process model that allows the indirect effect of *X* on *Y* through *M* to depend on *W* as a result of the moderation of the effect of *M* on *Y*.

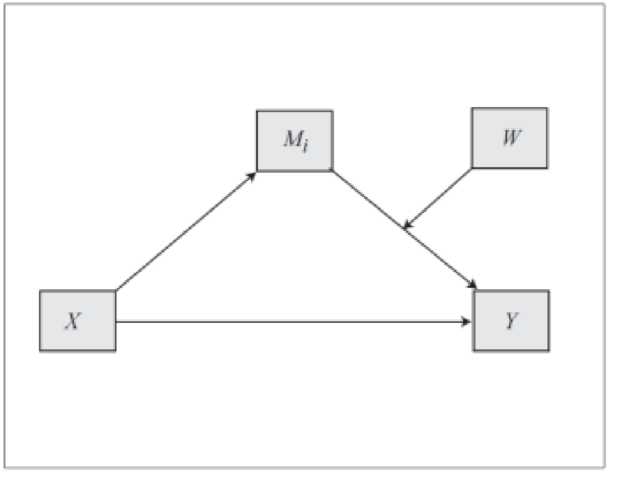
Model 14



Model template system

PROCESS has 55 model templates, referred to by number in PROCESS, which specify the form of the model (moderation, mediation, conditional process) and which variables play what roles.

Model 14



Minimum required specifications

- Which variables play which role in the model (**`y=`** **`x=`** **`m=`** and so forth)
- Model number (**`model=`**)
- In SAS and R only: Data file or data frame being analyzed (**`data=`**)

SPSS

```
process y=yvar/x=xvar/m=mvlist/w=wvar/model=14.
```

SAS

```
%process(data=filename,y=yvar,x=xvar,m=mvlist,w=wvar,model=14)
```

R

```
process(data=filename,y="yvar",x="xvar",m="mvlist",w="wvar",model=14)
```

Comments on syntax and usage

```
process y=withdraw/x=estress/m=affect/model=4/boot=10000/normal=1/total=1/
    stand=1/seed=17623.
```

```
%process (data=estress,y=withdraw,x=estress,m=affect,model=4,boot=10000,normal=1,
    total=1,stand=1,seed=17623)
```

```
process(data=estress,y="withdraw",x="estress",m="affect",model=4,boot=10000,
    normal=1,total=1,stand=1,seed=17623)
```

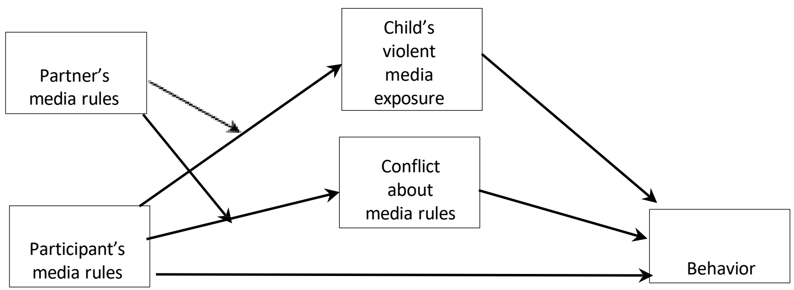
- In R, the data must be in the form of a data frame. In SAS and R, you must specify the data file/data frame you are analyzing. SPSS will analyze the active data file.
- In SPSS, separate options/specifications with a forward slash. In SAS and R, use a comma.
- R is a case-sensitive language. “PROCESS” ≠ “Process” ≠ “process”; “estress” ≠ “ESTRESS”. Type R commands exactly as they appear in these materials. SPSS and SAS are not case sensitive.
- In R, variable names must be in quotes. No quotes are used in SPSS and SAS
- In R, missing values must be represented in the data as “NA”
- PROCESS accepts only numeric data. Data must be numbers or numerical codes.
- PROCESS for R will not accept variables designated as factors.
- In SPSS, best to keep variable names to eight characters or fewer.
- In R, when more than one variable is listed as a mediator or in other multivariate specifications, use the c() operator when listing the names of the variables.

End of review

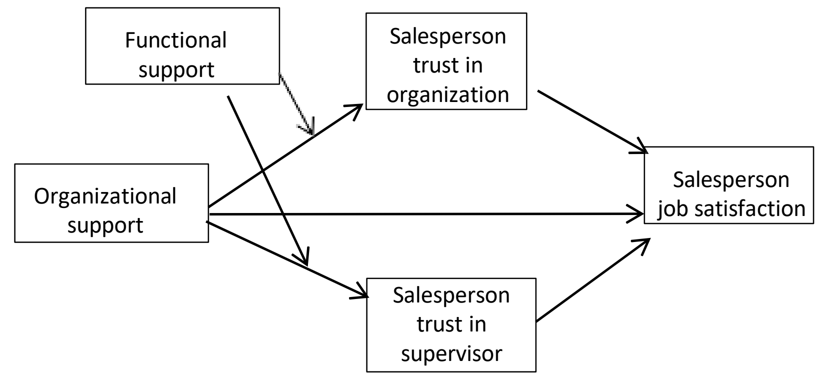
Differential Dominance

Multiple Mediators
One Moderator on Same Path

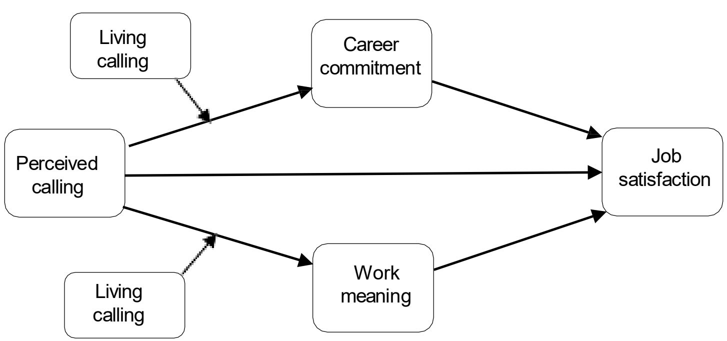
A first stage conditional process model with 2 mediators



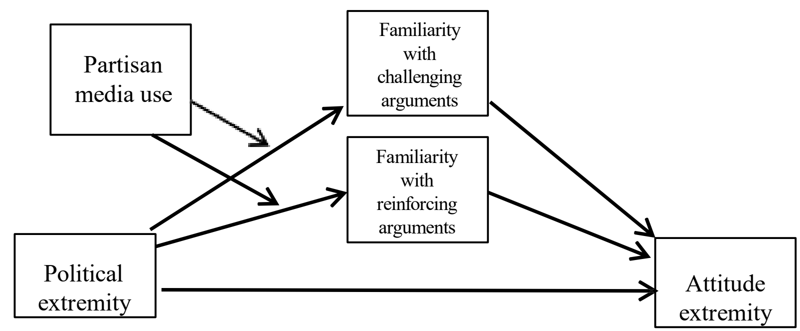
Mares, M-L., Stephenson, L., Martins, N., & Nathanson, A. I. (2018). A house divided: Parental disparity and conflict over media rules predict children's outcomes. ***Computers in Human Behavior**, 81*, 172-188.



Pomireleanu, N., & Mariadoss, B. J. (2015). The influence of organizational and functional support on the development of salesperson job satisfaction. ***Journal of Personal Selling and Sales Management***, 35, 33-50.



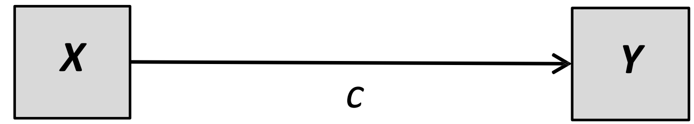
Duffy, R., Bott, E. M., Allan, B. A., Torryey, C. L., & Dik, B. (2012). Perceiving a calling, living a calling, and job satisfaction: Testing a moderated, multiple mediator model. ***Journal of Counseling Psychology***, 59, 50-59.



Gvirsman, S. D. (2014). It's not that we don't know, it's that we don't care: Explaining why selective exposure polarizes attitudes. ***Mass Communication and Society***, 17,74-97.

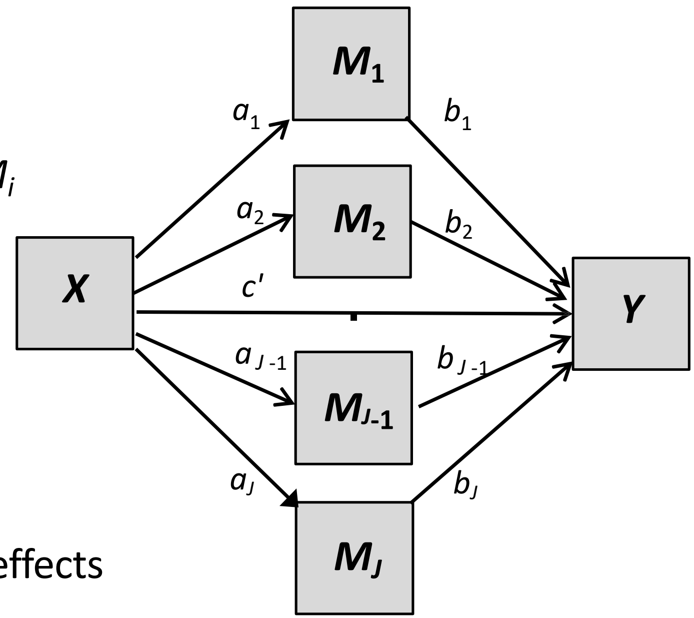
The parallel multiple mediator model

$$\widehat{Y}_i = c_0 + cX_i$$



$$\widehat{M_{ji}} = a_{0j} + a_j X_i$$

$$\widehat{Y}_i = c'_0 + c' X_i + \sum_{j=1}^{J} b_j M_{ji}$$



c = "total effect" of X on Y

$a_j \times b_j$ = "specific indirect effect" of X on Y through M_i

$\Sigma\,(a_j \times b_j)$= "total indirect effect" of X on Y

c' = "direct effect" of X on Y

total effect = direct effect + sum of specific indirect effects

$c \quad = \quad c' \quad + \quad \Sigma\,(a_j \times b_j)$

Example

Chen, C., Wen, P., & Hu, C. (2017). Role of formal mentoring in protégés' work-to-family conflict: A double-edged sword. ***Journal of Vocational Behavior,*** *100*, 101-110.

Artwork by Tumisu via Pixabay

193 employees of a machinery and equipment manufacturing company. Data collection occurred in two waves, with three month period between waves.

Time 1 measures

MENTOR: Extent to which an employee received formal job-related mentoring (higher = more).

Time 2 measures

CONFLICT: Feelings of conflict between the demands of work and the demands of home and family (higher = more). This was measured in the second wave.

RESOURCES: Perceptions of access to more job-related resources (higher = greater access to more).

WORKLOAD: Feelings of workload (higher = sense of greater workload).

The data file is MENTOR

The data: mentor

mentor.sav

*mentor.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing

	mentor	wforient	resource	workload	conflict
1	3.4	3.7	4.2	3.2	1.8
2	3.6	2.6	2.3	3.5	4.1
3	3.3	3.2	3.5	3.3	2.3
4	3.9	4.3	3.5	3.1	2.5
5	3.5	2.9	3.7	2.6	2.6
6	4.1	3.5	4.9	3.8	3.1
7	3.8	2.7	2.6	3.2	3.3
8	5.2	2.9	2.8	3.8	3.4
9	3.6	3.9	4.1	2.1	2.2
10	4.6	3.8	4.1	3.6	2.5

mentor.sas

mentor

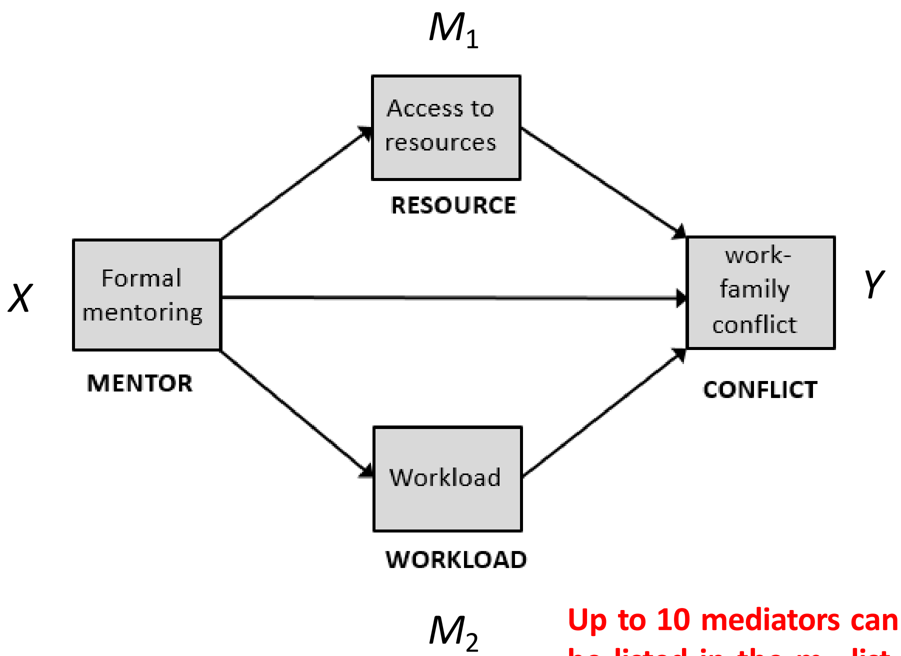
```
data mentor;
  input mentor wforient resource workload conflict;
  datalines;
  3.4 3.7 4.2 3.2 1.8
  3.6 2.6 2.3 3.5 4.1
  3.3 3.2 3.5 3.3 2.3
  3.9 4.3 3.5 3.1 2.5
  3.5 2.9 3.7 2.6 2.6
  4.1 3.5 4.9 3.8 3.1
  3.8 2.7 2.6 3.2 3.3
  5.2 2.9 2.8 3.8 3.4
  3.6 3.9 4.1 2.1 2.2
  4.6 3.8 4.1 3.6 2.5
```

In R: Don't forget to change the path below to where your **mentor.csv** file is located.

```
> mentor<-read.table("c:/mmcpa/mentor.csv", sep=",",header=TRUE)
> head(mentor)
  mentor wforient resource workload conflict
1    3.4      3.7      4.2      3.2      1.8
2    3.6      2.6      2.3      3.5      4.1
3    3.3      3.2      3.5      3.3      2.3
4    3.9      4.3      3.5      3.1      2.5
```

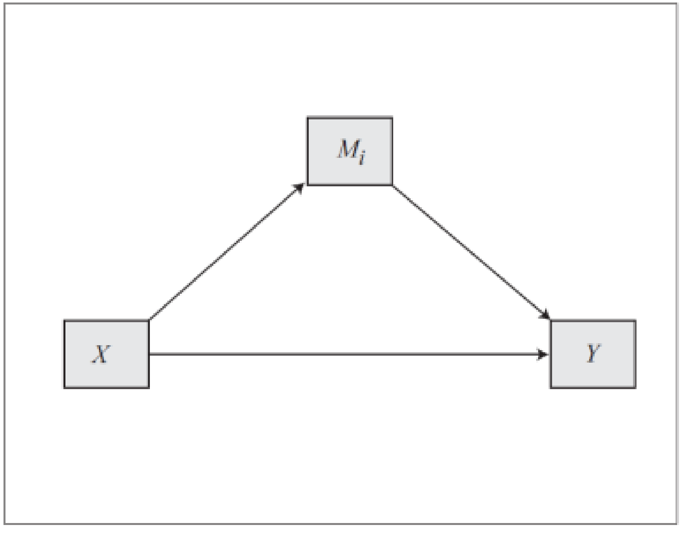
These aren't their actual data. But the analysis we do yields results similar to what they report.

Estimation and inference using PROCESS



PROCESS model 4 is used for simple and parallel multiple mediator models.

Model 4



Up to 10 mediators can be listed in the m= list. Order does not matter.

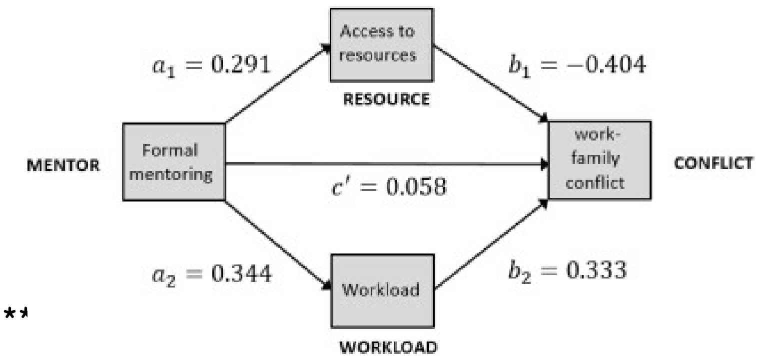
```
process y=conflict/x=mentor/m=resource workload/total=1/model=4/seed=92612.
```

```
%process (data=mentor,y=conflict,x=mentor,m=resource workload,total=1,
   model=4,seed=92612)
```

```
process(data=mentor,y="conflict",x="mentor",m=c("resource", "workload"),
   total=1,model=4,seed=92612)
```

PROCESS output

```
Model  : 4
    Y  : conflict
    X  : mentor
   M1  : resource
   M2  : workload

Sample Size:  193

Custom Seed:  92612
```




```
***********************************************************  **
OUTCOME VARIABLE:
 resource
```

$$\widehat{M_{1i}} = 2.310 + 0.291X_i$$

Model Summary

R	R-sq	MSE	F	df1	df2	p
.1903	.0362	.9833	7.1744	1.0000	191.0000	.0080

Model

	coeff	se	t	p	LLCI	ULCI
constant	2.3102	.4070	5.6760	.0000	1.5073	3.1130
mentor	.2910	.1086	2.6785	.0080	.0767	.5053

a_1 path

```
*****************************************************************************
OUTCOME VARIABLE:
 workload
```

$$\widehat{M_{2i}} = 2.007 + 0.344X_i$$

Model Summary

R	R-sq	MSE	F	df1	df2	p
.2356	.0555	.8778	11.2201	1.0000	191.0000	.0010

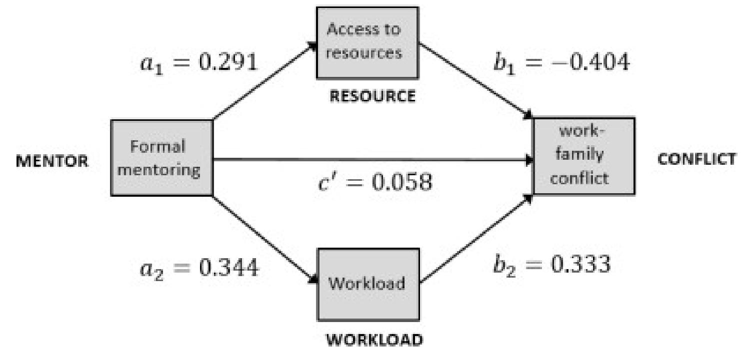
Model

	coeff	se	t	p	LLCI	ULCI
constant	2.0070	.3845	5.2193	.0000	1.2485	2.7655
mentor	.3438	.1026	3.3496	.0010	.1414	.5463

a_2 path

PROCESS output

$$\widehat{Y}_i = 3.111 + 0.058X_i - 0.404M_{1i} + 0.333M_{2i}$$



OUTCOME VARIABLE:
conflict

Model Summary

R	R-sq	MSE	F	df1	df2	p
.7658	.5864	.2444	89.3290	3.0000	189.0000	.0000

Model

	coeff	se	t	p	LLCI	ULCI
constant	3.1114	.2459	12.6538	.0000	2.6264	3.5965
mentor	.0583	.0579	1.0070	.3152	-.0559	.1726
resource	-.4036	.0383	-10.5387	.0000	-.4792	-.3281
workload	.3332	.0405	8.2204	.0000	.2533	.4132

c' path
b_1 path
b_2 path

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

OUTCOME VARIABLE:
conflict

$$\widehat{Y}_i = 2.848 + 0.056X_i$$

Model Summary

R	R-sq	MSE	F	df1	df2	p
.0479	.0023	.5835	.4392	1.0000	191.0000	.5083

Model

	coeff	se	t	p	LLCI	ULCI
constant	2.8478	.3135	9.0831	.0000	2.2294	3.4662
mentor	.0555	.0837	.6627	.5083	-.1096	.2205

c path

PROCESS output

```
************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Total effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .0555      .0837      .6627      .5083     -.1096      .2205

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .0583      .0579     1.0070      .3152     -.0559      .1726

Indirect effect(s) of X on Y:
             Effect     BootSE   BootLLCI   BootULCI
TOTAL        -.0029      .0716     -.1367      .1427
resource     -.1175      .0456     -.2085     -.0309
workload      .1146      .0391      .0453      .1988
```

c path

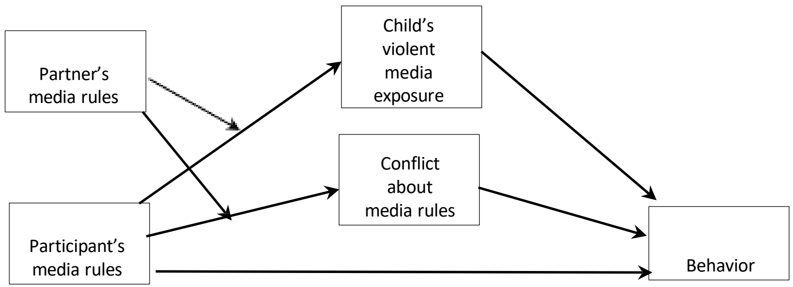
c′ path

a_1b_1 with bootstrap CI

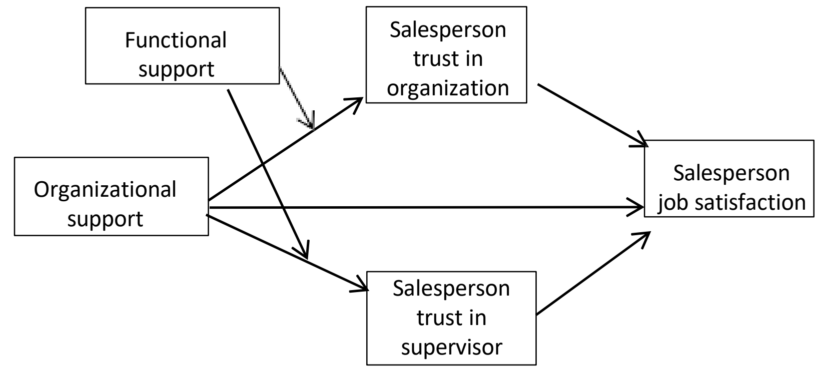
a_2b_2 with bootstrap CI

The data are consistent with the claim that formal mentoring negatively influences work-family conflict by increasing access to resources, which reduces conflict (-0.118; 95% CI = -0.209 to -0.031). Furthermore, greater levels of mentoring results in greater conflict through higher workload, with in turn elevates conflict (0.115, 95% CI = 0.045 to 0.199). Independent of these mechanisms, there is no relationship between mentoring and work-family conflict.

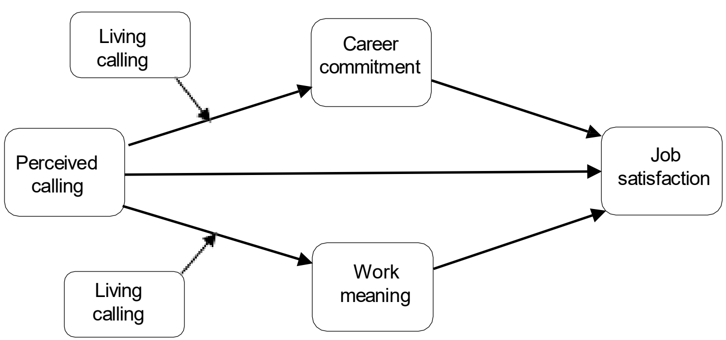
A first stage conditional process model with 2 mediators



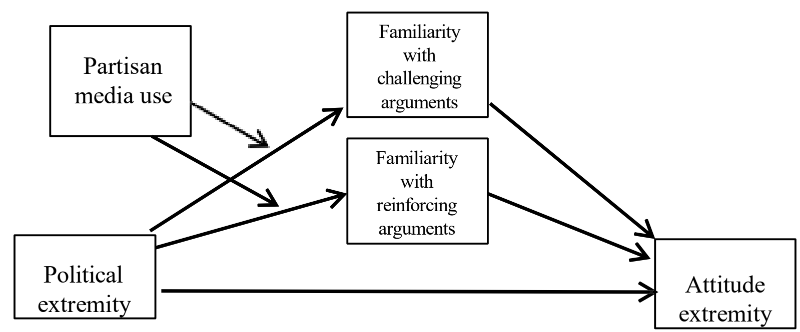
Mares, M-L., Stephenson, L., Martins, N., & Nathanson, A. I. (2018). A house divided: Parental disparity and conflict over media rules predict children's outcomes. ***Computers in Human Behavior**, 81*, 172-188.



Pomireleanu, N., & Mariadoss, B. J. (2015). The influence of organizational and functional support on the development of salesperson job satisfaction. ***Journal of Personal Selling and Sales Management***, 35, 33-50.



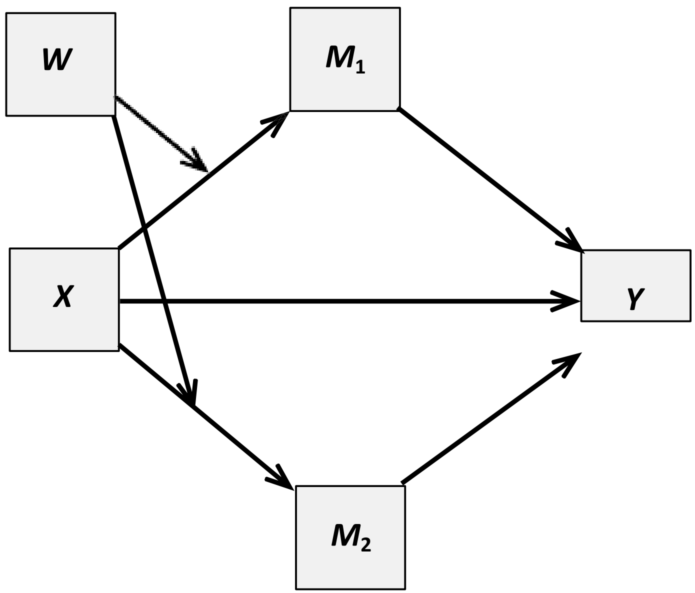
Duffy, R., Bott, E. M., Allan, B. A., Torryey, C. L., & Dik, B. (2012). Perceiving a calling, living a calling, and job satisfaction: Testing a moderated, multiple mediator model. ***Journal of Counseling Psychology***, 59, 50-59.



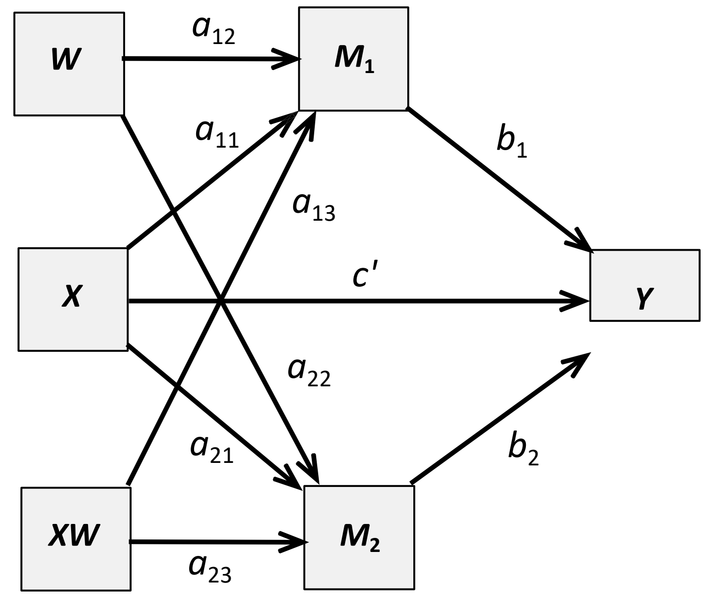
Gvirsman, S. D. (2014). It's not that we don't know, it's that we don't care: Explaining why selective exposure polarizes attitudes. ***Mass Communication and Society***, 17,74-97.

The model equations

Conceptual Model



Statistical Model



Write out the three equations for this model

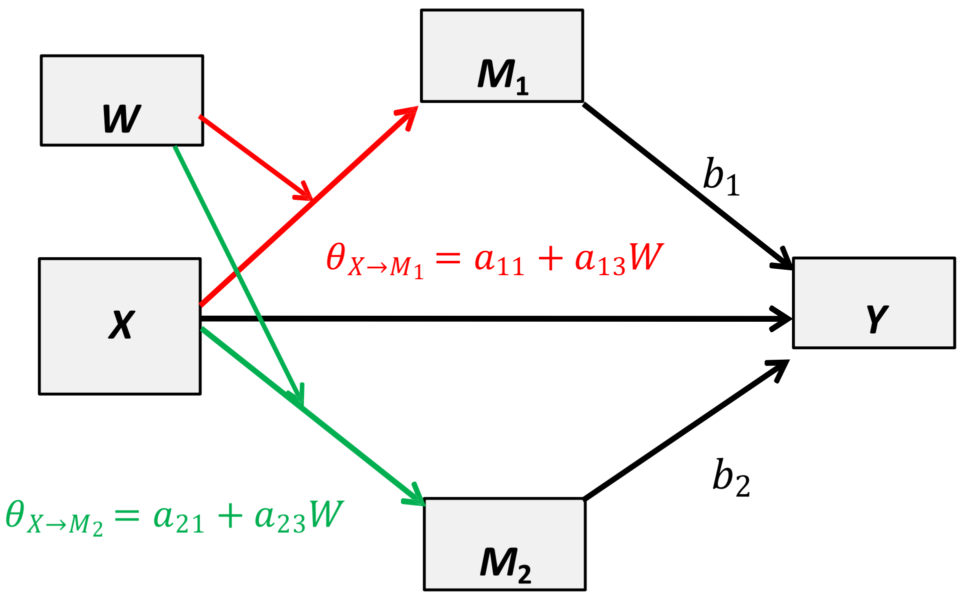
$$\widehat{M_{1i}} =$$

$$\widehat{M_{2i}} =$$

$$\widehat{Y_i} =$$

The conditional effects of *X* on M_1 and M_2

The effects of *X* on M_1 and M_2 are both linear functions of *W*.



$$\widehat{M_{1i}} = a_{10} + a_{11}X_i + a_{12}W_i + a_{13}X_iW_i$$

which can be written equivalently as

$$\widehat{M_{1i}} = a_{10} + (a_{11}+a_{13}W_i)X_i + a_{12}W_i$$

or

$$\widehat{M_{1i}} = a_{10} + \theta_{X\rightarrow M_1}X_i + a_{12}W_i$$

where $\boldsymbol{\theta_{X\rightarrow M_1} = a_{11} + a_{13}W}$

$$\widehat{M_{2i}} = a_{20} + a_{21}X_i + a_{22}W_i + a_{23}X_iW_i$$

which can be written equivalently as

$$\widehat{M_{2i}} = a_{20} + (a_{21} + a_{23}W_i)X_i + a_{22}W_i$$

or

$$\widehat{M_{2i}} = a_{20} + \theta_{X\rightarrow M_2}X_i + a_{22}W_i$$

where $\boldsymbol{\theta_{X\rightarrow M_2} = a_{21} + a_{23}W}$

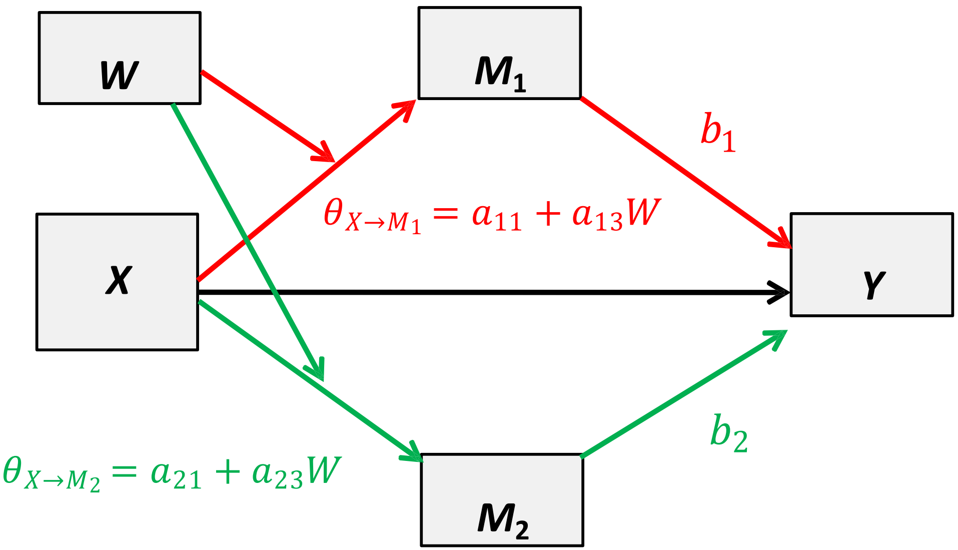
The conditional indirect effects

The specific indirect effects of *X* are also linear functions of *W*.

$$\widehat{M_{1i}} = a_{10} + (a_{11} + a_{13}W_i)X_i + a_{12}W_i$$

$$\widehat{M_{2i}} = a_{20} + (a_{21} + a_{23}W_i)X_i + a_{22}W_i$$

$$\widehat{Y_i} = c_0' + c'X_i + b_1M_{1i} + b_2M_{2i}$$



Indirect effect of *X* on *Y* through M_1 depends on *W*:

$$\theta_{X\rightarrow M_1}b_1 = (a_{11} + a_{13}W)b_1$$
$$= a_{11}\,b_1 + a_{13}b_1W$$

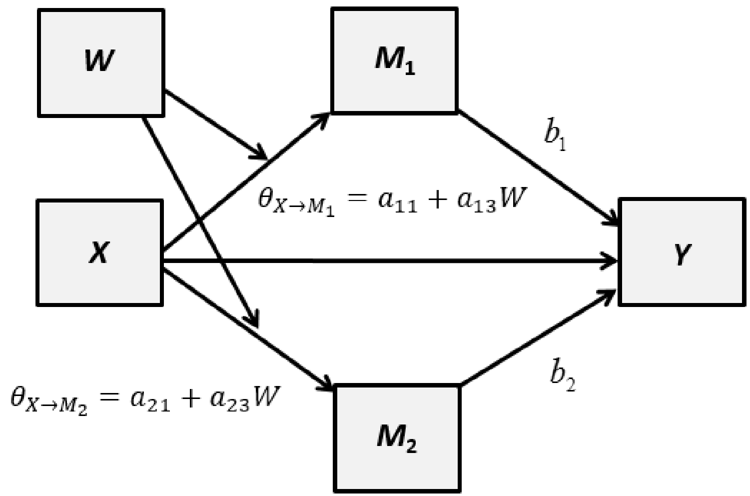
Indirect effect of *X* on *Y* through M_2 depends on *W*:

$$\theta_{X\rightarrow M_2}b_2 = (a_{21} + a_{23}W)b_2$$
$$= a_{21}\,b_2 + a_{23}b_2W$$

The indirect effects of *X* are linear functions of *W*. $a_{13}b_1$ and $a_{23}b_2$ are the indices of moderated mediation for this model. They quantify the relationship between *W* and the size of the indirect effect of X.

Differential dominance

This a *differential dominance conditional process model*. It is an intriguing model that allows mechanisms to be different for different "types" of people. For some types, mechanism 1 may be dominant, whereas for other types, mechanism 2 may dominate.



$$\widehat{M_{1i}} = a_{10} + (a_{11}+a_{13}W_i)X_i + a_{12}W_i$$

$$\widehat{M_{2i}} = a_{20} + (a_{21} + a_{23}W_i)X_i + a_{22}W_i$$

$$\widehat{Y_i} = c_0' + c'X_i + b_1M_{1i} + b_2M_{2i}$$

Suppose that upon estimation with some data, we find a_{11} = 0.40, a_{13} = -0.40, a_{21} = 0.00, a_{23} = 0.60, and so

$$\theta_{X\to M_1} = a_{11} + a_{13}W = 0.40 - 0.40W$$

$$\theta_{X\to M_2} = a_{21} + a_{23}W = 0.00 + 0.60W$$

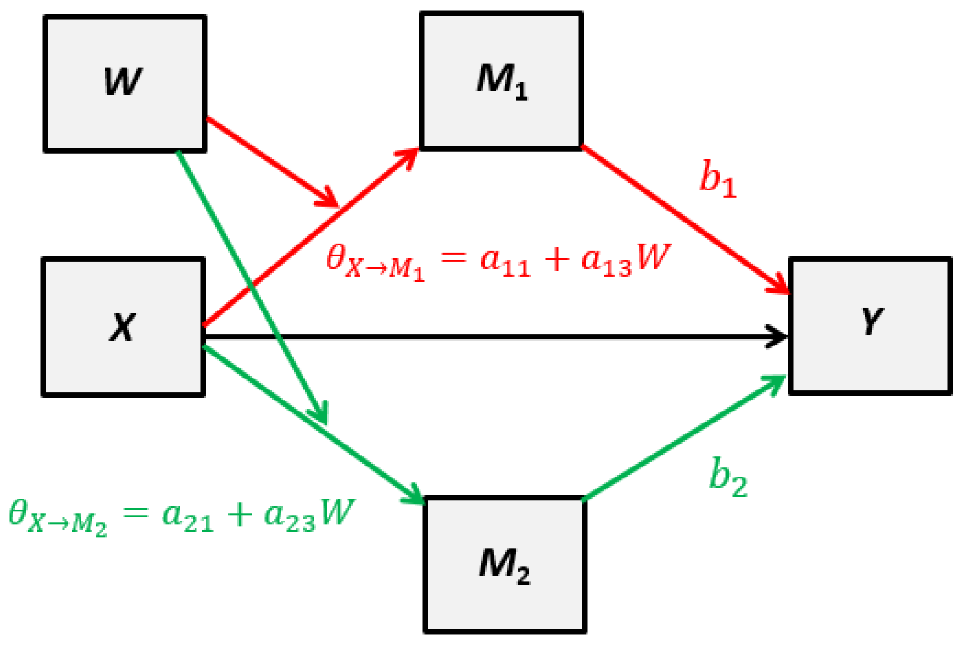
Further suppose that b_1 = 0.50, b_2 = 0.30.

Differential dominance

So the indirect effect of X on Y through M_1 and M_2 are, respectively,

$$\theta_{X\to M_1}b_1 = (a_{11} + a_{13}W)b_1 = (0.40 + 0.40W)0.50 = 0.20\ - 0.20W$$

$$\theta_{X\to M_2}b_2 = (a_{21} + a_{23}W)b_2 = (0.00 + 0.60W)0.30 = 0.00 + 0.18W$$



Differential dominance

Suppose W = 0 for people of "Type A" and W = 1 for people of "Type B". Therefore…

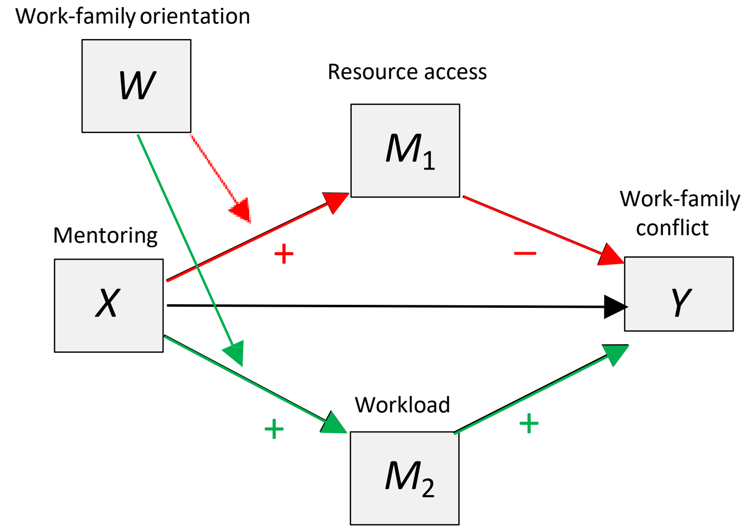
Specific indirect effect	Type A (W = 0)	Type B (W = 1)
through M_1	$(a_{11}+a_{13}W)b_1$ $=(0.40-0.40\times 0)0.50$ $=0.20-0.20\times 0=\mathbf{0.20}$	$(a_{11}+a_{13}W)b_1$ $=(0.40-0.40\times 1)0.50$ $=0.20-0.20\times 1=0.00$
through M_2	$(a_{21}+a_{23}W)b_2$ $=(0.00+0.60\times 0)0.30$ $=0.00+0.18\times 0=0.00$	$(a_{21}+a_{23}W)b_2$ $=(0.00+0.60\times 1)0.30$ $=0.00+0.18\times 1=\mathbf{0.18}$

So for people of "type A" (e.g., W = 0) X affects Y through M_1 but not through M_2. But for people of "type B" (e.g., W = 1) X affects Y through M_2 but not through M_1.

Recall the parallel multiple mediator model from earlier

Chen, C., Wen, P., & Hu, C. (2017). Role of formal mentoring in protégés' work-to-family conflict: A double-edged sword. ***Journal of Vocational Behavior,*** *100*, 101-110.

Artwork by Tumisu via Pixabay



Let's add moderation in the first stage, allowing the indirect effects of formal mentoring on work-family conflict through resource access and workload to both differ as a function of a person's **work-family orientation** (as the authors of this study did).

WFORIENT: Personal identity as work or family-oriented (higher = more work oriented identity).

The data: mentor

mentor.sav

*mentor.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing

	mentor	wforient	resource	workload	conflict
1	3.4	3.7	4.2	3.2	1.8
2	3.6	2.6	2.3	3.5	4.1
3	3.3	3.2	3.5	3.3	2.3
4	3.9	4.3	3.5	3.1	2.5
5	3.5	2.9	3.7	2.6	2.6
6	4.1	3.5	4.9	3.8	3.1
7	3.8	2.7	2.6	3.2	3.3
8	5.2	2.9	2.8	3.8	3.4
9	3.6	3.9	4.1	2.1	2.2
10	4.6	3.8	4.1	3.6	2.5

mentor.sas

mentor

```
data mentor;
  input mentor wforient resource workload conflict;
  datalines;
  3.4 3.7 4.2 3.2 1.8
  3.6 2.6 2.3 3.5 4.1
  3.3 3.2 3.5 3.3 2.3
  3.9 4.3 3.5 3.1 2.5
  3.5 2.9 3.7 2.6 2.6
  4.1 3.5 4.9 3.8 3.1
  3.8 2.7 2.6 3.2 3.3
  5.2 2.9 2.8 3.8 3.4
  3.6 3.9 4.1 2.1 2.2
  4.6 3.8 4.1 3.6 2.5
```

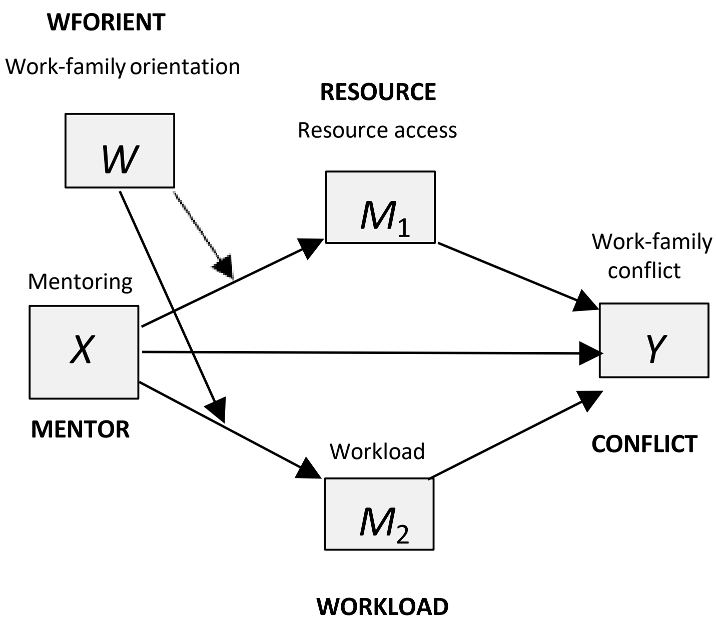
In R: Don't forget to change the path below to where your **mentor.csv** file is located.

```
> mentor<-read.table("c:/mmcpa/mentor.csv", sep=",",header=TRUE)
> head(mentor)
  mentor wforient resource workload conflict
1    3.4      3.7      4.2      3.2      1.8
2    3.6      2.6      2.3      3.5      4.1
3    3.3      3.2      3.5      3.3      2.3
4    3.9      4.3      3.5      3.1      2.5
```

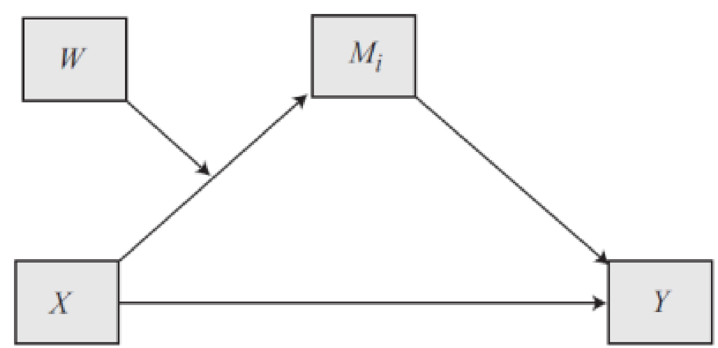
These aren't their actual data. But the analysis we do yields results similar to what they report.

Implementation in PROCESS

This is PROCESS model 7



Model 7



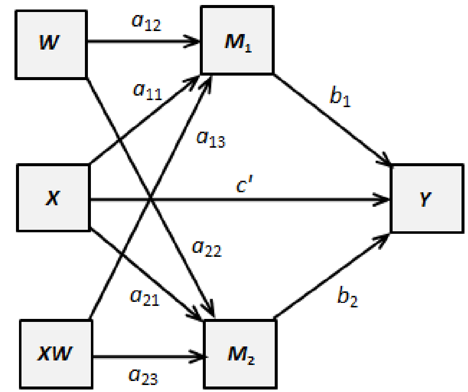
In conditional process with multiple mediators, PROCESS replicates moderated paths in diagram across all mediators

```
process y=conflict/x=mentor/m=resource workload/w=wforient/plot=1/model=7/seed=8231.
```

```
%process (data=mentor,y=conflict,x=mentor,m=resource workload,w=wforient,plot=1,
   model=7,seed=8231)
```

```
process(data=mentor,y="conflict",x="mentor",m=c("resource","workload"),w="wforient",
   plot=1,model=7,seed=8231)
```

PROCESS output




```
Model  : 7
    Y  : conflict
    X  : mentor
   M1  : resource
   M2  : workload
    W  : wforient

Sample
Size:  193

Custom
Seed:      8231

**************************************************************************
OUTCOME VARIABLE:
 resource

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .5140      .2642      .7586    22.6253     3.0000   189.0000      .0000

Model
               coeff         se          t          p       LLCI       ULCI
constant      3.6395     1.4671     2.4806      .0140      .7454     6.5335
mentor        -.6137      .3887    -1.5787      .1161    -1.3804      .1531
wforient      -.3177      .4497     -.7066      .4807    -1.2048      .5693
Int_1          .2609      .1179     2.2134      .0281      .0284      .4934

Product terms key:
 Int_1     :        mentor    x        wforient

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*W      .0191     4.8994     1.0000   189.0000      .0281
```

a_{11} = -0.614
a_{13} = 0.261

a_{11}

a_{13}

The effect of extent of formal mentoring on access to resources depends on work-family orientation.

PROCESS output

PROCESS sees that the moderator *W* is a continuum and generates estimates of the conditional effect of *X* on M_1 for values of *W* corresponding to the 16th, 50th, and 84th percentiles of the distribution (the default). The **plot** option generates data to help visualize the model of M_1.

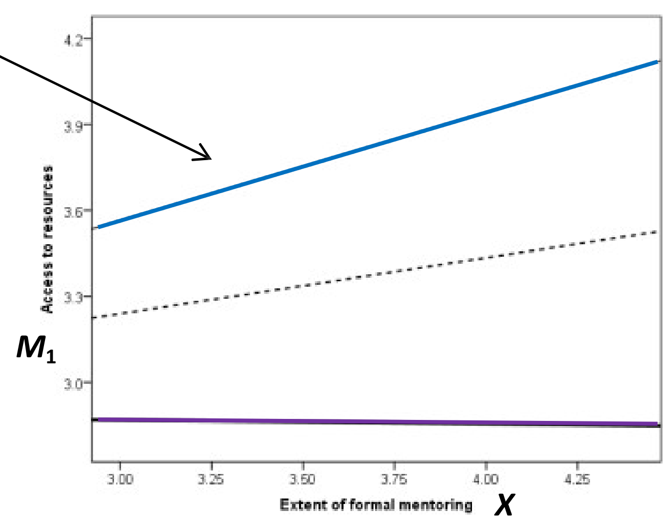
$$\theta_{X \rightarrow M_1} = a_{11} + a_{13}W = -0.614 + 0.261W$$

```
Focal predict: mentor (X)
      Mod var: wforient (W)

Conditional effects of the focal predictor at values of the moderator(s):
```

W	wforient	Effect	se	t	p	LLCI	ULCI
16th	2.3000	-.0136	.1427	-.0956	.9240	-.2951	.2678
50th	3.1000	.1951	.0966	2.0198	.0448	.0046	.3856
84th	3.8000	.3777	.1194	3.1620	.0018	.1421	.6133

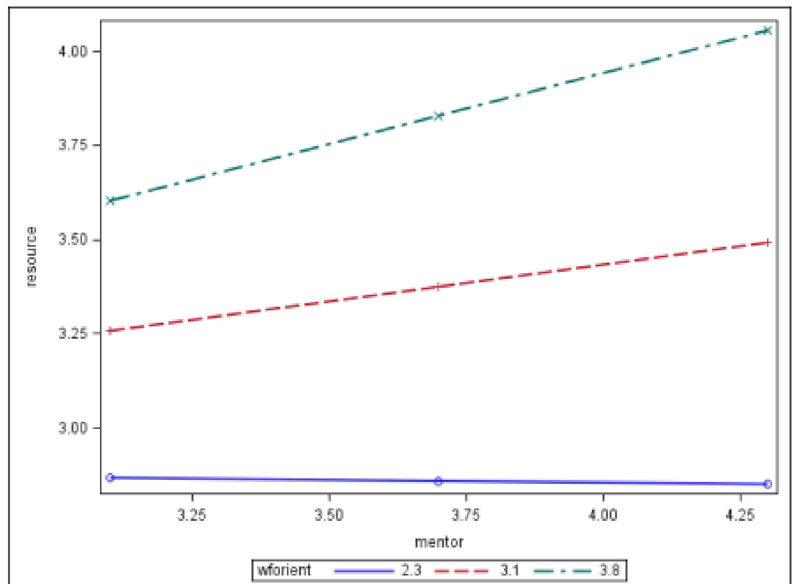
The slopes of these lines

```
Data for visualizing the conditional effect of the focal predictor:
Paste text below into a SPSS syntax window and execute to produce plot.

DATA LIST FREE/
         mentor wforient  resource.
 BEGIN DATA.
      3.1000     2.3000     2.8664
      3.7000     2.3000     2.8582
      4.3000     2.3000     2.8501
      3.1000     3.1000     3.2592
      3.7000     3.1000     3.3763
      4.3000     3.1000     3.4933
      3.1000     3.8000     3.6029
      3.7000     3.8000     3.8296
      4.3000     3.8000     4.0562
END DATA.
GRAPH/SCATTERPLOT=
 mentor    WITH      resource BY       wforient .
```



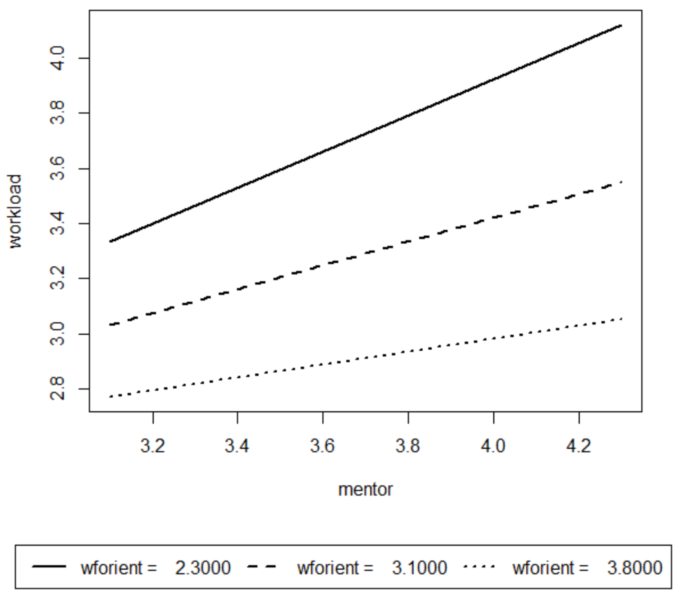
Visualization in SAS

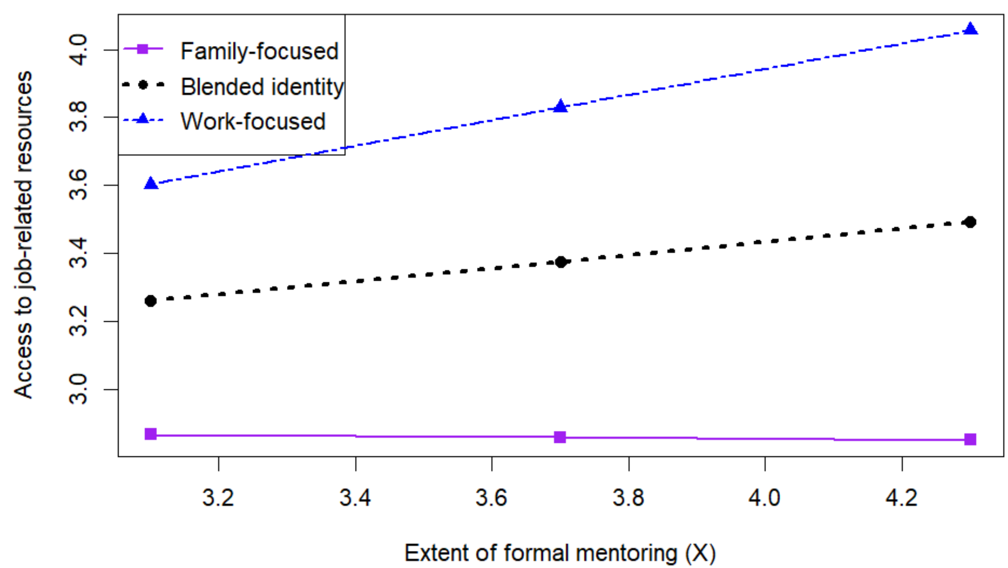
```
data;
input mentor wforient resource;
cards;
     3.1000      2.3000      2.8664
     3.7000      2.3000      2.8582
     4.3000      2.3000      2.8501
     3.1000      3.1000      3.2592
     3.7000      3.1000      3.3763
     4.3000      3.1000      3.4933
     3.1000      3.8000      3.6029
     3.7000      3.8000      3.8296
     4.3000      3.8000      4.0562

run;
proc sgplot;reg x=mentor y=resource/group=wforient;run;
```



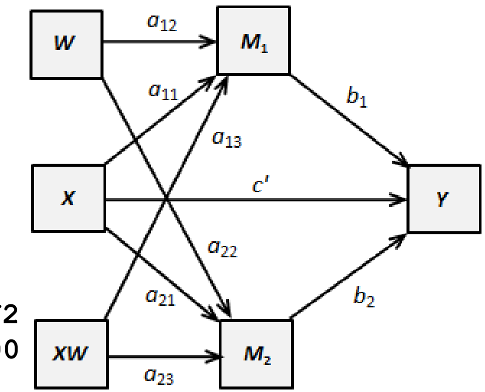
Visualization in R

```
modsave <- process(data=mentor,y="conflict",x="mentor",m=c("resource","workload"),
                   w="wforient",  plot=1,model=7,seed=8231, save = 2)
x <- modsave[10:18,1]
w <- modsave[10:18, 2]
y <- modsave[10:18, 3]
wmarker<-c(15,15,15,16,16,16,17,17,17)
plot(y=y,x=x,cex=1.2,pch=wmarker,  xlab="Extent of formal mentoring (X)", ylab="Access to
     job-related resources",col=c("purple","purple","purple","black","black","black",
     "blue","blue","blue"))
legend.txt<-c("Family-focused","Blended identity", "Work-focused")
legend("topleft", legend = legend.txt,cex=1,lty=c(1,3,6),lwd=c(2,3,2),
pch=c(15,16,17),col=c("purple","black","blue"))  lines(x[w==2.3],y[w==2.3],lwd=2,col="purple")
lines(x[w==3.1],y[w==3.1],lwd=3,lty=3,col="black")
lines(x[w==3.8],y[w==3.8],lwd=2,lty=6,col="blue")
```





PROCESS output




```
OUTCOME VARIABLE:
 workload

Model Summary
          R       R-sq        MSE          F        df1        df2
      .4921      .2422      .7117    20.1341     3.0000   189.0000

Model
              coeff         se          t          p       LLCI       ULCI
constant      .1593     1.4211      .1121      .9109    -2.6439     2.9625
mentor       1.3010      .3765     3.4555      .0007      .5583     2.0437
wforient      .4965      .4356     1.1399      .2558     -.3627     1.3557
 Int_1       -.2808      .1142    -2.4599      .0148     -.5060     -.0556

Product terms key:
 Int_1    :        mentor   x        wforient

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*W      .0243     6.0509     1.0000   189.0000      .0148
```

← a_{21}

← a_{23}

a_{21} = 1.301

a_{23} = -0.281

The effect of extent of formal mentoring on workload depends on work-family orientation.

PROCESS output

PROCESS sees that the moderator *W* is continuous, so it produces the conditional effect of *X* on *M* at three values of the moderator. The **plot** option generates data to help visualize the model of *M*.

$$\theta_{X \rightarrow M_2} = a_{21} + a_{23}W = 1.301 - 0.281W$$

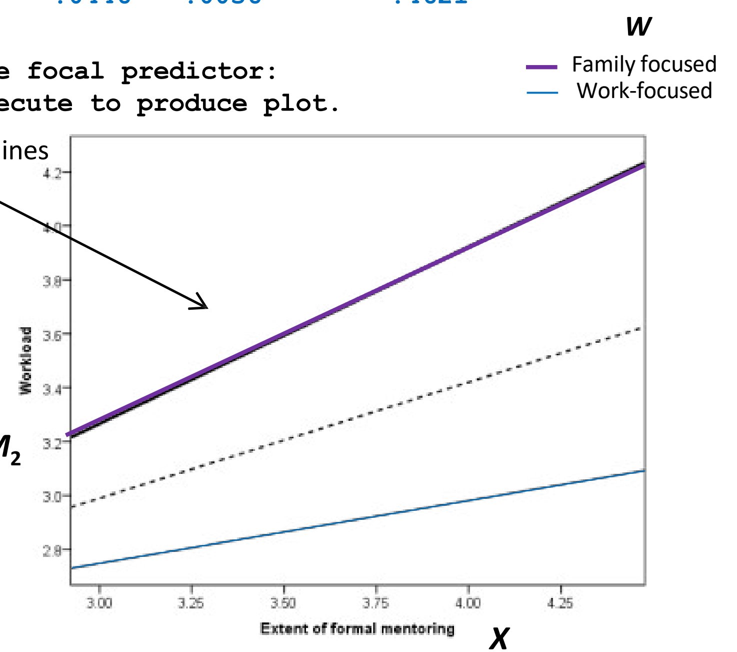
```
     Focal predict: mentor (X)
           Mod var: wforient (W)

 Conditional effects of the focal predictor at values of the moderator(s):
```

W

wforient	Effect	se	t	p	LLCI	ULCI
2.3000	.6551	.1382	4.7409	.0000	.3825	.9277
3.1000	.4305	.0935	4.6013	.0000	.2459	.6150
3.8000	.2339	.1157	2.0214	.0446	.0056	.4621

```
 Data for visualizing the conditional effect of the focal predictor:
 Paste text below into a SPSS syntax window and execute to produce plot.

 DATA LIST FREE/
       mentor    wforient   workload.
 BEGIN DATA.
       3.1000    2.3000     3.3321
       3.7000    2.3000     3.7251
       4.3000    2.3000     4.1182
       3.1000    3.1000     3.0328
       3.7000    3.1000     3.2911
       4.3000    3.1000     3.5494
       3.1000    3.8000     2.7710
       3.7000    3.8000     2.9113
       4.3000    3.8000     3.0516
 END DATA.
 GRAPH/SCATTERPLOT=
  mentor     WITH      workload BY        wforient .
```

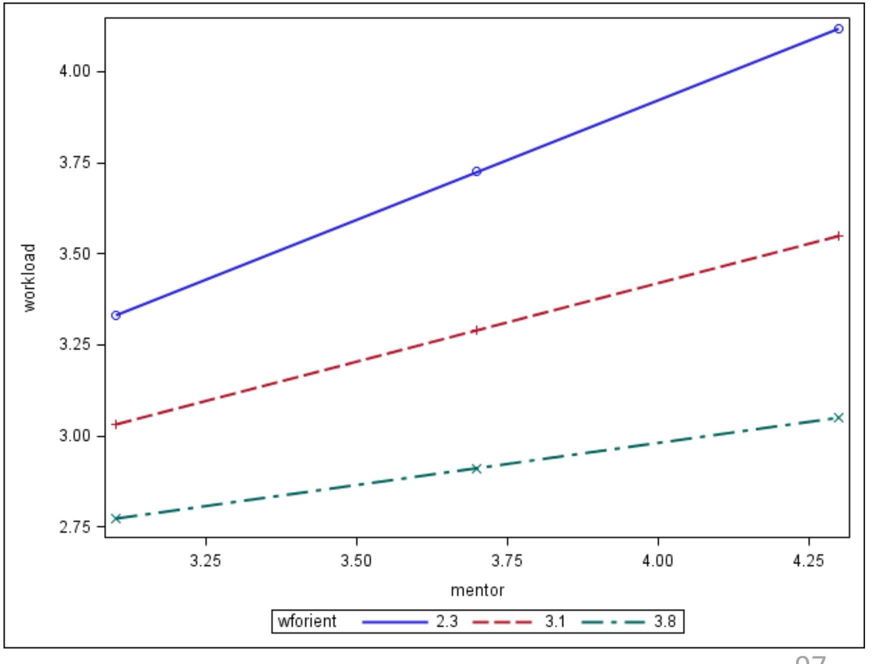
The slopes of these lines



M_2

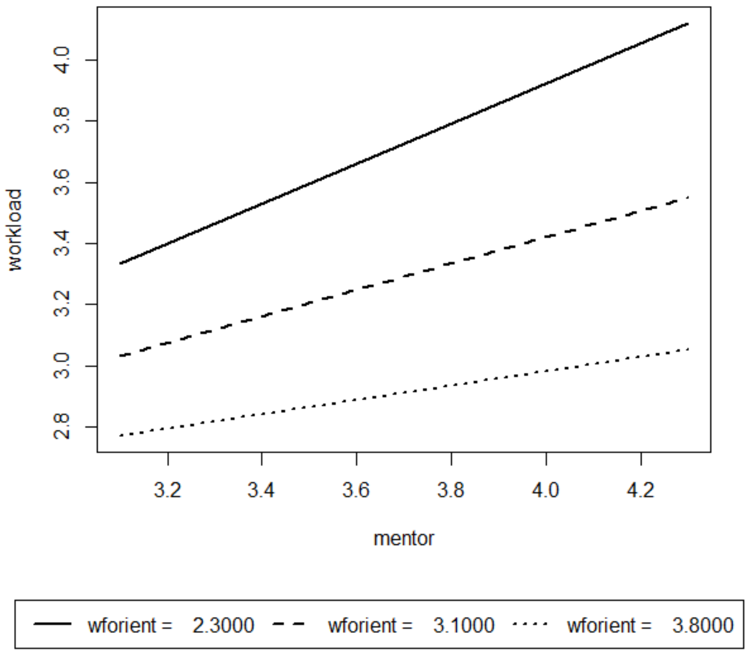
Visualization in SAS

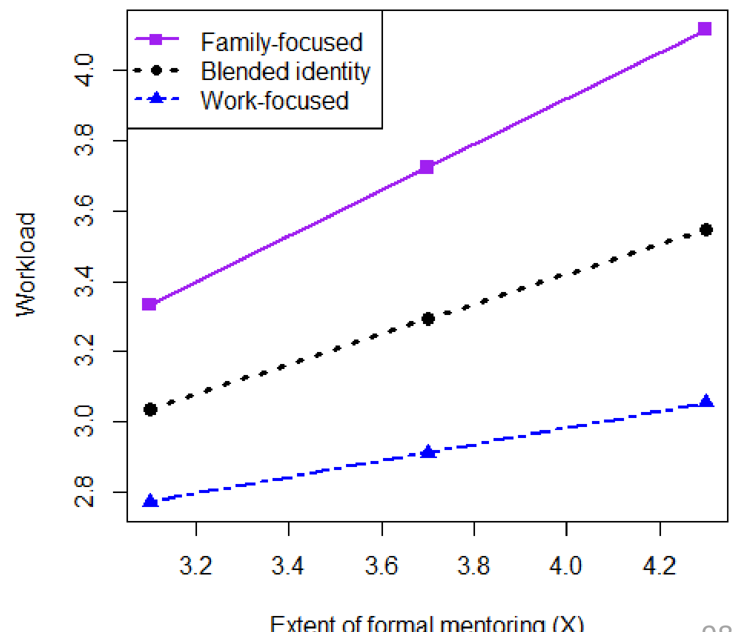
```
data;
input mentor wforient workload;
cards;
     3.1000      2.3000      3.3321
     3.7000      2.3000      3.7251
     4.3000      2.3000      4.1182
     3.1000      3.1000      3.0328
     3.7000      3.1000      3.2911
     4.3000      3.1000      3.5494
     3.1000      3.8000      2.7710
     3.7000      3.8000      2.9113
     4.3000      3.8000      3.0516

run;
proc sgplot;reg x=mentor y=workload/group=wforient;run;
```



Visualization in R

```
modsave <- process(data=mentor,y="conflict",x="mentor",m=c("resource","workload"),
                   w="wforient",  plot=1,model=7,seed=8231, save = 2)
x <- modsave[28:36,1]
w <- modsave[28:36, 2]
y <- modsave[28:36, 3]
wmarker<-c(15,15,15,16,16,16,17,17,17)
plot(y=y,x=x,cex=1.2,pch=wmarker,  xlab="Extent of formal mentoring (X)",
ylab="Workload",col=c("purple","purple","purple",1,1,1,4,4,4))  legend.txt<-c("Family-
focused","Blended identity", "Work-focused")  legend("topleft", legend =
legend.txt,cex=1,lty=c(1,3,6),lwd=c(2,3,2), pch=c(15,16,17),col=c("purple","black","blue"))
lines(x[w==2.3],y[w==2.3],lwd=2,col="purple")  lines(x[w==3.1],y[w==3.1],lwd=3,lty=3,col="black")
lines(x[w==3.8],y[w==3.8],lwd=2,lty=6,col="blue")
```





PROCESS output

$c' = 0.058$

$b_1 = -0.404$

$b_2 = 0.333$

```
OUTCOME VARIABLE:
 conflict

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .7658      .5864      .2444    89.3290     3.0000   189.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant     3.1114      .2459    12.6538      .0000     2.6264     3.5965
mentor        .0583      .0579     1.0070      .3152     -.0559      .1726
resource     -.4036      .0383   -10.5387      .0000     -.4792     -.3281
workload      .3332      .0405     8.2204      .0000      .2533      .4132
```

← c'

← b_1

← b_2

Controlling for extent of formal mentoring and access to resources, those who perceive greater workload feel greater work-family conflict.

Controlling for extent of formal mentoring and workload, those who perceive greater access to resources feel less work-family conflict.

The direct effect of extent of extent of formal mentoring on work-family conflict is not statistically significant

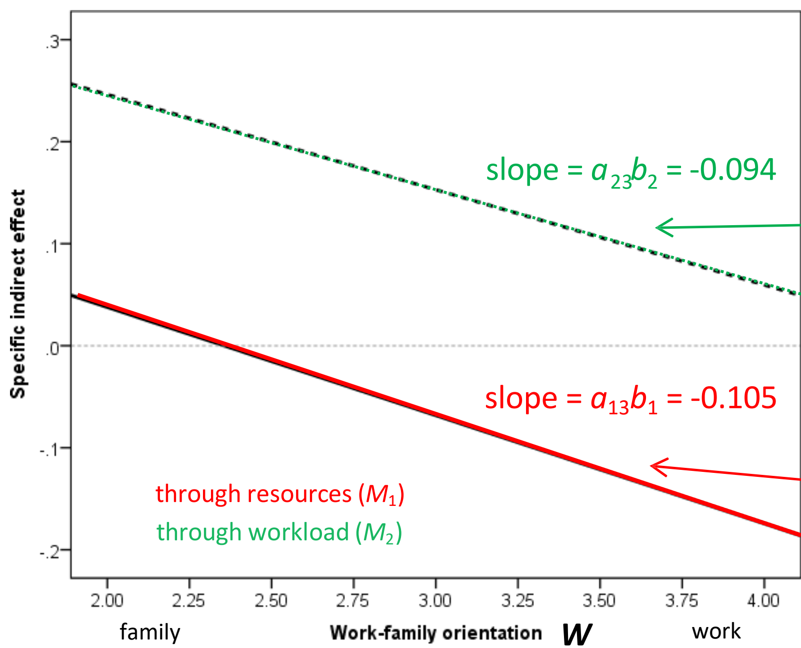
Conditional indirect effects

$\theta_{X\to M_2} = a_{21} + a_{23}W = 1.301 - 0.281W$

					through resources	through workload
W	$\theta_{X\to M_1}$	$\theta_{X\to M_2}$	b_1	b_2	$\theta_{X\to M_1} b_1$	$\theta_{X\to M_2} b_2$
2.30	-0.014	0.655	-0.404	0.333	0.006	0.218
3.10	0.195	0.431	-0.404	0.333	-0.079	0.144
3.80	0.378	0.234	-0.404	0.333	-0.153	0.078

$\theta_{X\to M_1} = a_{11} + a_{13}W = -0.614 + 0.261W$

Conditional specific indirect effects



The conditional indirect effects are a linear function of moderator W.

$$\theta_{X\to M_2} b_2 = (a_{21} + a_{23}W)b_2 = a_{21}b_2 + a_{23}b_2W = 0.433 - 0.094W$$

$$\theta_{X\to M_1} b_1 = (a_{11} + a_{13}W)b_1 = a_{11}b_1 + a_{13}b_1W = 0.248 - 0.105W$$

$a_{13}b_1$ and $a_{23}b_2$ are indices of moderated mediation. They quantify the relationship between W and the size of the specific indirect effects of X on Y through M_1 and M_2, respectively. An inference about these serves as a test of moderated mediation.

Tests of moderation of the indirect effects

```
****************** DIRECT AND INDIRECT EFFECTS OF X ON Y *****************

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .0583      .0579     1.0070      .3152     -.0559      .1726

Conditional indirect effects of X on Y:

INDIRECT EFFECT:
 mentor      ->    resource    ->    conflict

   wforient      Effect     BootSE   BootLLCI   BootULCI
     2.3000       .0055      .0473     -.0897      .0975
     3.1000      -.0787      .0363     -.1526     -.0068
     3.8000      -.1525      .0477     -.2484     -.0618

      Index of moderated mediation:
              Index     BootSE  BootLLCI  BootULCI
wforient      -.1053      .0410     -.1895     -.0284
---

INDIRECT EFFECT:
 mentor      ->    workload    ->    conflict

   wforient      Effect     BootSE   BootLLCI   BootULCI
     2.3000      .2183      .0495      .1340      .3287
     3.1000      .1434      .0355      .0795      .2171
     3.8000      .0779      .0373      .0086      .1565

      Index of moderated mediation:
              Index     BootSE    BootLLCI    BootULCI
wforient      -.0936      .0334     -.1680     -.0355
```

Direct effect

W

$a_{13}b_1$ = -0.105: the index of moderated mediation and 95% CI. The specific indirect effect through resource access is moderated by work-family orientation. It is larger (more negative among those more work oriented)

W

$a_{23}b_2$ = -0.094: the index of moderated mediation and 95% CI. The specific indirect effect through workload is moderated by work-family orientation. It is larger (more positive) among those more family oriented.

With evidence of moderation, we can probe

```
****************** DIRECT AND INDIRECT EFFECTS OF X ON Y *****************

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .0583      .0579     1.0070      .3152     -.0559      .1726

Conditional indirect effects of X on Y:

INDIRECT EFFECT:
 mentor      ->     resource      ->     conflict

  wforient       Effect     BootSE   BootLLCI   BootULCI
    2.3000        .0055      .0473     -.0897      .0975
    3.1000       -.0787      .0363     -.1526     -.0068
    3.8000       -.1525      .0477     -.2484     -.0618

      Index of moderated mediation:
                Index      BootSE   BootLLCI    BootULCI
wforient       -.1053       .0410    -.1895      -.0284
---

INDIRECT EFFECT:
 mentor      ->     workload      ->     conflict

   wforient      Effect     BootSE   BootLLCI   BootULCI
    2.3000       .2183      .0495     .1340      .3287
    3.1000       .1434      .0355     .0795      .2171
    3.8000        .0779      .0373     .0086      .1565

      Index of moderated   mediation:
                Index      BootSE    BootLLCI  BootULCI
wforient       -.0936       .0334         -.1680-.0355
```

Direct effect

W

$$\theta_{X\to M_1}b_1 = (a_{11} + a_{13}W)b_1$$
$$= a_{11}b_1 + a_{13}b_1W$$
$$= 0.248 - 0.105W$$

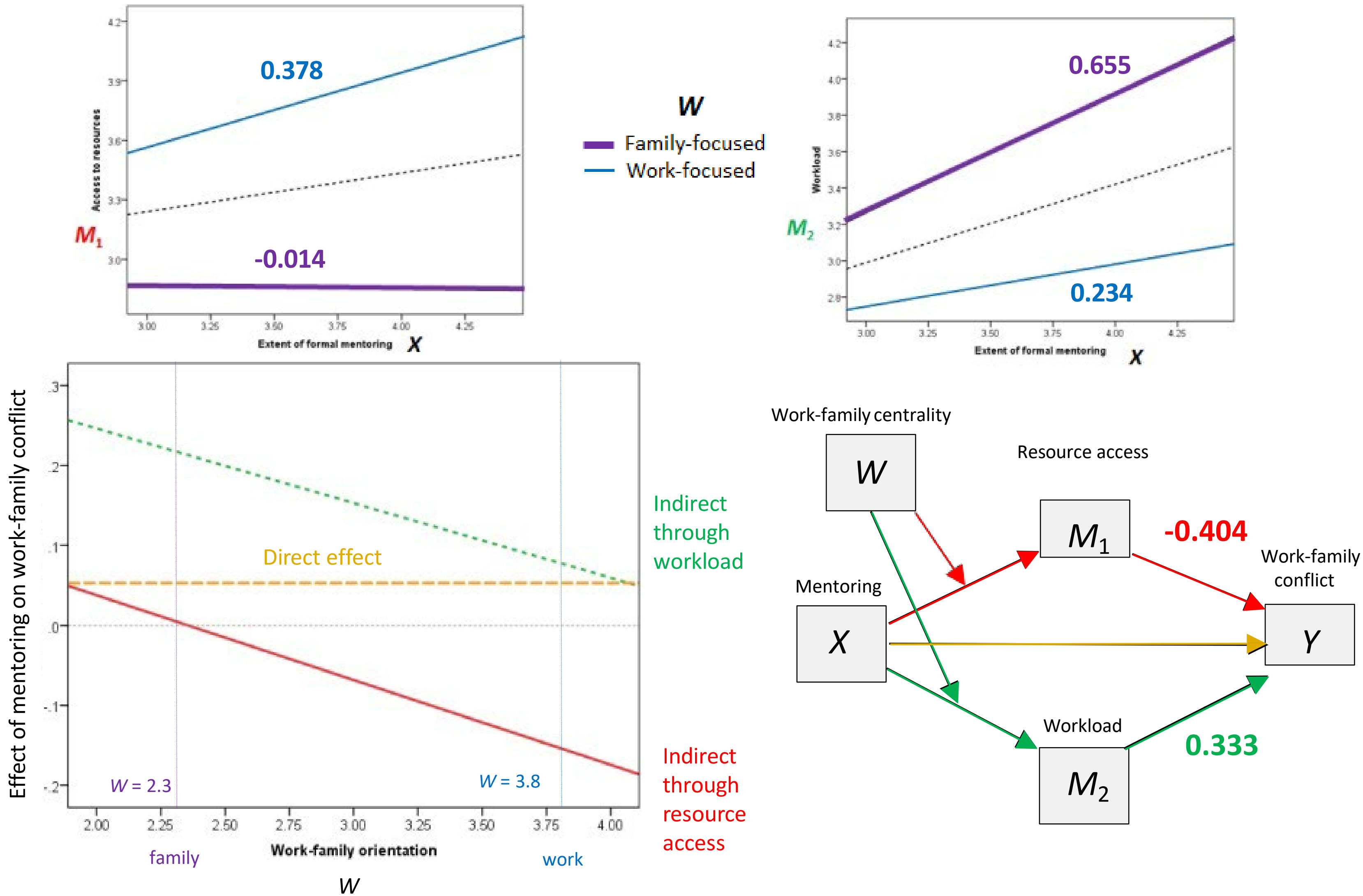
with 95% bootstrap CIs

W

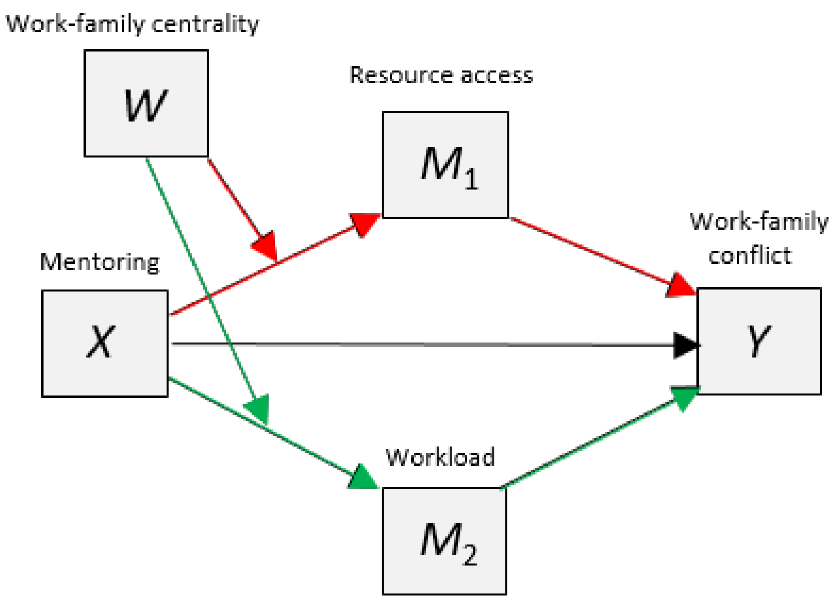
$$\theta_{X\to M_2}b_2 = (a_{21} + a_{23}W)b_2$$
$$= a_{21}b_2 + a_{23}b_2W$$
$$= 0.433 - 0.094W$$

with 95% bootstrap CIs

Summary


0.378
-0.014
Access to resources
M_1
Extent of formal mentoring
X
W
Family-focused
Work-focused
0.655
0.234
Workload
M_2
Extent of formal mentoring
X
Effect of mentoring on work-family conflict
Indirect through workload
Direct effect
Indirect through resource access
W = 2.3
W = 3.8
Work-family orientation
family
work
W
Work-family centrality
W
Resource access
M_1
-0.404
Work-family conflict
Mentoring
X
Y
Workload
0.333
M_2

Summary



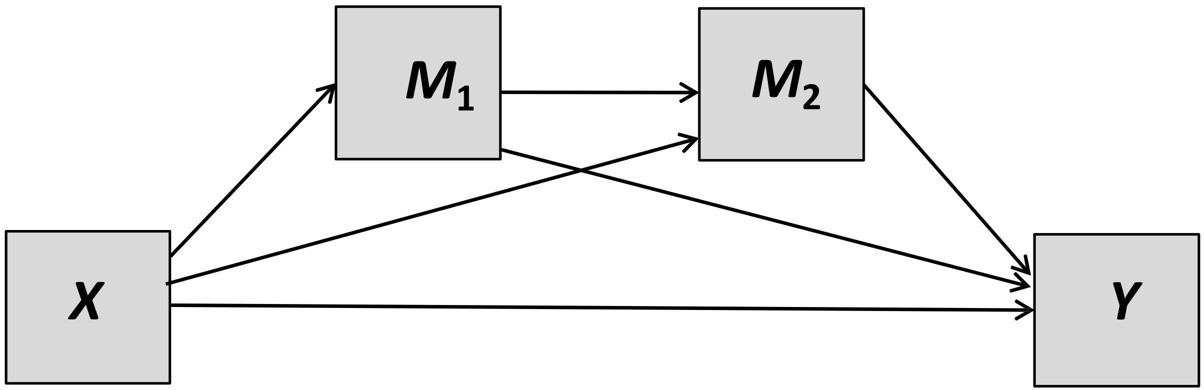
- The effect of formal mentoring on work-family conflict operates indirectly (negatively) through access to resources among the more work-focused employees, but not among the more family-focused employees.

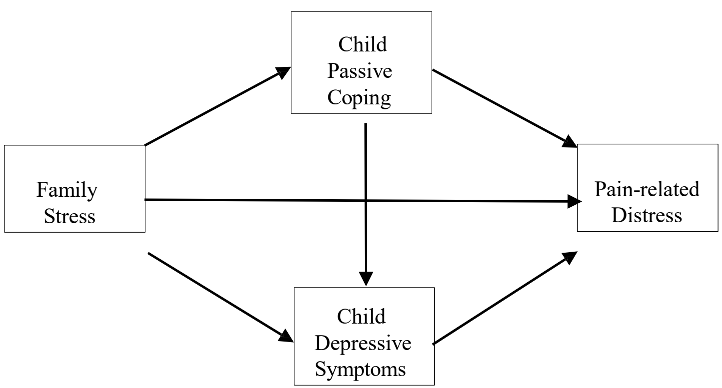
- The effect of formal mentoring on work-family conflict operates indirectly (positively) through feelings of greater workload moreso among the more family-focused employees.

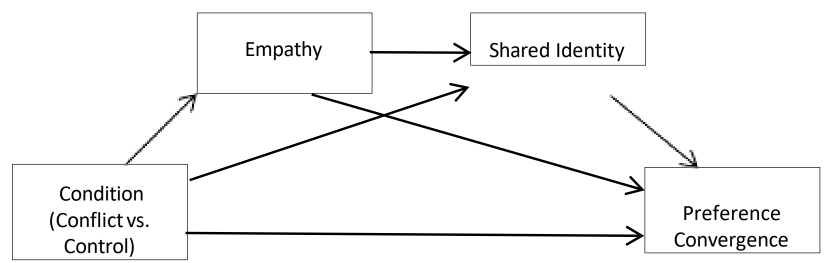
Moderation in Serial Mediation

The serial multiple mediator model

This is a serial multiple mediator model with two mediators and all possible direct and indirect effects freely estimated.



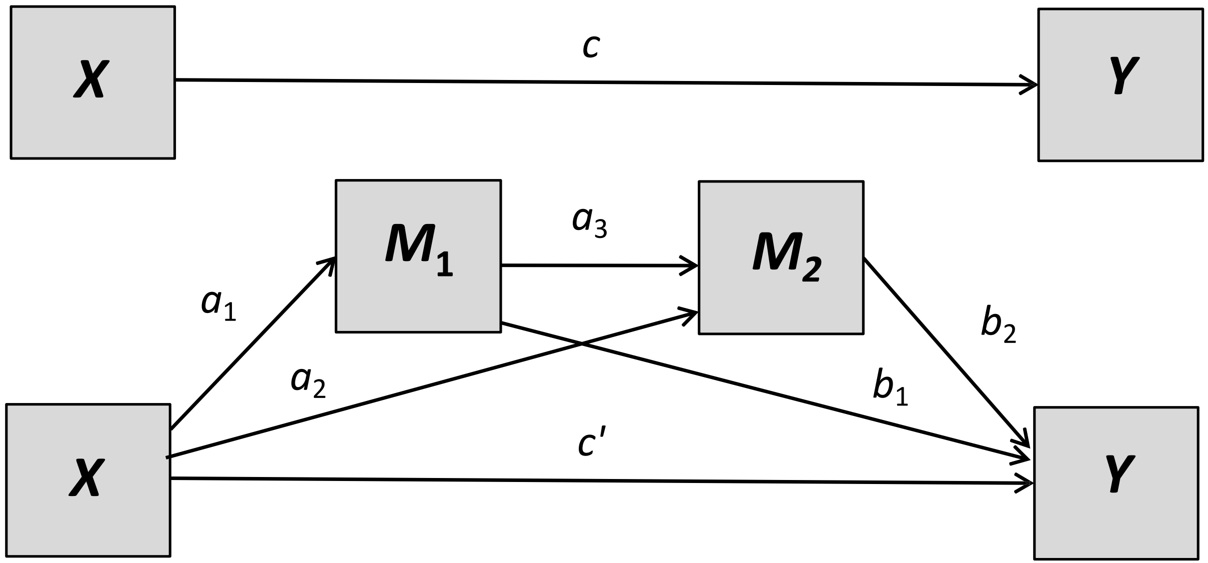




Reed-Knight, B., van Tilburg, P. A. L., Levy, R. L. et al. (2018). Maladaptive coping and depressive symptoms partially explain the association between family stress and pain-related distress in youth with IBD. ***Journal of Pediatric Psychology, 43***, 94-103.

Schrift, R. Y., & Moty, A. (2015). Pain and preferences: Observed Decisional conflict and the convergence of preferences. ***Journal of Consumer Research, 42***, 515-534..

Serial mediation: Path analysis rules



$$\hat{Y}_i = c_0 + cX_i$$
$$\widehat{M_{1i}} = a_{01} + a_1X_i$$
$$\widehat{M_{2i}} = a_{02} + a_2X_i + a_3M_{1i}$$
$$\hat{Y}_i = c'_0 + c'X_i + b_1M_{1i} + b_2M_{2i}$$

Direct effect of X: c'
Specific indirect effect of X through M_1: a_1b_1
Specific indirect effect of X through M_2: a_2b_2
Specific indirect effect of X through M_1 and M_2: $a_1a_3b_2$
Total indirect effect of X: $a_1b_1 + a_2b_2 + a_1a_3b_2$
Total effect of X: $c = c' + a_1b_1 + a_2b_2 + a_1a_3b_2$

Example

May 1st, 2011, 11:30PM

Professor Erik Nisbet (OSU School of Communication) had an national telephone survey in the field examining perceptions of Muslims in the U.S. when Obama announces the death of bin Laden; 390 respondents prior to announcement (BINLADEN = 0) and 271 after announcement (BINLADEN = 1).

Measures

STEREO: Stereotype endorsement, 4 items (5-pt semantic differential)

"Please tell us how much you associate each of the following sets of characteristics with Muslims"

e.g., Peaceful – Violent
Tolerant – Fanatical

RTHREAT: Realistic threat, 5 items (5-pt Likert)

"Below are a few statements expressing different views about Muslims living in the U.S. Please read and tell us how much you agree with each statement"

e.g., "Muslims in the U.S. sympathize with terrorists"
"Muslims make America a more dangerous place to live"

MCIVIL: Restriction of Muslim civil liberties, 5 items (5-pt Likert)

"Below are some statements people have expressed about Muslim civil liberties and terrorism in the U.S. Please read each and tell us how much you agree or disagree…"

e.g., "All Muslims in the U.S. should be required to carry a special ID card"
"Muslims in the U.S. should register their whereabouts with the U.S. government"

Higher reflects greater negative stereotype endorsement/threat/willingness to restrict…"

The data: binladen

binladen.sav

binladen.sas

*binladen.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	binladen	rthreat	stereo	mcivil	age	ideo	sex
1	0	3.00	2.80	2.80	7.5	8	0
2	1	2.00	1.80	3.20	3.3	4	1
3	1	2.25	2.00	2.80	5.6	6	0
4	1	2.00	2.60	3.40	4.0	5	1
5	1	4.00	4.20	4.00	5.9	8	1
6	0	1.00	1.40	4.80	7.9	6	0
7	0	4.00	5.00	5.00	5.8	5	0
8	0	4.00	2.80	3.80	7.2	6	1
9	0	3.25	3.00	2.80	5.7	4	1
10	1	4.00	5.00	5.00	5.9	5	1
11	0	2.50	2.20	2.60	5.0	7	1
12	1	1.50	3.20	1.60	3.9	4	1
13	1	3.25	3.00	3.40	5.6	9	0
14	1	4.25	2.80	5.00	4.6	4	1
15	1	2.50	3.80	3.20	2.6	5	0
16	1	2.75	3.40	3.20	5.6	8	1

binladen

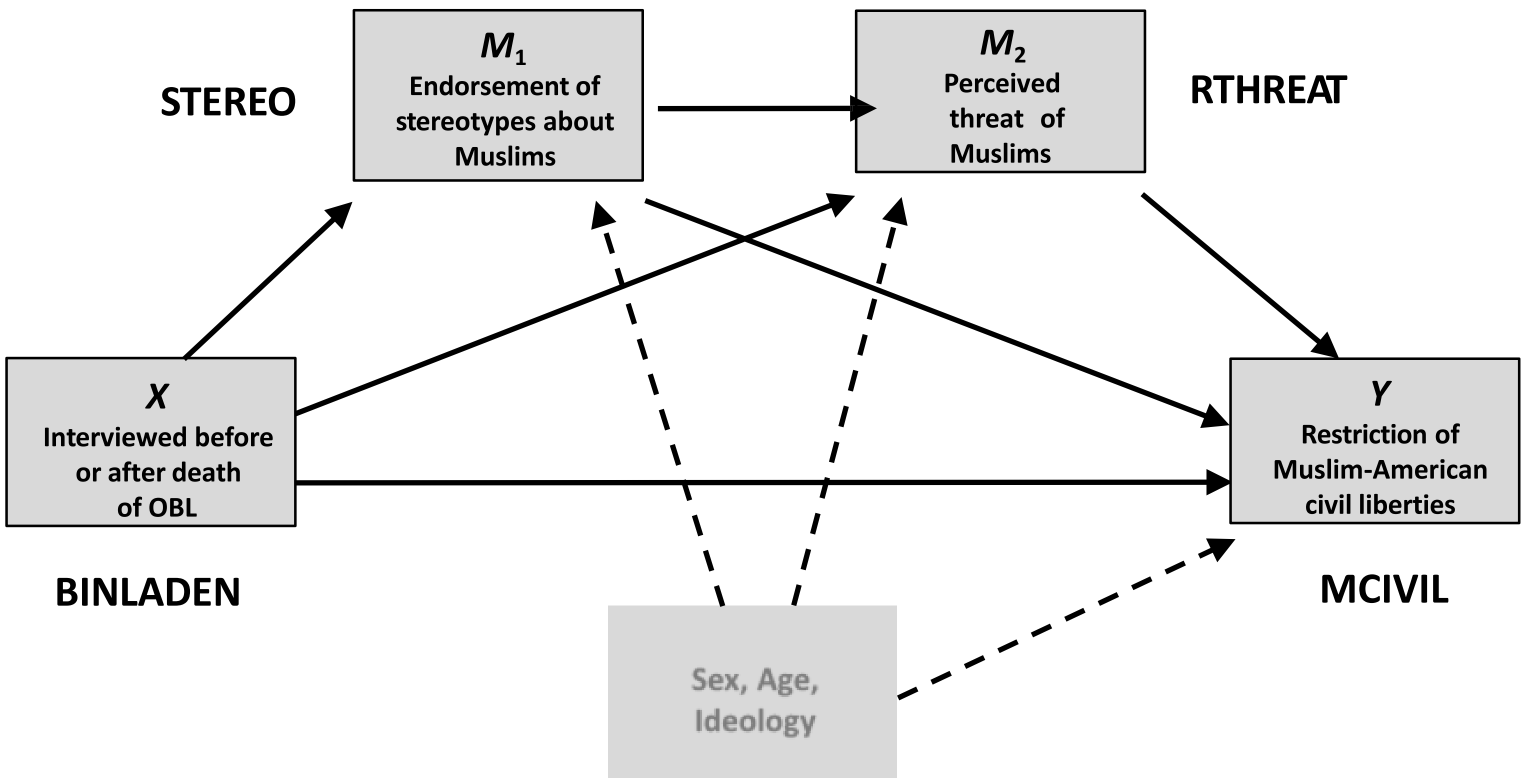
```
data binladen;
input binladen rthreat stereo mcivil age ideo sex;
datalines;
0   3.00   2.80   2.80   7.5 8   0
1   2.00   1.80   3.20   3.3 4   1
1   2.25   2.00   2.80   5.6 6   0
1   2.00   2.60   3.40   4.0 5   1
1   4.00   4.20   4.00   5.9 8   1
0   1.00   1.40   4.80   7.9 6   0
0   4.00   5.00   5.00   5.8 5   0
0   4.00   2.80   3.80   7.2 6   1
0   3.25   3.00   2.80   5.7 4   1
1   4.00   5.00   5.00   5.9 5   1
0   2.50   2.20   2.60   5.0 7   1
1   1.50   3.20   1.60   3.9 4   1
1   3.25   3.00   3.40   5.6 9   0
1   4.25   2.80   5.00   4.6 4   1
1   2.50   3.80   3.20   2.6 5   0
1   2.75   3.40   3.20   5.6 8   1
```

In R: Don't forget to change the path below to where your **binladen.csv** file is located.

```
binladen<-read.table("c:/mmcpa/binladen.csv", sep=",",header=TRUE)
head(binladen)
```

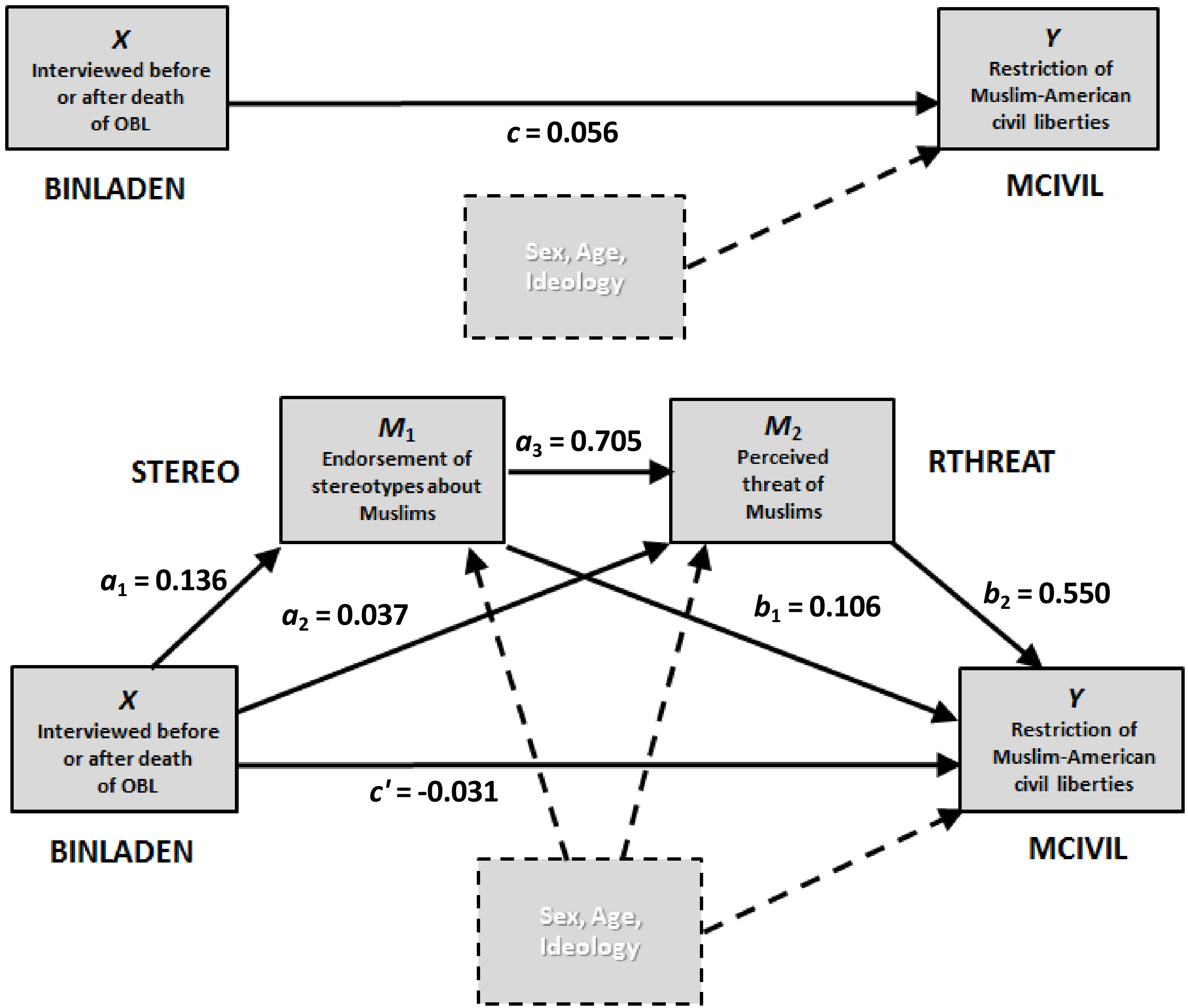
```
> binladen<-read.table("c:/mmcpa/binladen.csv", sep=",",header=TRUE)
> head(binladen)
  binladen rthreat stereo mcivil age ideo sex
1        0    3.00    2.8    2.8 7.5    8   0
2        1    2.00    1.8    3.2 3.3    4   1
3        1    2.25    2.0    2.8 5.6    6   0
4        1    2.00    2.6    3.4 4.0    5   1
```

A serial mediation model with no moderation



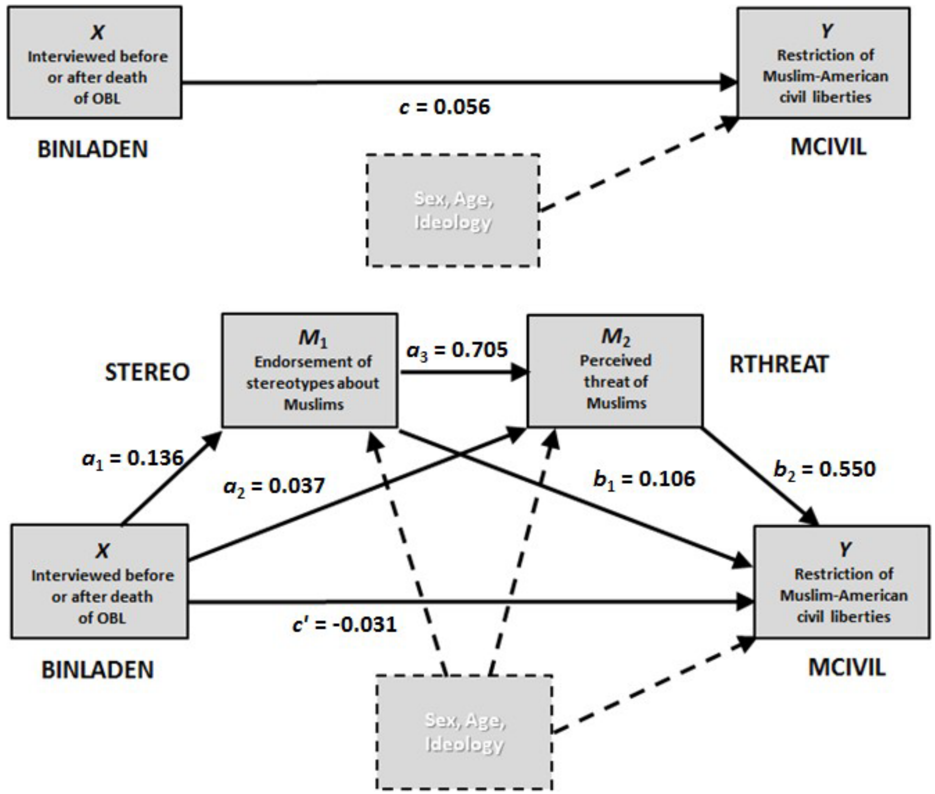
This model includes four pathways of influence of news coverage of OBL death, two through a single mediator, one through both mediators in serial, and one direct.

Example



X
Interviewed before
or after death
of OBL
BINLADEN
c = 0.056
Y
Restriction of
Muslim-American
civil liberties
MCIVIL
Sex, Age,
Ideology
STEREO
M1
Endorsement of
stereotypes about
Muslims
a3 = 0.705
M2
Perceived
threat of
Muslims
RTHREAT
a1 = 0.136
a2 = 0.037
b1 = 0.106
b2 = 0.550
X
Interviewed before
or after death
of OBL
c' = -0.031
BINLADEN
Y
Restriction of
Muslim-American
civil liberties
MCIVIL
Sex, Age,
Ideology

Example



Direct effect = -0.031

Specific indirect effect via stereotype endorsement: 0.136(0.106) = 0.014

Specific indirect effect via perceived threat: 0.037(0.550) = 0.020

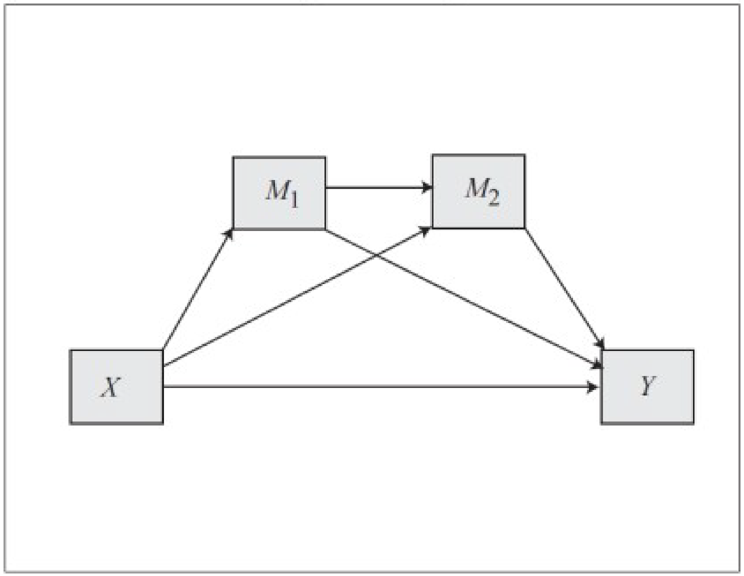
Specific indirect effect via stereotype endorsement and threat: 0.136(0.705)(0.550) = 0.053

Total indirect effect = 0.014 + 0.020 + 0.053 = 0.087

Total effect = -0.031 + 0.087 = 0.056

Estimation and inference using PROCESS

Model 6
(2 mediators)



PROCESS model 6 is the serial multiple mediator model.

In model 6, order of the variables in the "m=" list matters. Variables listed earlier are causally prior to those listed later. PROCESS allows up to six mediators to be linked in a causal chain. All possible indirect and direct effects are estimated.

```
process y=mcivil/x=binladen/m=stereo rthreat/cov=sex age ideo/model=6
  /total=1/seed=6234.
```

```
%process (data=binladen,y=mcivil,x=binladen,m=stereo rthreat,cov=sex age
  ideo,model=6,total=1,seed=6234)
```

```
process(data=binladen,y="mcivil",x="binladen",m=c("stereo","rthreat"),
  cov=c("sex","age","ideo"),model=6,total=1,seed=6234)
```

PROCESS output

```
Model  : 6
    Y  : mcivil
    X  : binladen
   M1  : stereo
   M2  : rthreat

Covariates:
 sex       age       ideo

Sample
Size:  661

Custom
Seed:      6234
```

$$\widehat{M_{1i}} = 1.90 + 0.14X_i + 0.04U_{1i} + 0.05U_{2i} + 0.13U_{3i}$$

```
**************************************************************************
OUTCOME VARIABLE:
 stereo
```

a_1 path

```
Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .3557      .1265      .6495    23.7609     4.0000   656.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant     1.9045      .1322    14.4084      .0000     1.6449     2.1640
binladen      .1358      .0639     2.1258      .0339      .0104      .2613
sex           .0398      .0635      .6262      .5314     -.0849      .1644
age           .0504      .0192     2.6220      .0089      .0127      .0882
ideo          .1293      .0143     9.0483      .0000      .1012      .1574
**************************************************************************
```

PROCESS output

$$\widehat{M_{2i}} = -0.25 + 0.04X_i + 0.70M_{1i} + 0.13U_{1i} + 0.05U_{2i} + 0.09U_{3i}$$

```
**************************************************************************
OUTCOME VARIABLE:
 rthreat

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .6764      .4575      .6076   110.4916     5.0000   655.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
Constant      -.2548      .1467    -1.7369      .0829     -.5428      .0332
binladen       .0374      .0620      .6038      .5462     -.0843      .1592
stereo         .7047      .0378    18.6630      .0000      .6306      .7789
sex            .1286      .0614     2.0938      .0367      .0080      .2492
age            .0451      .0187     2.4135      .0161      .0084      .0818
ideo           .0898      .0147     6.1257      .0000      .0610      .1186
**************************************************************************
```

a_2 path

a_3 path

Practice writing an interpretation for a_2 and a_3

PROCESS output

$$\hat{Y}_i = 0.72 - 0.03X_i + 0.11M_{1i} + 0.55M_{2i} - 0.10U_{1i} - 0.01U_{2i} + 0.05U_{3i}$$

```
***************************************************************************
OUTCOME VARIABLE:
 mcivil

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .6727      .4526      .5890    90.1100     6.0000   654.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant      .7165      .1448     4.9499      .0000      .4323     1.0008
binladen     -.0311      .0611     -.5095      .6106     -.1510      .0888
Stereo        .1057      .0460     2.2965      .0220      .0153      .1960
rthreat       .5491      .0385    14.2732      .0000      .4736      .6247
sex          -.1001      .0607    -1.6504      .0993     -.2193      .0190
age          -.0103      .0185     -.5599      .5758     -.0466      .0259
ideo          .0545      .0148     3.6696      .0003      .0253      .0836
***************************************************************************
```

c' path

b_1 path

b_2 path

Practice writing an interpretation for b_1 and b_2

PROCESS output

$$\widehat{Y}_i = 1.51 + 0.06X_i - 0.01U_{1i} + 0.04U_{2i} + 0.17U_{3i}$$

```
************************* TOTAL EFFECT MODEL ***************************
OUTCOME VARIABLE:
 mcivil

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .3675      .1351      .9278    25.6100     4.0000   656.0000      .0000

Model
                coeff         se          t          p       LLCI       ULCI
constant       1.5149      .1580     9.5894      .0000     1.2047     1.8251
binladen        .0564      .0764      .7380      .4608     -.0936      .2063
sex            -.0099      .0759     -.1310      .8958     -.1589      .1391
age             .0393      .0230     1.7085      .0880     -.0059      .0844
ideo            .1675      .0171     9.8053      .0000      .1339      .2010
************************************************************************
```

c path

PROCESS output

```
************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Total effect of X on Y
     Effect           se            t            p         LLCI         ULCI
      .0564        .0764        .7380        .4608       -.0936        .2063

Direct effect of X on Y
     Effect           se            t            p         LLCI         ULCI
     -.0311        .0611       -.5095        .6106       -.1510        .0888

Indirect effect(s) of X on Y:
           Effect     BootSE   BootLLCI   BootULCI
TOTAL       .0875      .0462     -.0017      .1812
Ind1        .0144      .0100     -.0004      .0376
Ind2        .0206      .0331     -.0428      .0882
Ind3        .0526      .0251      .0039      .1015

Indirect effect key:
Ind1  binladen    ->     stereo      ->    mcivil
Ind2  binladen    ->     rthreat     ->    mcivil
Ind3 binladen     ->     stereo      ->    rthreat      ->    mcivil
```

c path

c' path

a_1b_1

a_2b_2

$a_1a_3b_2$

and bootstrap CIs

The data are consistent with the claim that after coverage of OBL's death endorsement of restriction of Muslim civil liberties is higher serially through the effect of stereotype endorsement and perceived threat of Muslims (0.053, 95% CI=0.004 to 0.102) but not through perceived threat independent of stereotype endorsement (0.021, 95% CI = -0.043 to 0.088) or through stereotype endorsement independent of perceived threat (0.014, 95% CI = -0.004 to 0.038) after controlling for age, sex, and ideology. There is no significant direct effect of coverage of his death.

Causality in Serial Mediation

In a simple mediation model with no random assignment there are 6 possible orderings of the variables in the model.

X → M → Y, X → Y → M, M → X → Y, M → Y → X, Y → X → M, Y → M → X

With randomization of X, you can eliminate many of these, and are left with 2 possible orders. But also any of these could be confounded as well.

X → M → Y, X → Y → M,

Serial mediation with 2 mediators and no random assignment there are $6 \times 4 = 24$ possible orderings of the variables in the model.

X → M1 → M2 → Y, X → M1 → Y → M2, X → M2 → M1 → Y, X → M2 → Y → M1, X → Y → M1 → M2, X → Y → M2 → M1
M1 → M2 → X → Y, M1 → M2 → Y → X, M1 → X → M2 → Y, M1 → X → Y → M2, M1 → Y → M2 → X, M1 → Y → X → M2
M2 → M1 → X → Y, M2 → M1 → Y → X, M2 → X → M1 → Y, M2 → X → Y → M1, M2 → Y → M1 → X, M2 → Y → X → M1
Y → X → M1 → M2, Y → X→ M2 → M1, Y → M1 → M2 → X, Y → M1 → X → M2, Y → M2 → M1 → X, Y → M2 → X → M1

With randomization of X you limit to 6 but also confounding. This number continues to increase as you add mediators.

Customizing the assignment of covariates

By default, all covariates are assigned to all equations, but this can be overridden with the **cmatrix** option, new in PROCESS v3. The assignment of covariates to equations is internally represented with the *C* matrix.

```
PROCESS y=yvar/x=xvar/m=med1 med2 ... medk/cov=cov1 cov2 … covk/model= ...
```

	cov1	cov2	…	covk
med1	0/1	0/1	…	0/1
med2	0/1	0/1	…	0/1
.	.	.	…	.
.	.	.	…	.
.	.	.	…	.
medk	0/1	0/1	…	0/1
yvar	0/1	0/1	…	0/1

C matrix

A one in the cell means the covariate in that column is to be included in the model of the variable in that row. A zero means the covariate in that column is to be excluded from the model of the variable in that row. By default, all entries in the *C* matrix are set to one.

Customizing the assignment of covariates

For example, from the analysis just completed, the regression coefficients for sex in the model of stereotype endorsement and civil liberties restrictions are nonsignificant. And age is nonsignificantly related to civil liberties restrictions. If we wanted to, we could fix these partial regression weights to zero by excluding them from the corresponding equations.

C matrix

	sex	age	ideo
stereo	0	1	1
rthreat	1	1	1
mcivil	0	0	1

The **cmatrix** option does the assignment of covariates to equations. Read the matrix left to right, top to bottom, assigning the zeros and ones. Separate by commas in SPSS and R.

```
process y=mcivil/x=binladen/m=stereo rthreat/cov=sex age ideo/model=6
 /seed=6234/cmatrix=0,1,1,1,1,1,0,0,1.
```

```
%process (data=binladen,y=mcivil,x=binladen,m=stereo rthreat,  cov=sex age
  ideo,model=6,seed=6234,cmatrix=0 1 1 1 1 1 0 0 1);
```

```
process(data=binladen,y="mcivil",x="binladen",m=c("stereo","rthreat"),
  cov=c("sex","age","ideo"),model=6,seed=6234,cmatrix=c(0,1,1,1,1,1,0,0,1))
```

Try This Model

```
process y=mcivil/x=binladen/m=stereo rthreat/cov=sex age ideo/model=6
 /seed=6234/cmatrix=0,1,1,1,1,1,0,0,1.
```

```
%process (data=binladen,y=mcivil,x=binladen,m=stereo rthreat,  cov=sex age
  ideo,model=6,seed=6234,cmatrix=0 1 1 1 1 1 0 0 1);
```

```
process(data=binladen,y="mcivil",x="binladen",m=c("stereo","rthreat"),
  cov=c("sex","age","ideo"),model=6,seed=6234,cmatrix=c(0,1,1,1,1,1,0,0,1))
```

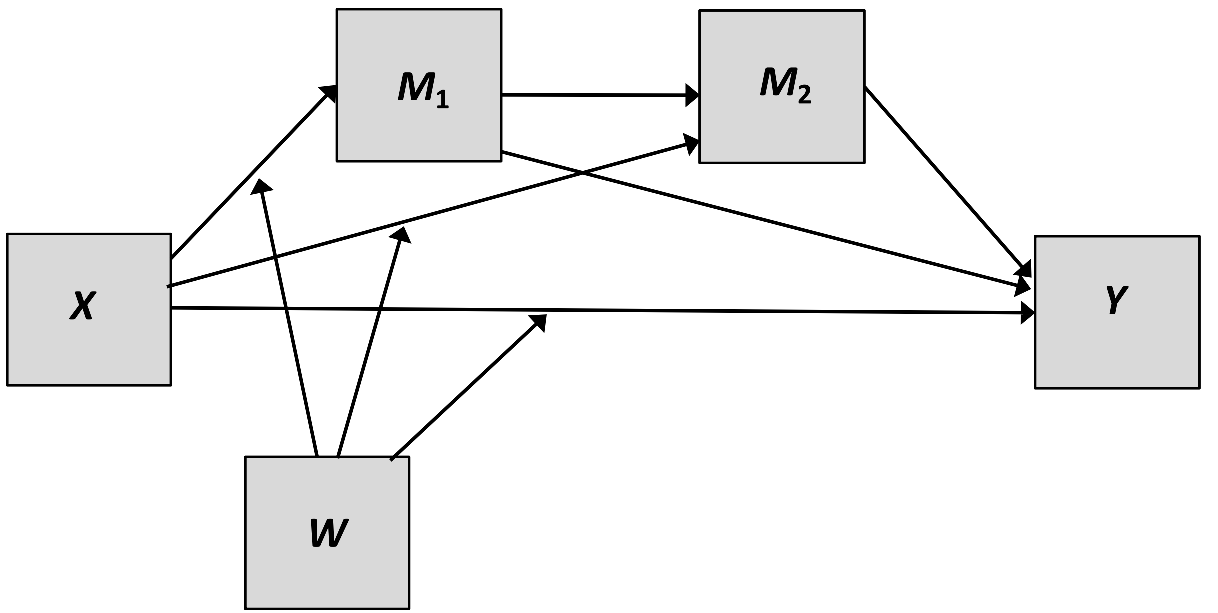
Look at whether the total effect = direct + (total indirect effect).

Imagine we think that age is a confounder of the effect of stereotype endorsement on realistic threat but not restrictions on civil liberties. What are the pros and cons of not controlling for age in the model of mcivil?

Combining moderation with serial mediation

Serial mediation and moderation can easily be integrated into a conditional process model. There are many permutations possible.

Example:

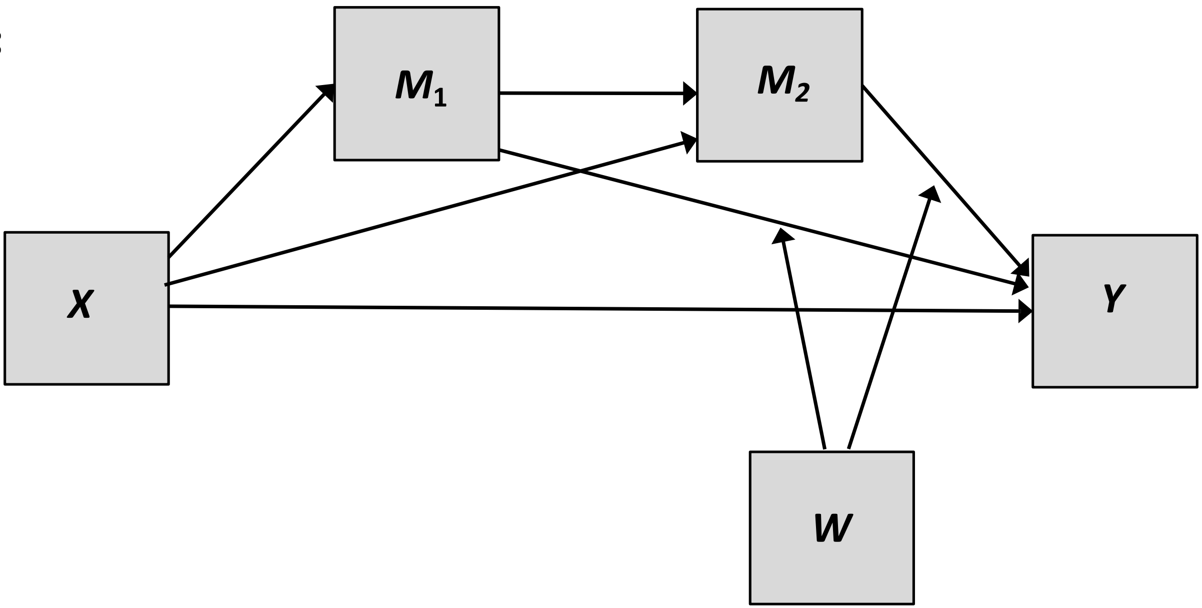


A serial multiple mediator model with moderation of the effect of X on all variable's causally later (M_1, M_2, and Y) by a common moderator W.

Combining moderation with serial mediation

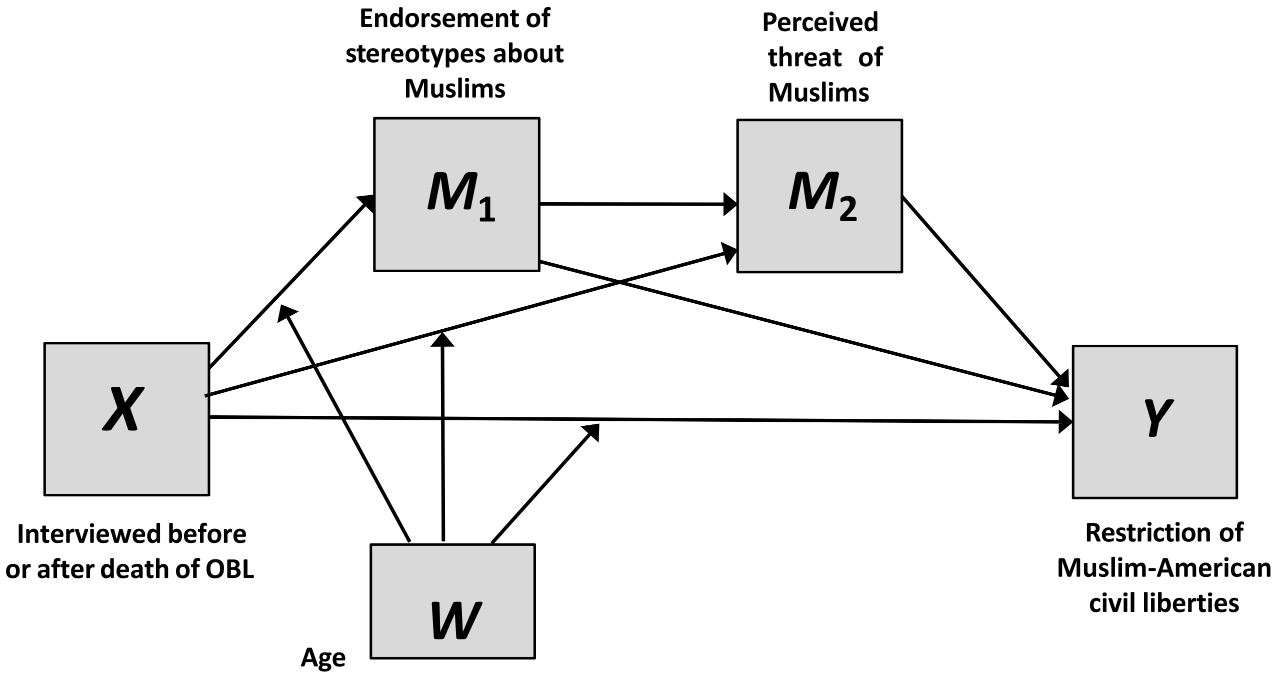
A serial multiple mediator model with moderation of the effects of M_1 and M_2 on Y by a common moderator W.

Example:

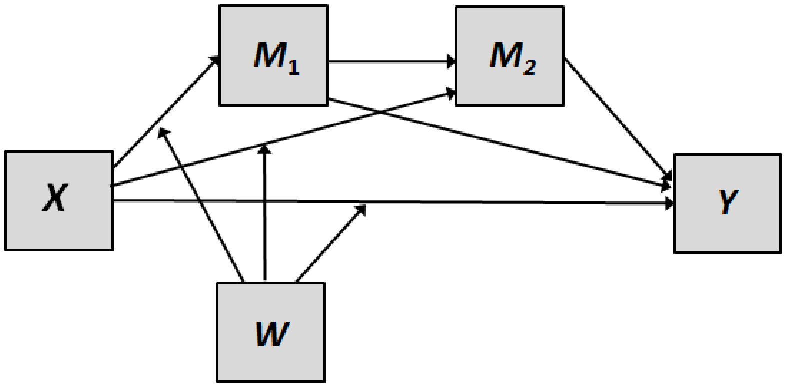


There is very little in the literature about the estimation and interpretation of models that combine moderation with serial mediation. PROCESS v2 could not estimate models such as this. PROCESS v3+ can. It has many preprogrammed versions, and later you'll learn how to create your own models.

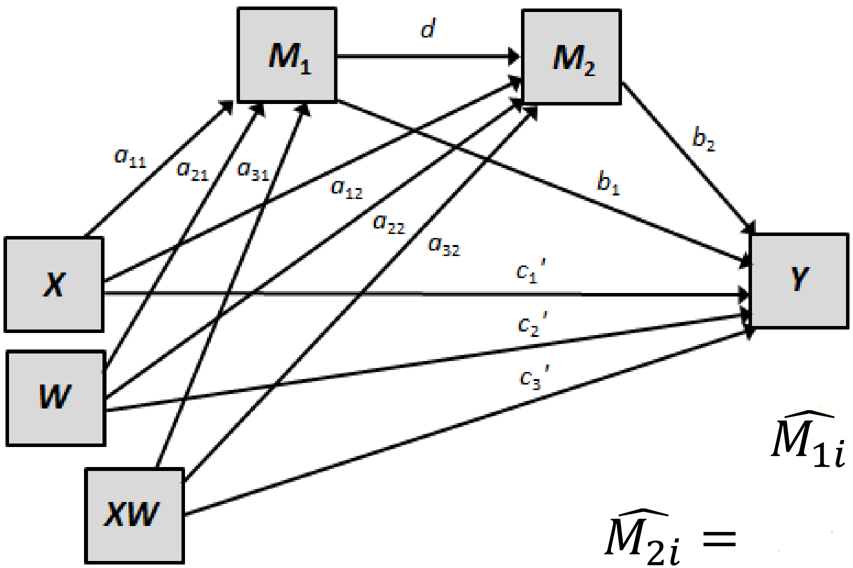
Moderated serial mediation



Let's allow the effect of coverage of bin Laden's death (*X*) on negative stereotype endorsement (M_1), perceived threat (M_2), and willingness to restrict Muslim-American civil liberties (*Y*) to vary with age. This allows for moderation of all three indirect effects, as well as the direct effect of *X*. We will include sex and ideology as covariates in all equations. PROCESS v3 can do this, and many other forms of moderated serial mediation.



Conceptual Model



Statistical Model

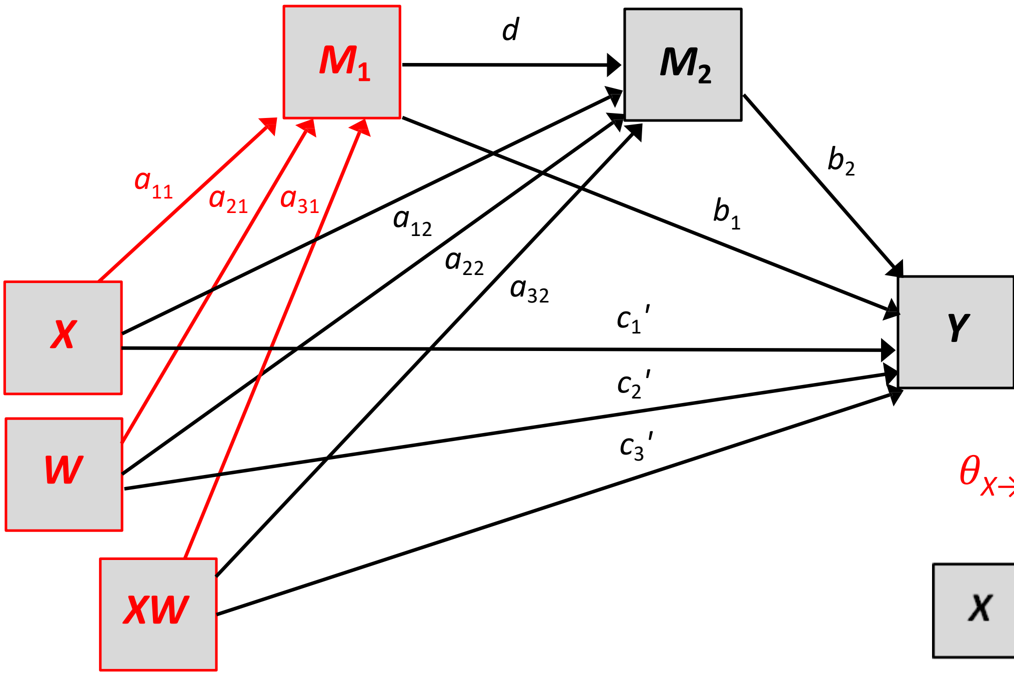
Write out the equations for the model:

$$\widehat{M_{1i}} =$$

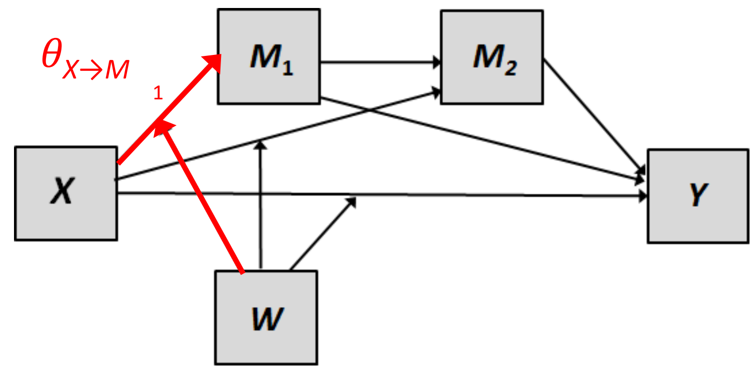
$$\widehat{M_{2i}} =$$

$$\widehat{Y_i} =$$

The model in statistical form



$\theta_{X \to M_1} = a_{11} + a_{31}W$

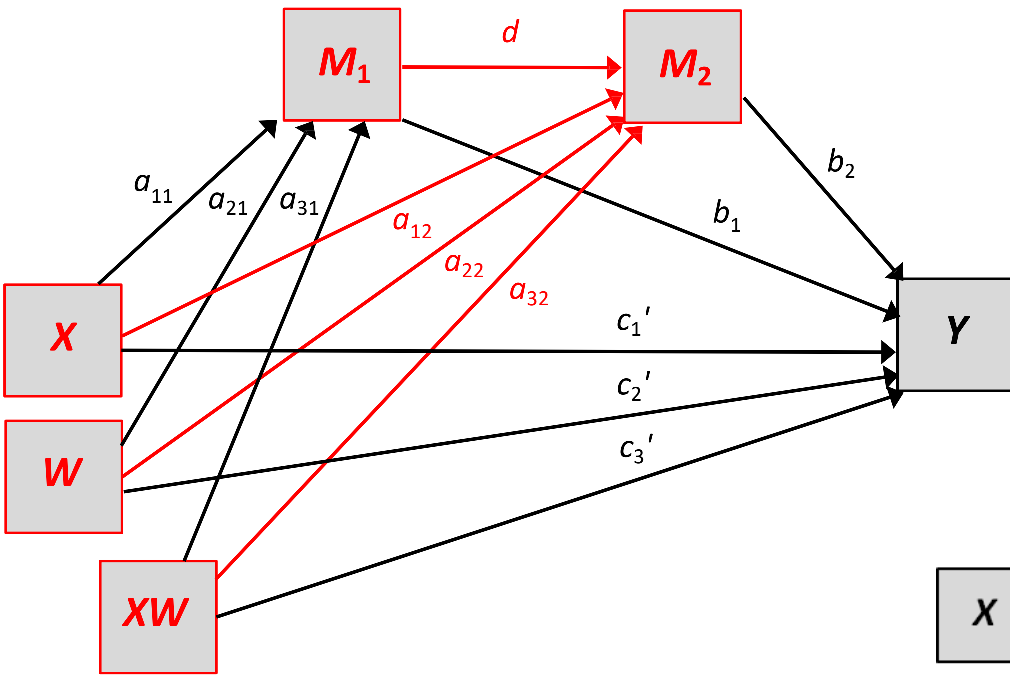


$$\widehat{M_{1i}} = a_{01} + a_{11}X_i + a_{21}W_i + a_{31}X_iW_i$$

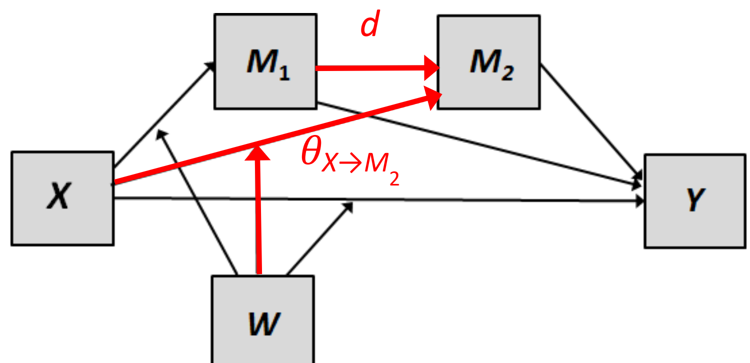
can be written as $\widehat{M_{1i}} = a_{01} + (a_{11}+a_{31}W_i)X_i + a_{21}W_i$ or, alternatively, as

$\widehat{M_{1i}} = a_{01} + \theta_{X\to M_1}X_i + a_{21}W_i$ where $\theta_{X\to M_1} = (a_{11} + a_{31}\ W)$ is the conditional effect of X on M_1. X's effect on M_1 is a function of W.

The model in statistical form



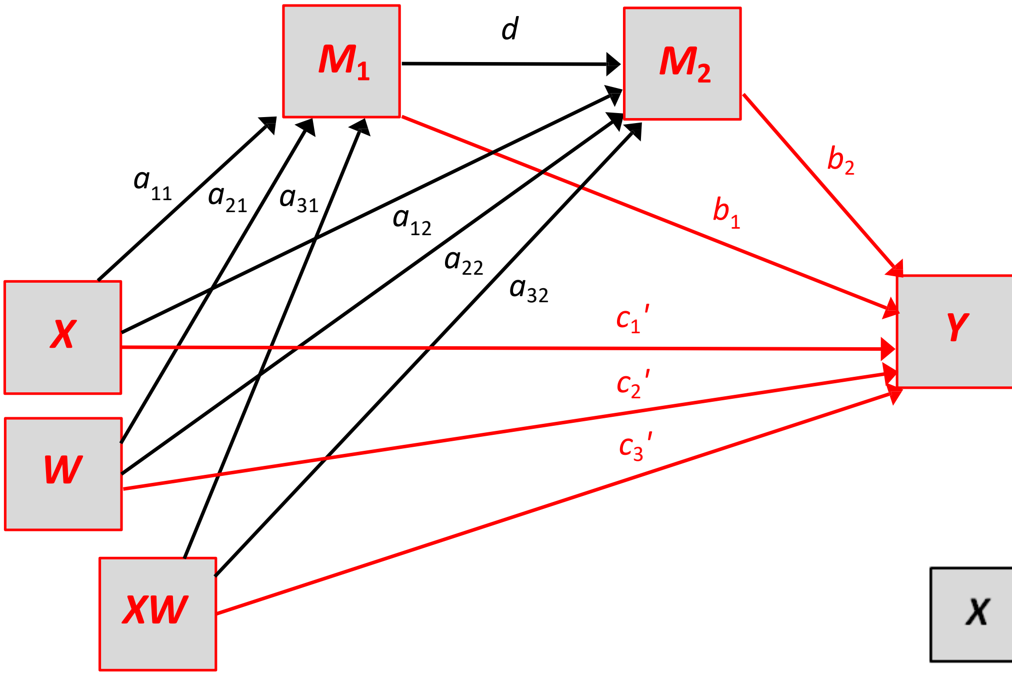
$\theta_{X \rightarrow M_2} = a_{12} + a_{32}W$



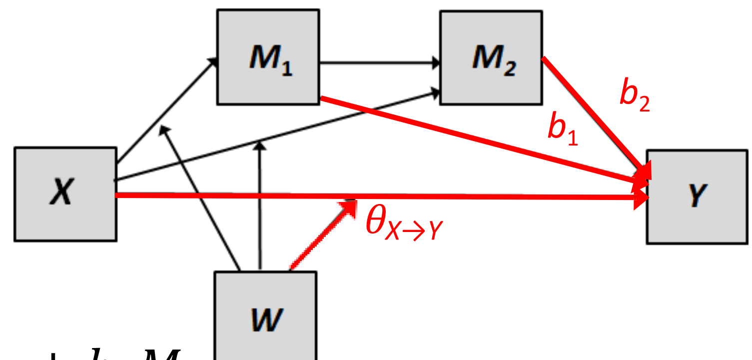
$$\widehat{M_{2i}} = a_{02} + a_{12}X_i + a_{22}W_i + a_{32}X_iW_i + dM_{1i}$$

can be written as $\widehat{M_{2i}} = a_{02} + (a_{12}+a_{32}W_i)X_i + a_{22}W_i + dM_{1i}$ or, alternatively, as $\widehat{M_{2i}} = a_{02} + \theta_{X \rightarrow M_2}W_i + a_{22}W_i + dM_{1i}$ where $\theta_{X \rightarrow M_2} = (a_{12} + a_{32}\, W)$ is the conditional effect of X on M_2. X's effect on M_2 independent of M_1 is a function of W.

The model in statistical form



$\theta_{X \to Y} = c'_1 + c'_3 W$

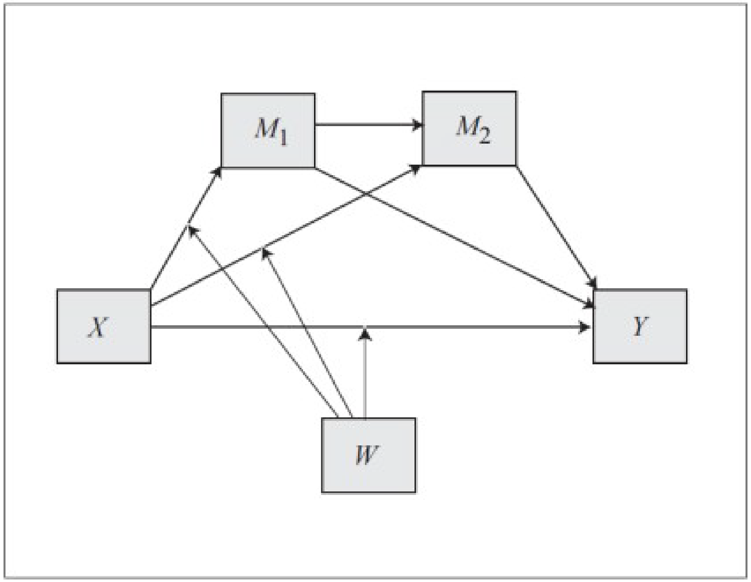


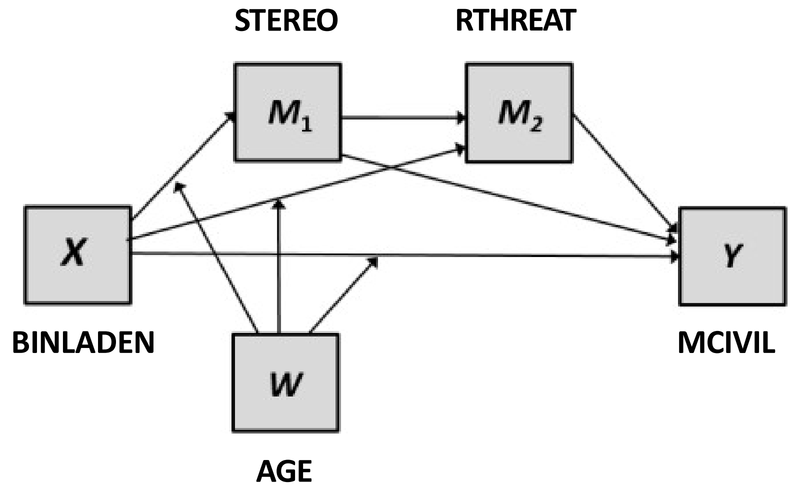
$$\widehat{Y}_i = c'_0 + c'_1 X_i + c'_2 W_i + c'_3 X_i W_i + b_1 M_{1i} + b_2 M_{2i}$$

can be written as $\widehat{Y}_i = c'_0 + (c'_1 + c'_3 W_i) X_i + c'_2 W_i + b_1 M_{1i} + b_2 M_{2i}$ or, alternatively, as $\widehat{Y}_i = c'_0 + \theta_{X \to Y} X_i + c'_2 W_i + b_1 M_{1i} + b_2 M_{2i}$ where $\theta_{X \to Y} = c'_1 + c'_3 W_i$ is the conditional direct effect of *X* on *Y*. *X*'s effect on *Y* independent of M_1 and M_2 is a function of *W*.

Estimation using PROCESS

Model 85





Note: We are including sex and ideology as covariates. They are not depicted here to reduce visual clutter.

```
process y=mcivil/x=binladen/m=stereo rthreat/w=age/cov=sex ideo
   /model=85/plot=1/seed=63234.
```

```
%process (data=binladen,y=mcivil,x=binladen,m=stereo rthreat,w=age,
    cov=sex ideo,model=85,plot=1,seed=63234);
```

```
process(data=binladen,y="mcivil",x="binladen",w="age",m=c("stereo","rthreat"),
    cov=c("sex","ideo"),model=85,plot=1,seed=63234)
```

PROCESS output

```
Model  : 85
    Y  : mcivil
    X  : binladen
   M1  : stereo
   M2  : rthreat
    W  : age

Covariates:
 sex       ideo
```

$$\widehat{M_{1i}} = a_{01} + a_{11}X_i + a_{21}W_i + a_{31}X_iW_i$$

```
***************************************************************************
OUTCOME VARIABLE:
 stereo

Model Summary
          R       R-sq       MSE          F        df1        df2          p
      .3643      .1327     .6459    20.0448     5.0000   655.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant     1.7422      .1518    11.4775      .0000     1.4441     2.0402
binladen      .5402      .1980     2.7279      .0065      .1513      .9290
age           .0831      .0245     3.4001      .0007      .0351      .1312
Int_1        -.0834      .0387    -2.1567      .0314     -.1594     -.0075
sex           .0386      .0633      .6095      .5424     -.0857      .1629
ideo          .1300      .0143     9.1235      .0000      .1021      .1580

Product terms key:
 Int_1    :        binladen x        age

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*W      .0062     4.6514     1.0000   655.0000      .0314
```

a_{11} = 0.540
a_{31} = -0.083

a_{11} path

a_{31} path

The effect of OBL death on stereotype endorsement depends on age.

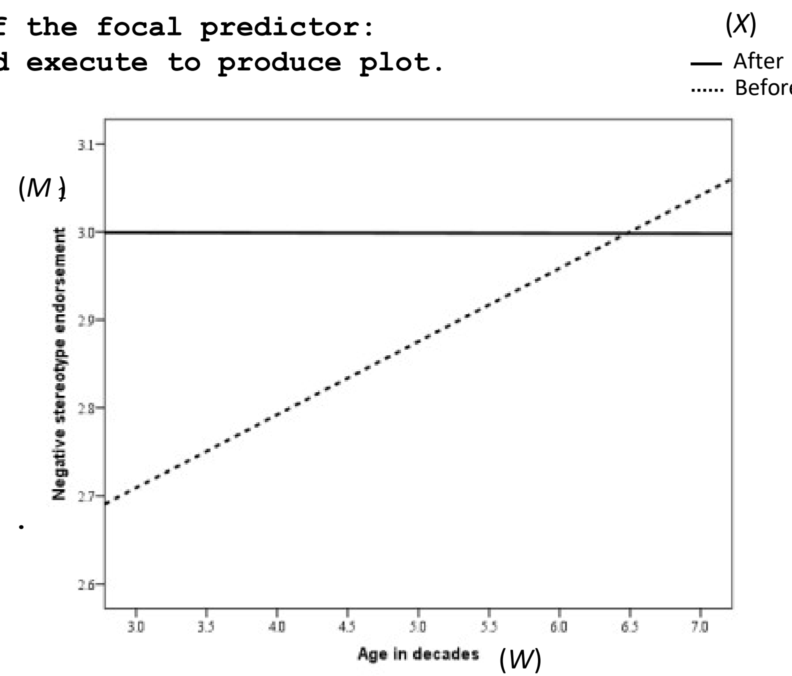
PROCESS output

PROCESS automatically probes the interaction, and the plot option helps visualize it.

```
Focal predict: binladen (X)
      Mod var: age      (W)
```

$$\theta_{X \rightarrow M_1} = 0.540 - 0.083W$$

```
Conditional effects of the focal predictor at values of the moderator(s):

          age    Effect          se          t          p       LLCI       ULCI
16th   3.0000     .2899       .0957     3.0284      .0026      .1019      .4778
50th   4.8000     .1397       .0637     2.1918      .0287      .0145      .2648
84th   6.7000    -.0188       .0959     -.1963      .8444     -.2072      .1695

Data for visualizing the conditional effect of the focal predictor:
Paste text below into a SPSS syntax window and execute to produce plot.

DATA LIST FREE/
   binladen   age        stereo     .
 BEGIN DATA.
      .0000     3.0000   2.7092
     1.0000     3.0000   2.9991
      .0000     4.8000   2.8588
     1.0000     4.8000   2.9985
      .0000     6.7000   3.0168
     1.0000     6.7000   2.9980
END DATA.
GRAPH/SCATTERPLOT=
 age       WITH      stereo    BY       binladen .
```



PROCESS output

$$\widehat{M_{2i}} = a_{02} + a_{12}X_i + a_{22}W_i + a_{32}X_iW_i + dM_{1i}$$

a_{12} = 0.338
a_{32} = -0.062
d = 0.700

```
OUTCOME VARIABLE:
 rthreat

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .6781      .4598      .6060    92.7650     6.0000   654.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant     -.3650      .1611    -2.2654      .0238     -.6814     -.0486
binladen      .3376      .1929     1.7500      .0806     -.0412      .7163
stereo        .6995      .0378    18.4832      .0000      .6252      .7738
age           .0696      .0239     2.9135      .0037      .0227      .1165
Int_1        -.0618      .0376    -1.6428      .1009     -.1356      .0121
sex           .1279      .0613     2.0857      .0374      .0075      .2484
ideo          .0910      .0147     6.2097      .0000      .0622      .1198

Product terms key:
 Int_1    :        binladen x        age

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*W      .0022     2.6989     1.0000   654.0000      .1009
```

a_{12} path

d path

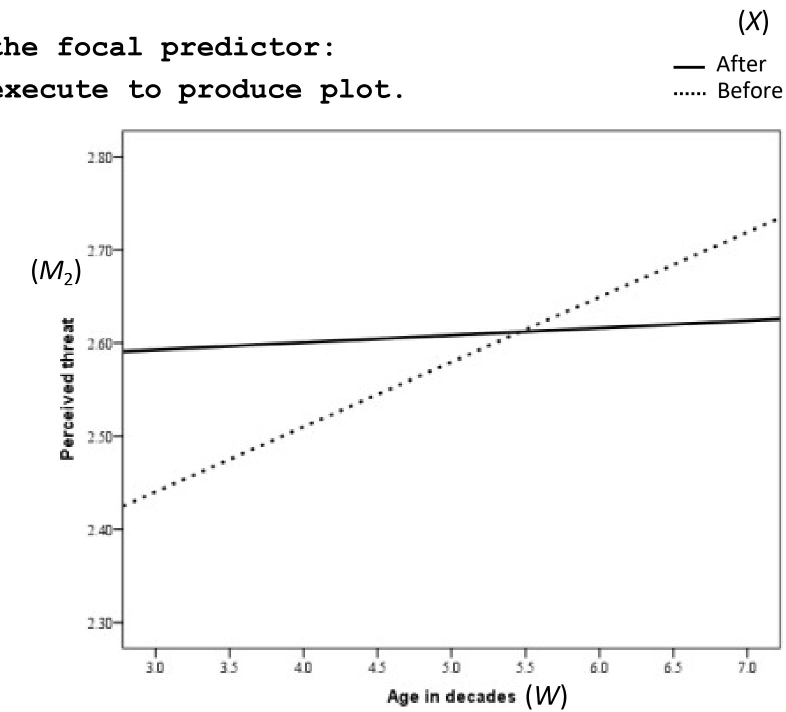
a_{32} path

Non-significant moderation of the effect of OBL death on perceived threat by age.

PROCESS output

PROCESS does not show us the results from probing the interaction because the *p*-value for the interaction is > 0.10. But the plot option still generates what you need to visualize the model. Set the .10 cutoff to something higher if you want, using the **intprobe** option (e.g., **intprobe = 0.2**; **intprobe=1** always probes, regardless of whether significant).

```
   Focal predict: binladen (X)
         Mod var: age      (W)

Data for visualizing the conditional effect of the focal predictor:
Paste text below into a SPSS syntax window and execute to produce plot.

DATA LIST FREE/
   binladen   age         rthreat    .
 BEGIN DATA.
        .0000   3.0000       2.4404
      1.0000   3.0000       2.5926
        .0000   4.8000       2.5657
      1.0000   4.8000       2.6067
        .0000   6.7000       2.6980
      1.0000   6.7000       2.6216

END DATA.
GRAPH/SCATTERPLOT=
 age      WITH     rthreat  BY       binladen .
```



PROCESS output

OUTCOME VARIABLE:
mcivil

$$\widehat{Y}_i = c_0' + c'_1 X_i + c_2' W_i + c'_3 X_i W_i + b_1 M_{1i} + b_2 M_{2i}$$

Model Summary

R	R-sq	MSE	F	df1	df2	p
.6731	.4531	.5894	77.2887	7.0000	653.0000	.0000

Model

	coeff	se	t	p	LLCI	ULCI
constant	.6626	.1595	4.1534	.0000	.3493	.9758
binladen	.1145	.1907	.6004	.5484	-.2599	.4889
stereo	.1046	.0461	2.2703	.0235	.0141	.1950
rthreat	.5471	.0386	14.1883	.0000	.4714	.6229
age	.0016	.0237	.0685	.9454	-.0449	.0482
Int_1	-.0300	.0372	-.8061	.4205	-.1029	.0430
sex	-.1002	.0607	-1.6510	.0992	-.2194	.0190
ideo	.0552	.0149	3.7131	.0002	.0260	.0844

b_1 path

b_2 path

Product terms key:
Int_1 : binladen x age

Test(s) of highest order unconditional interaction(s):

	R2-chng	F	df1	df2	p
X*W	.0005	.6498	1.0000	6 53.0000	.4205

$b_1 =$ **0.105**
$b_2 =$ **0.547**

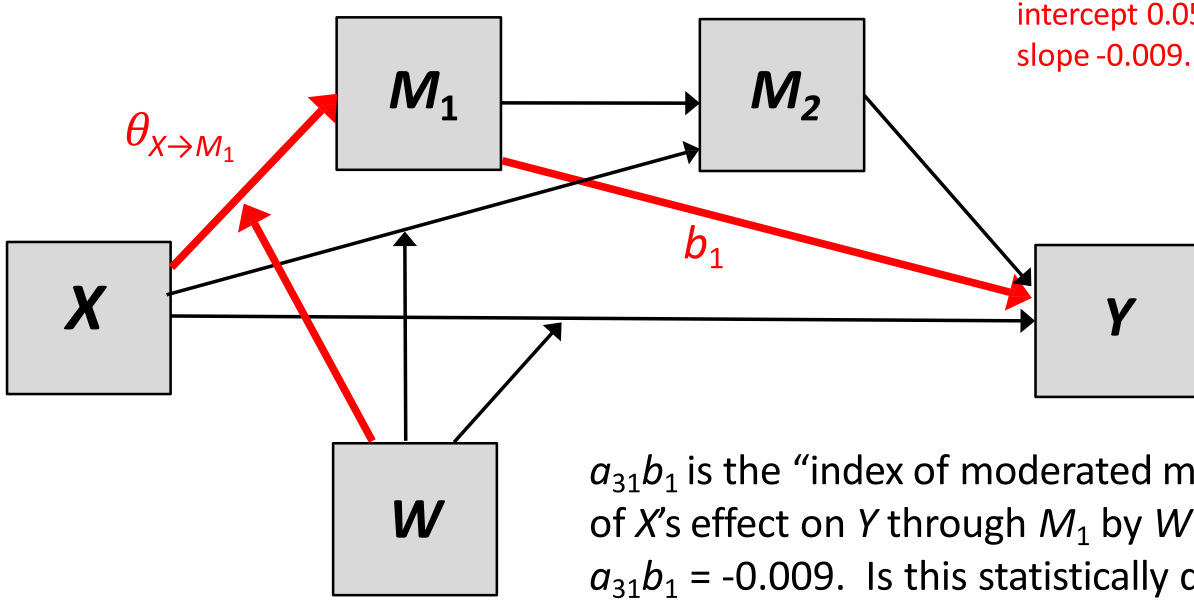
The index and test of moderated mediation

The specific indirect effects are linear functions of age in this model.

The specific indirect effect through **stereotype endorsement** only is

$$\omega_{M_1} = \theta_{X \to M_1} b_1 = (a_{11} + a_{31} W) b_1 = \underline{a_{11} b_1 + a_{31} b_1 W}$$
$$= (0.540 - 0.083\,W)\, 0.105 = \underline{0.057 - 0.009\,W}$$

A linear function with intercept 0.057 and slope -0.009.



$a_{31} b_1$ is the "index of moderated mediation" of *X*'s effect on *Y* through M_1 by *W*. Here, $a_{31} b_1$ = -0.009. Is this statistically different from zero?

The index of moderated mediation

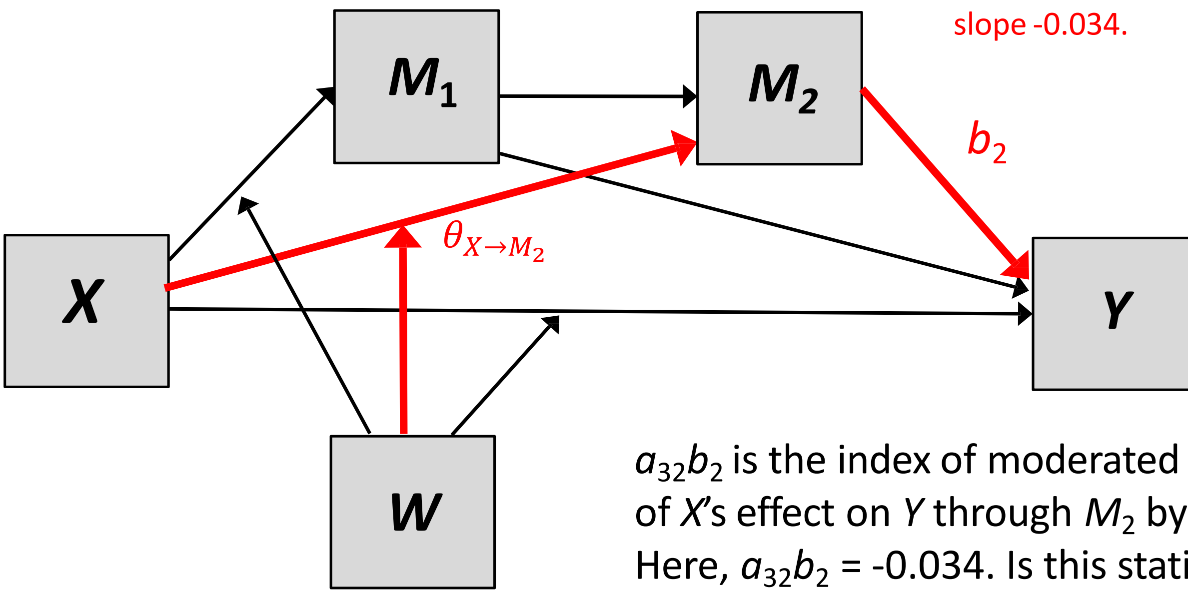
The specific indirect effects are linear functions of age in this model.

The specific indirect effect through **perceived threat** only is

$$\omega_{M_2} = \theta_{X \to M_2} b_2 = (a_{12} + a_{32}W)b_2 = \underline{a_{12}b_2 + a_{32}b_2W}$$

$$= (0.338 - 0.062\,W)\,0.547 = \underline{0.185 - 0.034\,W}$$

A linear function with intercept 0.185 and slope -0.034.



$a_{32}b_2$ is the index of moderated mediation of *X*'s effect on *Y* through M_2 by *W*. Here, $a_{32}b_2$ = -0.034. Is this statistically different from zero?

The index of moderated mediation

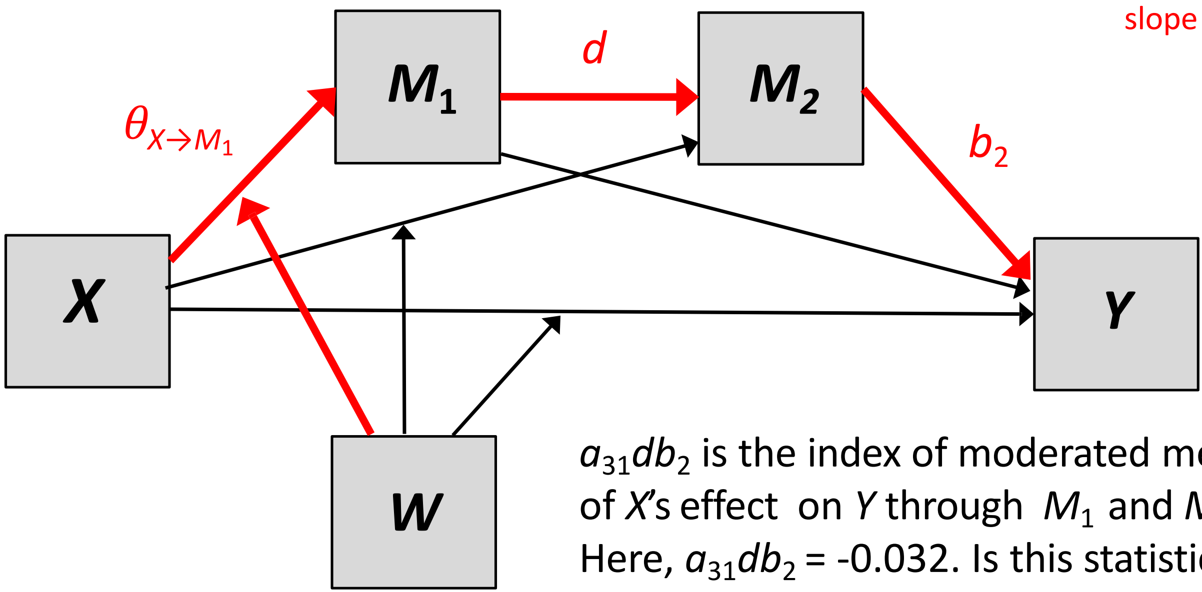
The specific indirect effects are linear functions of age in this model.

The serial indirect effect through **stereotype endorsement** and **perceived threat** is

$$\omega_{M_1M_2} = \theta_{X\rightarrow M_1} db_2 = (a_{11} + a_{31}W)db_2 = \underline{a_{11}db_2 + a_{31}db_2W}$$

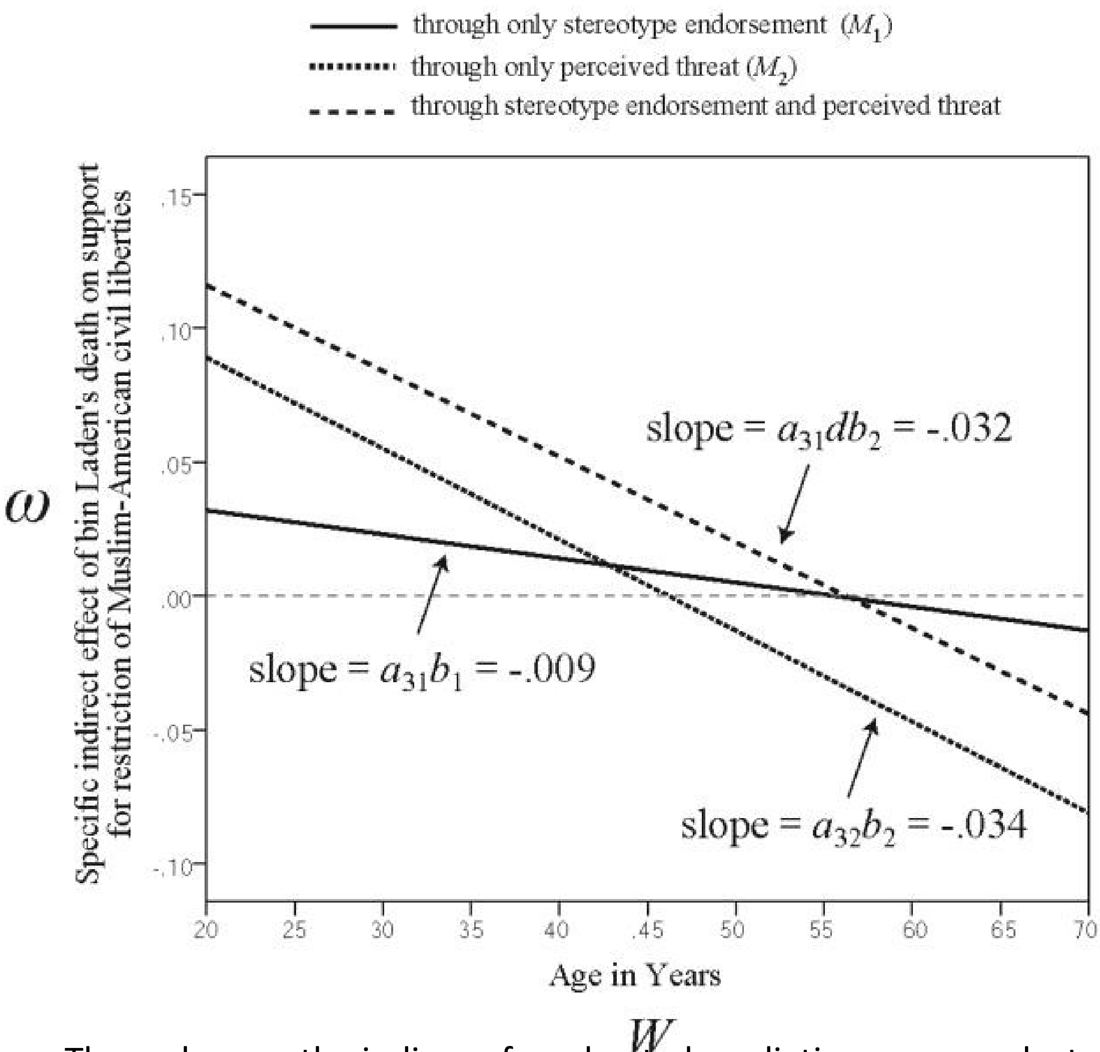
$$= (0.540 - 0.083\,W)(0.700)0.547 = \underline{0.207 - 0.032\,W}$$

A linear function with intercept 0.207 and slope -0.032.



$a_{31}db_2$ is the index of moderated mediation of X's effect on Y through M_1 and M_2 by W. Here, $a_{31}db_2$ = -0.032. Is this statistically different from zero?

A visual representation



These lines visually represent the relationship between the moderator (age) and the specific indirect effect. The slopes are the indices of moderated mediation. If a slope is different from zero, this means there is a linear relationship between the moderator and that specific indirect effect. If the slope is not different from zero, then that specific indirect effect is not linearly moderated.

These slopes—the indices of moderated mediation—are products of regression coefficients. **A bootstrap confidence interval is a good candidate for statistical inference.**

Conditional indirect effects

The indirect effects are linear functions of W and thus will depend on the value of W you condition the estimate on. Let's estimate the conditional indirect effects of X among people "relatively younger", "relatively moderate" in age, and "relatively older". We'll define these as the 16th percentile (3.000 = 30 years), the 50th percentile (4.800 = 48 years) and the 84th percentile (6.700 = 67 years). These are arbitrary, of course. We could use other values.

$$\omega_{M_1} = \theta_{X\to M_1} b_1 = (a_{11} + a_{31}W)\, b_1 = a_{11}b_1 + a_{31}b_1W = 0.057 - 0.009W$$

$$\omega_{M_2} = \theta_{X\to M_2} b_2 = (a_{12} + a_{32}W)\, b_2 = a_{12}b_2 + a_{32}b_2W = 0.185 - 0.034W$$

$$\omega_{M_1M_2} = \theta_{X\to M_1} db_2 = (a_{11} + a_{31}W)db_2 = a_{11}db_2 + a_{31}db_2W = 0.207 - 0.032W$$

W	$\theta_{X\to M_1}$	$\theta_{X\to M_2}$	d	b_1	b_2	ω_{M_1}	ω_{M_2}	$\omega_{M_1M_2}$
3.000	0.290	0.152	0.700	0.105	0.547	0.030	0.083	0.110
4.800	0.140	0.041	0.700	0.105	0.547	0.015	0.022	0.054
6.700	-0.019	-0.076	0.700	0.105	0.547	-0.002	-0.042	-0.007

Conditional indirect effects

W	$\theta_{X\rightarrow M_1}$	$\theta_{X\rightarrow M_2}$	d	b_1	b_2	ω_{M_1}	ω_{M_2}	$\omega_{M_1M_2}$
3.000	0.290	0.152	0.700	0.105	0.547	0.030	0.083	0.110
4.800	0.140	0.041	0.700	0.105	0.547	0.015	0.022	0.054
6.700	-0.019	-0.076	0.700	0.105	0.547	-0.002	-0.042	-0.007

Greater perceived threat leads to greater restriction of Muslim civil liberties

Greater stereotype endorsement leads to greater restriction of Muslim civil liberties

Greater stereotype endorsement leads to greater perceived threat

Bigger difference in perceived threat following his death relative to before among the younger, but little difference in perceived threat before and after among the older.

Bigger difference in stereotype endorsement following his death relative to before among the younger, but little difference in stereotype endorsement before and after among the older.

The conditional indirect effects need to be interpreted considering the sign of their components (i.e., each path in the causal system)

Conditional indirect effects

W	$\theta_{X \to M_1}$	$\theta_{X \to M_2}$	d	b_1	b_2	ω_{M_1}	ω_{M_2}	$\omega_{M_1 M_2}$
3.000	0.290	0.152	0.700	0.105	0.547	0.030	0.083	0.110
4.800	0.140	0.041	0.700	0.105	0.547	0.015	0.022	0.054
6.700	-0.019	-0.076	0.700	0.105	0.547	-0.002	-0.042	-0.007

Larger positive indirect effect of OBL death on restriction of Muslim civil liberties through stereotype endorsement among the relatively younger. Little to no indirect effect among the older.

Larger positive indirect effect of OBL death on restriction of Muslim civil liberties through perceived threat among the relatively younger. Little to no indirect effect among the older.

Larger positive indirect effect of OBL death on restriction of Muslim civil liberties through stereotype endorsement and perceived threat among the relatively younger. Weaker indirect effect among the moderate in age, no apparent indirect effect among the older.

These indirect effects are on the same scale and so can be compared meaningfully. Most of the mechanism operates through the serial mediation pathway.

$X \rightarrow M_1 \rightarrow Y$

The specific indirect effect through **stereotype endorsement** only is

$$\omega_{M_1} = \theta_{X \to M_1} b_1 = (a_{11} + a_{31} W) b_1 = a_{11} b_1 + a_{31} b_1 W$$

$$= (0.540 - 0.083\,W)\; 0.105 = 0.057 - \boxed{0.009\,W}$$

```
INDIRECT EFFECT:
 binladen     ->     stereo       ->     mcivil

        age       Effect      BootSE    BootLLCI     BootULCI
     3.0000        .0303       .0178       .0015        .0698
     4.8000        .0146       .0099      -.0004        .0378
     6.7000       -.0020       .0118      -.0276        .0217

      Index of moderated mediation:
             Index      BootSE     BootLLCI     BootULCI
age         -.0087       .0062       -.0238        .0003
```

As this specific indirect effect is not significantly moderated based on the index of moderated mediation, we can't say that these differ from each other.

W	ω_{M_1}	ω_{M_2}	$\omega_{M_1M_2}$
3.000	0.030	0.083	0.110
4.800	0.015	0.022	0.054
6.700	-0.002	-0.042	-0.007

$X \rightarrow M_2 \rightarrow Y$

The specific indirect effect through **perceived threat** only is

$$\omega_{M_2} = \theta_{X \to M_2} b_2 = (a_{12} + a_{32}W)b_2 = a_{12}b_2 + a_{32}b_2W$$

$$= (0.338 - 0.062\,W)\,0.547 = 0.185 - 0.034\,W$$

```
INDIRECT EFFECT:
 binladen     ->     rthreat      ->     mcivil

       age      Effect     BootSE   BootLLCI   BootULCI
    3.0000       .0833      .0482    -.0070      .1794
    4.8000       .0224      .0336    -.0420      .0911
    6.7000      -.0418      .0502    -.1396      .0575

      Index of moderated mediation:
           Index     BootSE   BootLLCI   BootULCI
age      -.0338      .0195    -.0724      .0040
```

As this specific indirect effect is not significantly moderated based on the index of moderated mediation, we can't say that these differ from each other.

W	ω_{M_1}	ω_{M_2}	$\omega_{M_1M_2}$
3.000	0.030	0.083	0.110
4.800	0.015	0.022	0.054
6.700	-0.002	-0.042	-0.007

$X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$

The specific indirect effect through **stereotype endorsement** and **perceived threat** is

$$\omega_{M_1M_2} = \theta_{X\rightarrow M_1} db_2 = (a_{11} + a_{31}W)db_2 = a_{11}db_2 + a_{31}db_2W$$

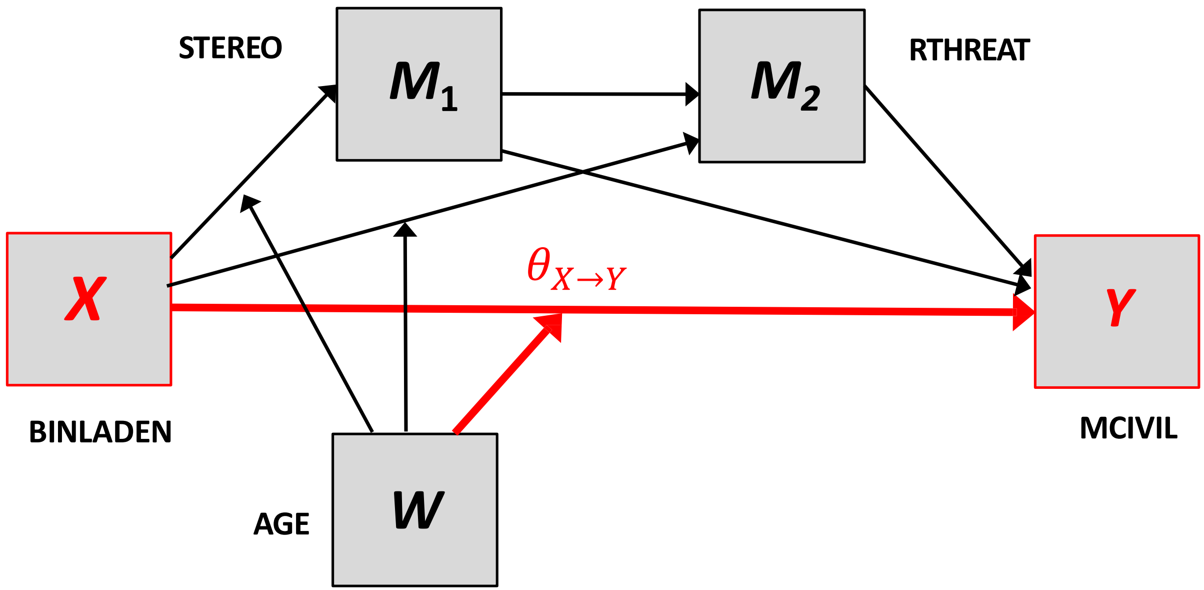
$$= (0.540 - 0.083\,W)(0.700)0.547 = 0.207 - 0.032\,W$$

```
INDIRECT EFFECT:
 binladen     ->     stereo        ->     rthreat      ->     mcivil

     Age           Effect      BootSE    BootLLCI    BootULCI
    3.0000          .1109       .0367      .0424       .1877
    4.8000          .0535       .0248      .0064       .1047
    6.7000         -.0072       .0389     -.0841       .0701

       Index of moderated mediation:
               Index      BootSE     BootLLCI     BootULCI
     age      -.0319       .0154      -.0628       -.0019
```

As this specific indirect effect is significantly moderated based on the index of moderated mediation, we can say that these differ from each other.

W	ω_{M_1}	ω_{M_2}	$\omega_{M_1M_2}$
3.000	0.030	0.083	0.110
4.800	0.015	0.022	0.054
6.700	-0.002	-0.042	-0.007

The direct effect of *X*



$$\hat{Y}_i = c'_0 + c'_1 X_i + c'_2 W_i + c'_3 X_i W_i + b_1 M_{1i} + b_2 M_{2i}$$

can be written as $\hat{Y}_i = c'_0 + (c'_1 + c'_3 W_i) X_i + c'_2 W_i + b_1 M_{1i} + b_2 M_{2i}$ or, alternatively, $\hat{Y}_i = c'_0 + \theta_{X\rightarrow Y} X_i + c'_2 W_i + b_1 M_{1i} + b_2 M_{2i}$ where $\theta_{X\rightarrow Y} = c'_1 + c'_3 W_i$ is the conditional direct effect of *X* on *Y*. *X*'s effect on *Y* independent of M_1 and M_2 is a function of *W*.

The direct effect of *X*

OUTCOME VARIABLE:
mcivil

$$\widehat{Y}_i = c_0' + c'_1 X_i + c_2' W_i + c'_3 X_i W_i + b_1 M_{1i} + b_2 M_{2i}$$

Model Summary

R	R-sq	MSE	F	df1	df2	p
.6731	.4531	.5894	77.2887	7.0000	653.0000	.0000

Model

	coeff	se	t	p	LLCI	ULCI
constant	.6626	.1595	4.1534	.0000	.3493	.9758
binladen	.1145	.1907	.6004	.5484	-.2599	.4889
stereo	.1046	.0461	2.2703	.0235	.0141	.1950
rthreat	.5471	.0386	14.1883	.0000	.4714	.6229
age	.0016	.0237	.0685	.9454	-.0449	.0482
Int_1	-.0300	.0372	-.8061	.4205	-.1029	.0430
sex	-.1002	.0607	-1.6510	.0992	-.2194	.0190
ideo	.0552	.0149	3.7131	.0002	.0260	.0844

c'_1

c'_3

$c'_1 =$ 0.115
$c'_3 =$ -0.030

Product terms key:
Int_1 : binladen x age

Test(s) of highest order unconditional interaction(s):

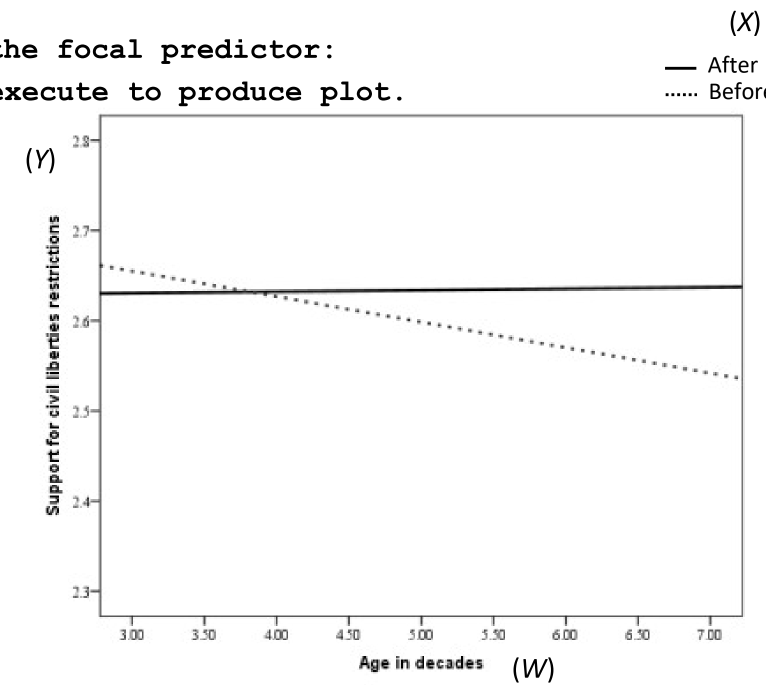
	R2-chng	F	df1	df2	p
X*W	.0005	.6498	1.0000	653.0000	.4205

The direct effect of OBL death on restriction of Muslim civil liberties does not seem to be moderated by age.

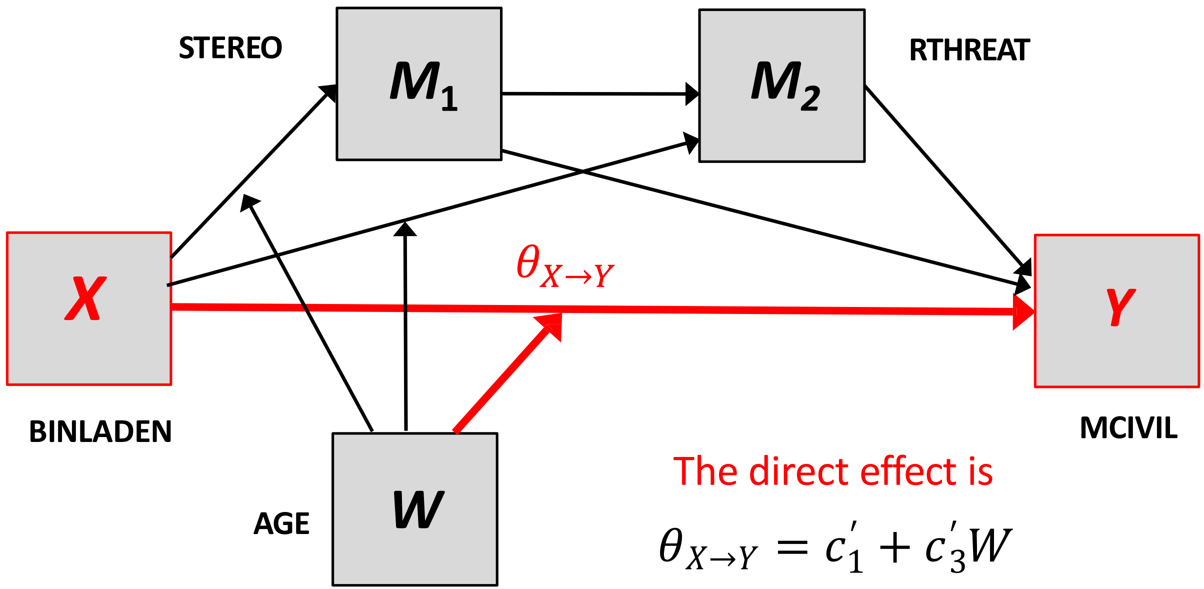
PROCESS output

PROCESS does not show us the results from probing the interaction in the model of Y because the *p*-value for the interaction is > 0.10. But the plot option still generates what you need to visualize the model. Set the .10 cutoff to something higher if you want, using the **intprobe** option.

```
    Focal predict: binladen (X)
          Mod var: age      (W)

 Data for visualizing the conditional effect of the focal predictor:
 Paste text below into a SPSS syntax window and execute to produce plot.

 DATA LIST FREE/
  binladen    age         mcivil.
 BEGIN DATA.
        .0000     3.0000   2.6307
       1.0000     3.0000   2.6553
        .0000     4.8000   2.6336
       1.0000     4.8000   2.6043
        .0000     6.7000   2.6367
       1.0000     6.7000   2.5505
 END DATA.
 GRAPH/SCATTERPLOT=
  age       WITH     mcivil   BY       binladen .
```



The direct effect of *X*



The direct effect is

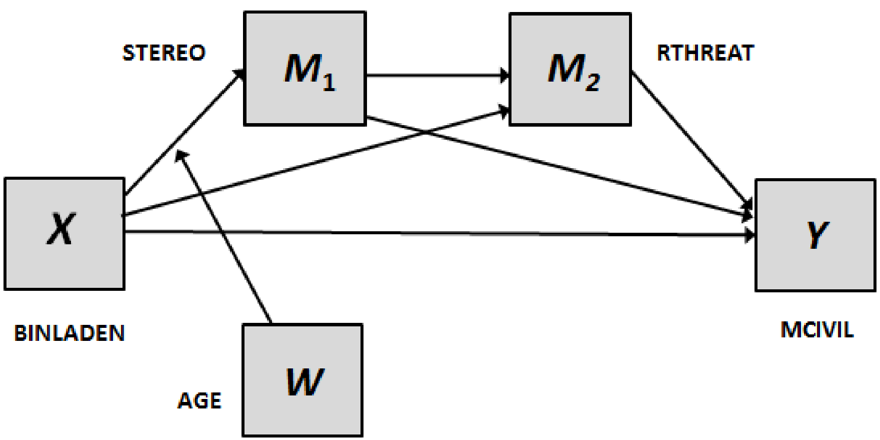
$$\theta_{X \to Y} = c_1' + c_3' W$$
$$= 0.115 - 0.030\, W$$

c'_3 is not statistically different from zero, so we can't say the direct effect is moderated. PROCESS shows us the conditional effects at different values of *W* in the summary section.

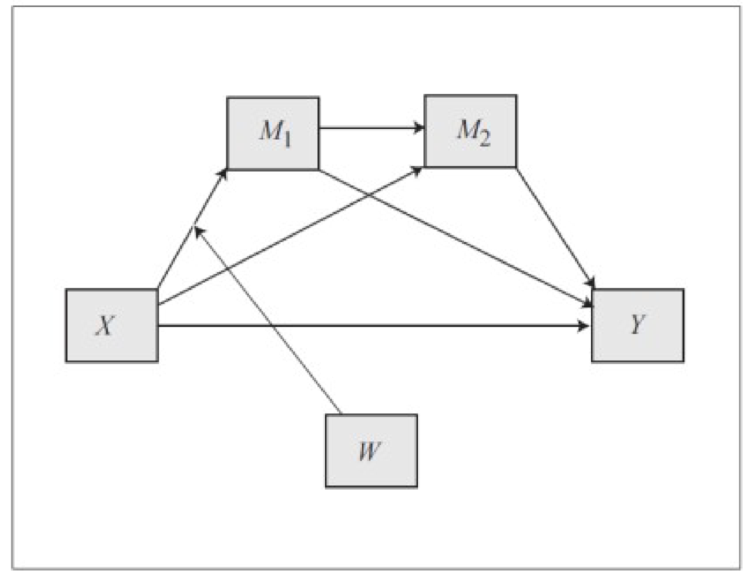
```
Conditional direct effect(s) of X on Y:
        age     Effect         se          t          p       LLCI       ULCI
     3.0000      .0246      .0923      .2669      .7897     -.1565      .2058
     4.8000     -.0293      .0611     -.4794      .6318     -.1493      .0907
     6.7000     -.0862      .0917     -.9405      .3473     -.2662      .0938
```

“Pruning" the model

One option is to fix the direct effect of X on Y, and perhaps the effect of X on M_2 to be independent of W, but allow the effect of X on M_1 to vary with W. Model 83 does the job.



Model 83



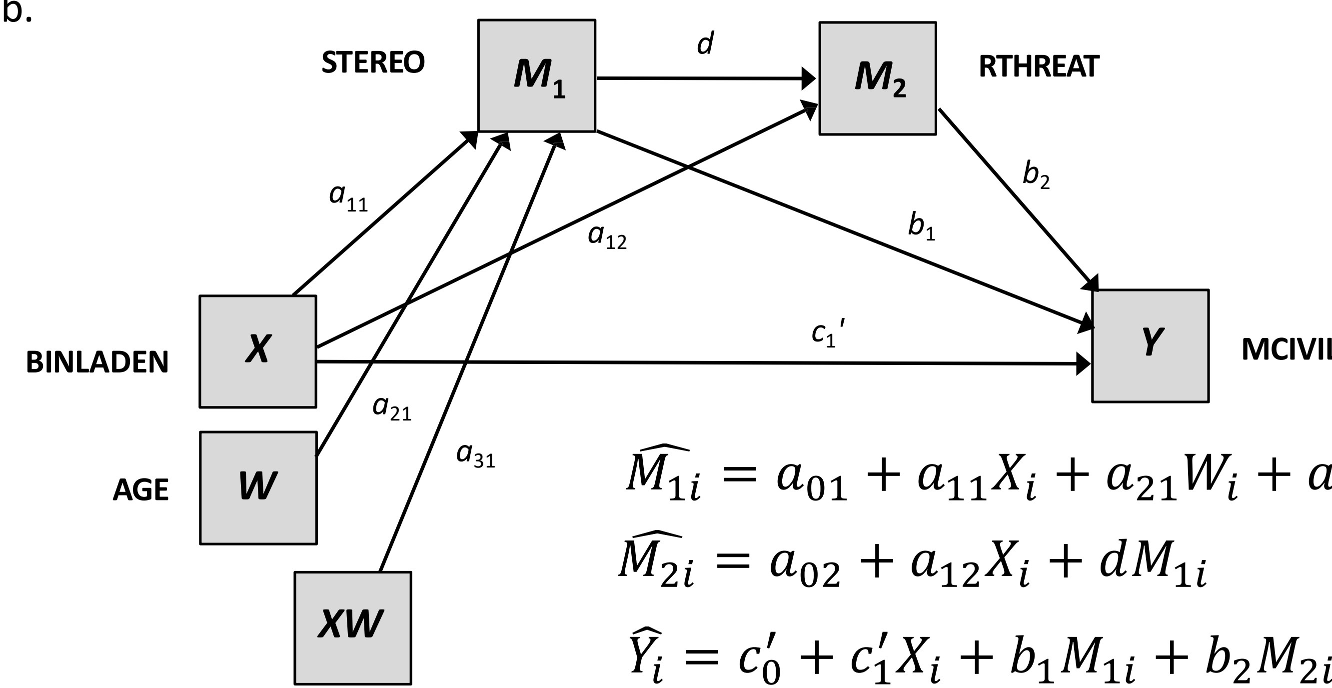

```
process y=mcivil/x=binladen/m=stereo rthreat/w=age/cov=sex ideo/model=83
 /plot=1/seed=63234.
```

```
%process (data=binladen,y=mcivil,x=binladen,m=stereo rthreat,w=age, cov=sex
  ideo,model=83,plot=1,seed=63234)
```

```
process(data=binladen,y="mcivil",x="binladen",m=c("stereo","rthreat"),
 w="age",cov=c("sex","ideo"),model=83,plot=1,seed=63234)
```

The statistical model

It seems clear that we should probably fix the effect of X on M_2 and the direct effect of X on Y to be independent of W, but allow the effect of X on M_1 to vary with W. Model 83 does the job.



$$\widehat{M_{1i}} = a_{01} + a_{11}X_i + a_{21}W_i + a_{31}X_iW_i$$

$$\widehat{M_{2i}} = a_{02} + a_{12}X_i + dM_{1i}$$

$$\hat{Y}_i = c_0' + c_1'X_i + b_1M_{1i} + b_2M_{2i}$$

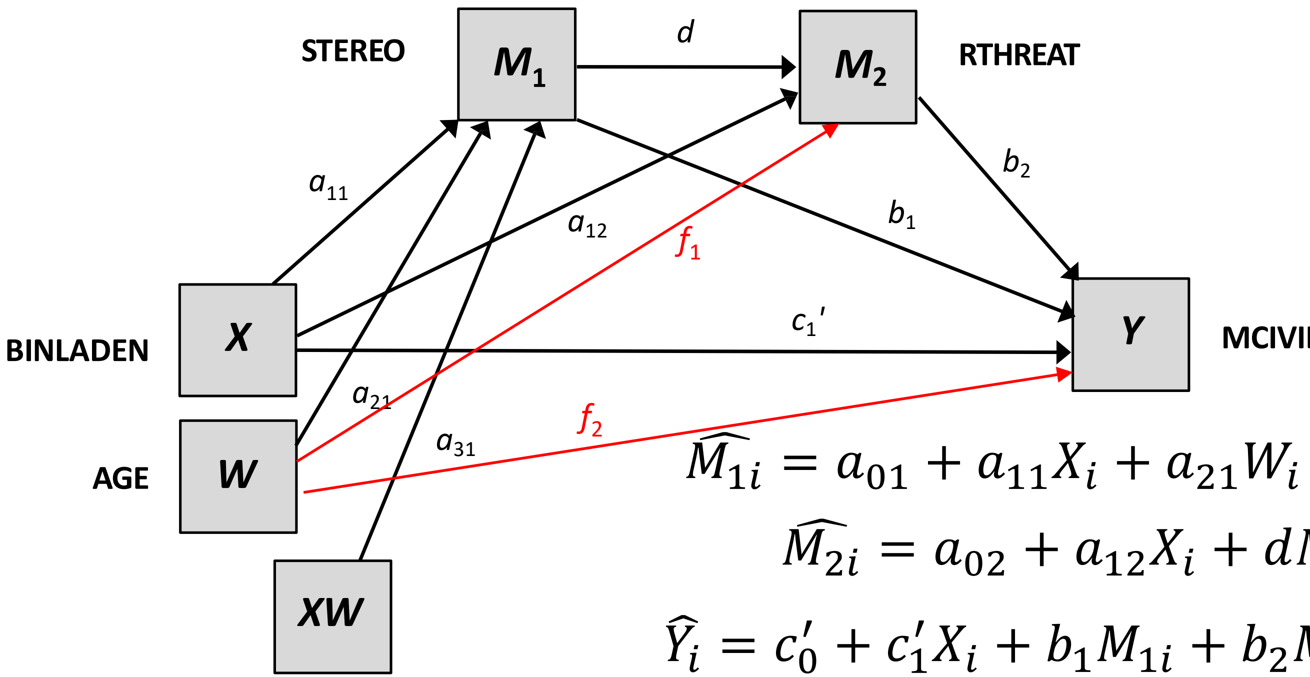
```
process y=mcivil/x=binladen/m=stereo rthreat/w=age/cov=sex ideo /model=83/plot=1/seed=63234.
```

```
%process (data=binladen,y=mcivil,x=binladen,m=stereo rthreat,w=age,cov=sex ideo,model=83,
 plot=1,seed=63234)
```

```
process (data=binladen,y="mcivil",x="binladen",m=c("stereo","rthreat"),w="age",
 cov=c("sex","ideo"),model=83,plot=1,seed=63234)
```

But what if we want age as a covariate elsewhere?

Removing the moderation of X's effect on M_2 and on Y by W by switching from model 85 to model 83 eliminates W from the models of M_2 and Y entirely. If you want age in the other equations, list it as a covariate (requires PROCESS version 3.2 or later).



$$\widehat{M_{1i}} = a_{01} + a_{11}X_i + a_{21}W_i + a_{31}X_iW_i$$

$$\widehat{M_{2i}} = a_{02} + a_{12}X_i + dM_{1i} + f_1W_i$$

$$\widehat{Y_i} = c_0' + c_1'X_i + b_1M_{1i} + b_2M_{2i} + f_2W_i$$

```
process y=mcivil/x=binladen/m=stereo rthreat/w=age/cov=age sex ideo/model=83/plot=1/seed=63234.
```

```
%process (data=binladen,y=mcivil,x=binladen,m=stereo rthreat,w=age,cov=age sex ideo,model=83,
 plot=1,seed=63234)
```

```
process(data=binladen,y="mcivil",x="binladen",m=c("stereo","rthreat"),w="age",cov=c("age",
 "sex","ideo"),model=83,plot=1,seed=63234)
```

Multicategorical Variables

Mediation (Mcat X)
Moderation (Mcat X, Mcat W)
Moderated Mediation (Mcat X, Mcat W)

A study of sex discrimination

Garcia, D. M., Schmitt, M. T., Branscombe, N. R., & Ellemers, N. (2010). Women's reactions to in-group members who protest discriminatory treatment: The importance of beliefs about inequality and response appropriateness. ***European Journal of Social Psychology, 49***, 733-745.

- Participants (all female) read a narrative about a female attorney who lost a promotion at her firm to a much less qualified male through discriminatory actions of the senior partners.

- Participants randomly assigned to one of two "protest" conditions were then told she protested the decision by presenting an argument to the partners about how unfair the decision was. Some were told she framed her protest around herself (individual protest condition), whereas others were told she framed her protest around the collective of women (collective protest condition).

- Participants randomly assigned to the "no protest" condition were told that although she was disappointed, she accepted the decision and continued working at the firm.

PROTEST: Experimental condition (0 = no protest, 1 = individual protest, 2 = collective protest)

A study of sex discrimination

Garcia, D. M., Schmitt, M. T., Branscombe, N. R., & Ellemers, N. (2010). Women's reactions to in-group members who protest discriminatory treatment: The importance of beliefs about inequality and response appropriateness. ***European Journal of Social Psychology, 49***, 733-745.

- After reading the narrative, the participants responded to a set of questions measuring how appropriate she felt the attorney's behavior was for the circumstance.

  APPROP: Perceived appropriateness of her response (higher = more appropriate)

- The participants evaluated the characteristics of the attorney, the responses to which were aggregated to produce a measure of positive evaluation or "liking."

  EVAL : Evaluation of the lawyer (higher = more positive evaluation, i.e., more likable)

- The participants also filled out the Modern Sexism Scale.

  SEXISM: Beliefs about the pervasiveness of sex discrimination in society (higher = sex discrimination perceived as more pervasive in society).

The data: lawyer2

lawyer2.sav

subnum	protest	sexism	angry	eval	approp
209	2	4.87	2	4.83	4.25
44	0	4.25	1	4.50	5.75
124	2	5.00	3	5.50	4.75
232	2	5.50	1	5.66	7.00
30	2	5.62	1	6.16	6.75
140	1	5.75	1	6.00	5.50
27	2	5.12	2	4.66	5.00
64	0	6.62	1	6.50	6.25
67	0	5.75	6	1.00	3.00
182	0	4.62	1	6.83	5.75
85	2	4.75	2	5.00	5.25

lawyer2.sas

lawyer2

```
data lawyer2;
 input protest sexism angry eval approp;
 datalines;
 2   4.87   2   4.83   4.25
 0   4.25   1   4.50   5.75
 2   5.00   3   5.50   4.75
 2   5.50   1   5.66   7.00
 2   5.62   1   6.16   6.75
 1   5.75   1   6.00   5.50
 2   5.12   2   4.66   5.00
 0   6.62   1   6.50   6.25
 0   5.75   6   1.00   3.00
 0   4.62   1   6.83   5.75
 2   4.75   2   5.00   5.25
```

In R: Don't forget to change the path below to where your **lawyer2.csv** file is located.

```
> lawyer2<-read.table("c:/mmcpa/lawyer2.csv", sep=",",header=TRUE)
> head(lawyer2)
  subnum protest sexism angry eval approp
1    209       2   4.87     2 4.83   4.25
2     44       0   4.25     1 4.50   5.75
3    124       2   5.00     3 5.50   4.75
4    232       2   5.50     1 5.66   7.00
5     30       2   5.62     1 6.16   6.75
6    140       1   5.75     1 6.00   5.50
```

Our objectives with these data

	Perceived Response Appropriateness (M)		Evaluation (Y)		
	M	SD	Y	SD	$\overline{Y}^*$
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

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

We'll examine whether the effect of the attorney's response on how she is evaluated (3 conditions)…

… operates through the mechanism of how appropriate her response to the situation is perceived as being (mediation)

… depends on perceived pervasiveness of sex discrimination in society (moderation).

… as mediated by the mechanism of perceived appropriateness of her response is moderated by perceived pervasiveness of sex discrimination (moderated mediation)

Single-factor ("one-way") analysis of variance

Did the lawyer's choice as to how to respond (not at all, individual protest, collective protest) influence how she was perceived? That is, is there a difference between conditions, on average, in how positively she was evaluated? Most would answer this using a single-factor (a.k.a. "one-way") ANOVA.

```
means tables = eval by protest/statistics anova.
```

```
proc anova data=lawyer2;
class protest;model eval=protest;means protest;run;
```

```
aggregate(lawyer2$eval,list(lawyer2$protest),mean)
protest.f<-factor(lawyer2$protest)
summary(aov(eval~protest.f,data=lawyer2))
```

H_0: all means equal

H_a: all means not equal

```
  Group.1        x
1       0 5.310244
2       1 5.826047
3       2 5.753333
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
protest.f     2   6.52   3.262   3.055 0.0506 .
Residuals   126 134.52   1.068
```

The lawyer's response to the discrimination affected how positively she was perceived on average, $F(2,126)$ =3.055, p = .051. She was most positively evaluated when she protested individually (Mean = 5.83, SD = 0.82), next most when protesting collectively (Mean = 5.75, SD = 0.94), and least when she didn't protest at all (Mean = 5.31, SD = 1.30).

Mediation

Does perceived response appropriateness mediate this effect?

	Perceived Response Appropriateness (M)		Evaluation (Y)		
	M	SD	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

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

Here, the presumed cause is a multicategorical variable with three levels. How does one conduct a mediation analysis in such a design?

Mediation of the effect of a multicategorical independent variable

Does perceived response appropriateness mediate this effect?

The causal steps approach

	Perceived Response Appropriateness (M)		Evaluation (Y)		
	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

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

From a single-factor ANOVA, the effect of experimental condition **on**…evaluation: $F(2,126) = 3.055$, $p = .051$ (**total effect**)

…response appropriateness: $F(2,126) = 22.219$, $p < .001$ (‘**a**’ **effect**)

From a single-factor ANCOVA…the relationship between perceived response appropriateness and evaluation is positive, $b = 0.412$, $p < 0.001$ (‘***b***’ **effect**), and…

…the effect of condition on evaluation disappears after controlling for response appropriateness, $F(2,125) = 0.729$, $p = .485$ (**direct effect**)

This approach has all the problems of the causal steps (“Baron and Kenny”) approach. There is a better way.

Representing a multicategorical predictor in a linear model

Predictor variables in a linear model must be quantitative or dichotomous (i.e., categorical with only two values). What if we want to include multicategorical variables? **How do we proceed?**

Any multicategorical variable with k categories can be represented with $k - 1$ variables in a regression model. For example, we can code membership in a category with a set of *dummy variables* and all of these $k - 1$ dummy variables in the model.

Dummy coding (a.k.a. "Indicator coding")

Set D_1 to 1 for cases in category 1, 0 otherwise
D_2 to 1 for cases in category 2, 0 otherwise
.
.
$D_{(k-1)}$ to 1 for cases in category $k - 1$, 0 otherwise

Category k is called the "reference category," for reasons that will be clear soon. It is represented here in the coding system, but it doesn't seem so.

Dummy coding condition

```
frequencies variables = protest.
```

```
proc freq data=lawyer2;tables protest;run;
```

```
table(lawyer2$protest)
```

No protest	Individual protest	Collective protest
0	1	2
41	43	45

One possible dummy variable coding system for condition (k = 3 categories):

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

The "reference category" is the one with zeros on all $k - 1$ dummy variables. In this example, those told the lawyer did not protest are the reference category.

Constructing dummy variables

There is a variety of ways of constructing dummy codes in a computing platform, each with its dangers, assumptions, and conveniences.

		D_1	D_2
protest = 0	**No protest**	0	0
protest = 1	**Individual**	1	0
protest = 2	**Collective**	0	1

```
compute d1 = (protest=1).
compute d2 = (protest=2).
```

```
data lawyer2;set lawyer2;
d1 = (protest=1);
d2 = (protest=2);
if (protest=.) then d1=.;
if (protest=.) then d2=.;
run;
```

```
d1<-as.numeric(lawyer2$protest==1)
d2<-as.numeric(lawyer2$protest==2)
lawyer2.dummy<-data.frame(lawyer2,d1,d2)
```

Estimating evaluation from experimental condition using regression

```
compute d1 = (protest=1).
compute d2 = (protest=2).
regression/dep = eval/method = enter d1 d2.
```

```
data lawyer2;set lawyer2;
d1 = (protest=1);d2 = (protest=2);run;
proc reg data=lawyer2;model eval = d1 d2;run;
```

We know there are no missing data on **protest**, so this is ok.

```
d1<-as.numeric(lawyer2$protest==1);d2<-as.numeric(lawyer2$protest==2)
lawyer2.dummy<-data.frame(lawyer2,d1,d2)
summary(lm(eval~d1+d2,data=lawyer2.dummy))
```

$$\widehat{Y}_i = 5.310 + 0.516D_{1i} + 0.443D_{2i}$$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.3102     0.1614  32.908   <2e-16 ***
d1            0.5158     0.2255   2.287   0.0239 *
d2            0.4431     0.2231   1.986   0.0492 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.033 on 126 degrees of freedom
Multiple R-squared:  0.04625,	Adjusted R-squared:  0.03111
F-statistic: 3.055 on 2 and 126 DF,  p-value: 0.05062
```

The model reproduces the group means

$$\widehat{Y}_i = 5.310 + 0.516D_{1i} + 0.443D_{2i}$$

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

No protest $\widehat{Y}_i = 5.310 + 0.516(0) + 0.443(0) = 5.310 = Y_{NP}$

Individual protest $\widehat{Y}_i = 5.310 + 0.516(1) + 0.443(0) = 5.826 = Y_{IP}$

Collective protest $\widehat{Y}_i = 5.310 + 0.516(0) + 0.443(1) = 5.753 = Y_{CP}$

Report

EVAL: evaluation of attorney

PROTEST: experimental condition	Mean	N	Std. Deviation
no protest	5.3102	41	1.30158
individual	5.8260	43	.81943
collective	5.7533	45	.93601
Total	5.6367	129	1.04970

Interpretation of the coefficients

$$\widehat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i}$$

$$\widehat{Y}_i = 5.310 + 0.516 D_{1i} + 0.443 D_{2i}$$

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

b_0 = 5.310 This is the mean evaluation among those assigned to the no protest condition ($D_1 = 0$, $D_2 = 0$).

b_1 = 0.516 This is the mean difference in evaluation between those in the individual protest condition ($D_1 = 1$, $D_2 = 0$) and those in the no protest condition ($D_1 = 0$, $D_2 = 0$)

$$b_1 = \overline{Y}_{IP} - \overline{Y}_{NP} = 5.826 - 5.310 = 0.516$$

b_2 = 0.443 This is the mean difference in evaluation between those in the collective protest condition ($D_1 = 0$, $D_2 = 1$) and those in the no protest condition ($D_1 = 0$, $D_2 = 0$)

$$b_2 = \overline{Y}_{CP} - \overline{Y}_{NP} = 5.753 - 5.310 = 0.443$$

When D_1 and D_2 are indicator codes constructed in this fashion, b_1 estimates the mean difference in *Y* between the group coded by D_1 and the reference group, and b_2 estimates the mean difference in *Y* between the group coded by D_2 and the reference group.

Statistical inference

We are estimating the coefficients of a model of the form

$$_T\widehat{Y}_i = {}_Tb_0 + {}_Tb_1 D_{1i} + {}_Tb_2 D_{2i}$$

If there is no actual difference, on average, between these groups on *Y*, this implies that both “true” regression coefficients $_Tb_1$ and $_Tb_2$ are equal to zero. The null hypothesis can be tested by converting the obtained R^2 to an *F*-ratio and then deriving a *p*-value.

H_0: $_Tb_1 = {}_Tb_2 = 0$
H_a: at least one is different from zero

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.3102     0.1614  32.908   <2e-16 ***
d1            0.5158     0.2255   2.287   0.0239 *
d2            0.4431     0.2231   1.986   0.0492 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.033 on 126 degrees of freedom
Multiple R-squared:  0.04625,   Adjusted R-squared:  0.03111
F-statistic: 3.055 on 2 and 126 DF,  p-value: 0.05062
```

$$F(k-1, df_{residual}) = \frac{df_{residual} R^2}{(k-1)(1-R^2)}$$

$$F(2,126) = \frac{126(0.046)}{2(1-0.046)}$$

$$F(2,126) = 3.055$$

F(2,126) = 3.055, *p* = .051. Fail to Reject H_0. The three group means do not differ from each other by more than could be explained by just ‘chance’. Compare this to the one-way ANOVA from earlier.

Statistical inference

We are estimating the coefficients of a model of the form

$$_T\widehat{Y}_i = {_T}b_0 + {_T}b_1 D_{1i} + {_T}b_2 D_{2i}$$

b_1 and b_2 can also be used to test hypotheses about differences between groups—specifically, between the group a dummy variable codes and the reference group.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.3102     0.1614  32.908   <2e-16 ***
d1            0.5158     0.2255   2.287   0.0239 *
d2            0.4431     0.2231   1.986   0.0492 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.033 on 126 degrees of freedom
Multiple R-squared:  0.04625,   Adjusted R-squared:  0.03111
F-statistic: 3.055 on 2 and 126 DF,  p-value: 0.05062
```

H_0: ${_T}b_1 = 0$
H_a: ${_T}b_1 \neq 0$
$b_1 = 0.516$, $t(126) = 2.287$, $p = .024$

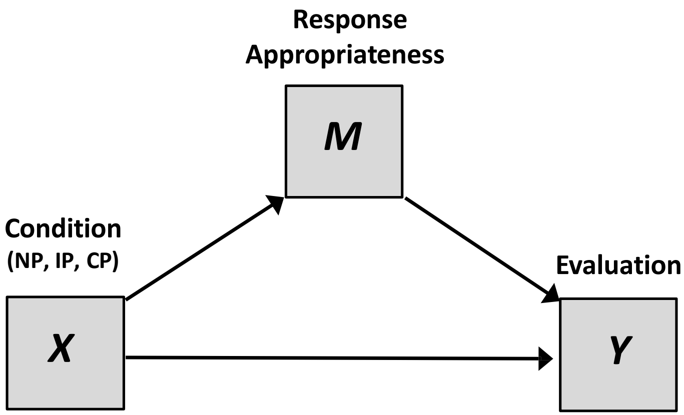
Those told she individually protested liked her more on average than those told she did not protest.

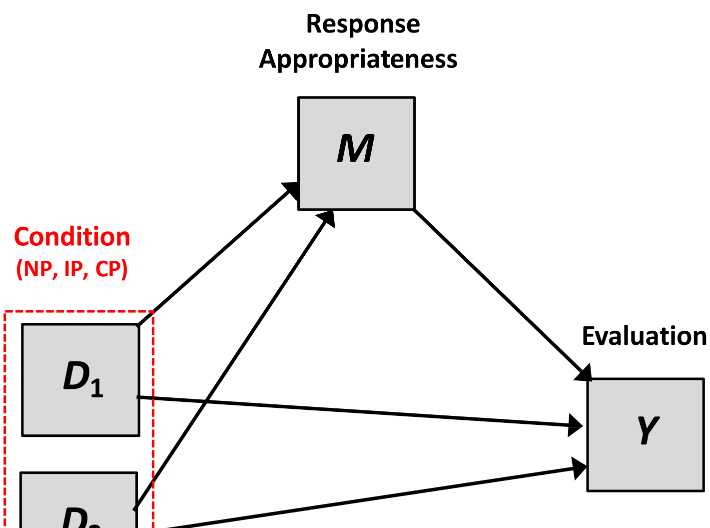
H_0: ${_T}b_2 = 0$
H_a: ${_T}b_2 \neq 0$
$b_2 = 0.443$, $t(126) = 1.986$, $p = .049$

Those told she collectively protested liked her more on average than those told she did not protest.

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.

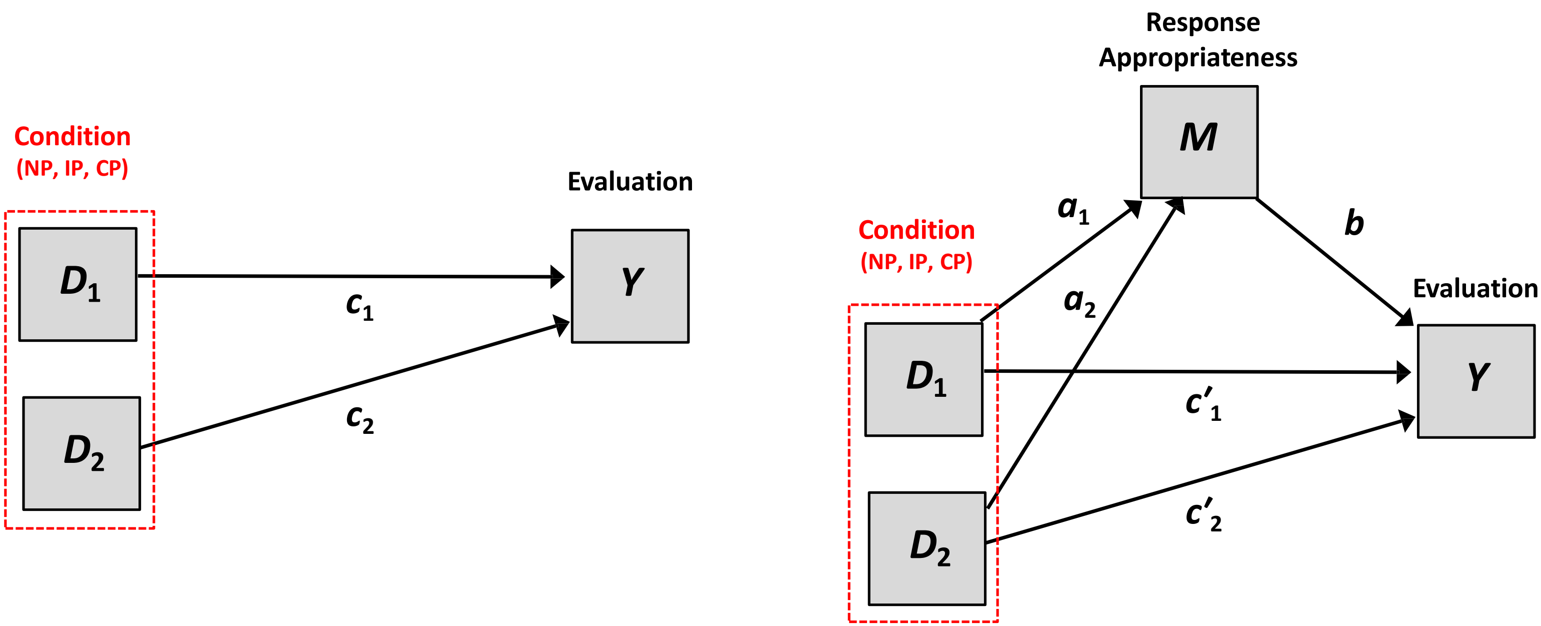
The simple mediation model, the focus of this paper, is diagrammed in Figure 1(b). This model reflects a causal sequence in which *X* affects *Y* indirectly through mediator variable *M*. In this model, *X* is postulated to affect *M*, and this effect then propagates causally to *Y*. This *indirect effect* 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).

Hayes and Preacher (2014, BJMSP). Also see Chapter 6 of IMMCPA2

(Relative) total, direct, and indirect effects



c_1 and c_2: *Relative* total effects of experimental condition on evaluation of the attorney
c'_1 and c'_2: *Relative* direct effects of experimental condition on evaluation of the attorney
a_1b and a_2b: *Relative* indirect effects of condition on evaluation 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.

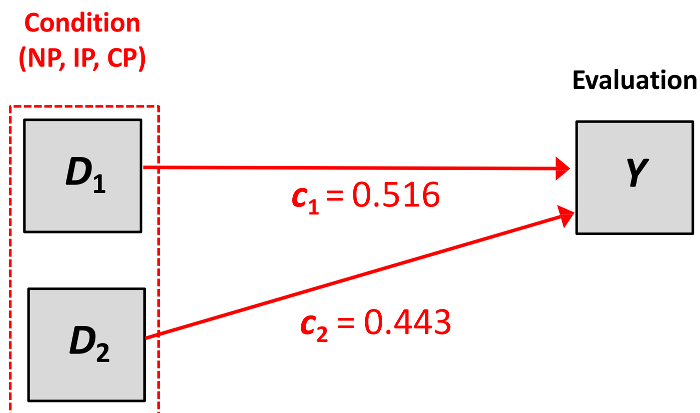
The total effect of experimental condition on evaluation (*c* paths)

```
compute d1 = (protest=1).
compute d2 = (protest=2).
regression/dep = eval/method = enter d1 d2.
```

```
data lawyer2;set lawyer2;
d1 = (protest=1);d2 = (protest=2);run;
proc reg data=lawyer2;model eval=d1 d2;run;
```

```
d1<-as.numeric(lawyer2$protest==1)
d2<-as.numeric(lawyer2$protest==2)
lawyer2.dummy<-data.frame(lawyer2,d1,d2)
summary(lm(eval~d1+d2,data=lawyer2.dummy))
```

We did this already!




```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.3102     0.1614  32.908   <2e-16 ***
d1            0.5158     0.2255   2.287   0.0239 *
d2            0.4431     0.2231   1.986   0.0492 *
```

Relative total effects

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

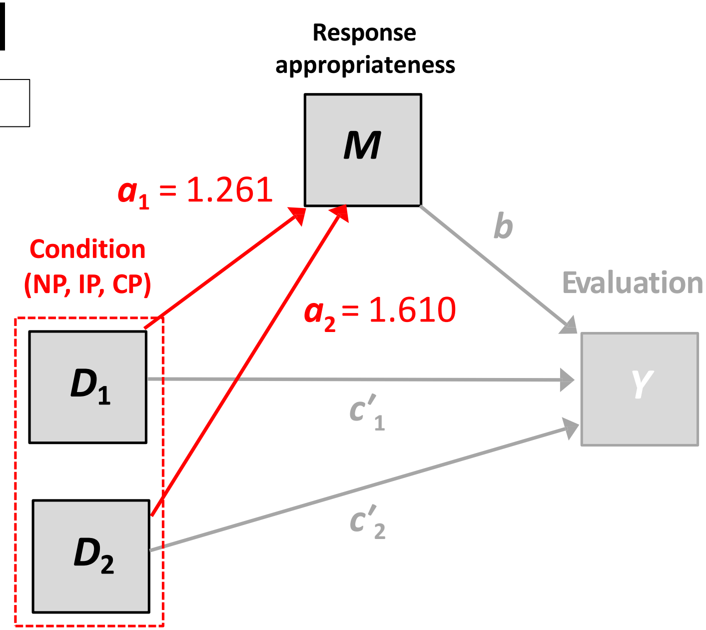
The effect of experimental condition on perceived response appropriateness (*a* paths)

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

```
proc reg data=lawyer2;model approp=d1 d2;run;
```

```
summary(lm(approp~d1+d2,data=lawyer2.dummy))
```

The code above assumes you already created d1 and d2 as in the prior slide and they are available in the data file/frame.




```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   3.8841     0.1825  21.288  < 2e-16 ***
d1            1.2612     0.2550   4.946 2.37e-06 ***
d2            1.6103     0.2522   6.384 3.01e-09 ***
```

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 evaluation (c' paths)

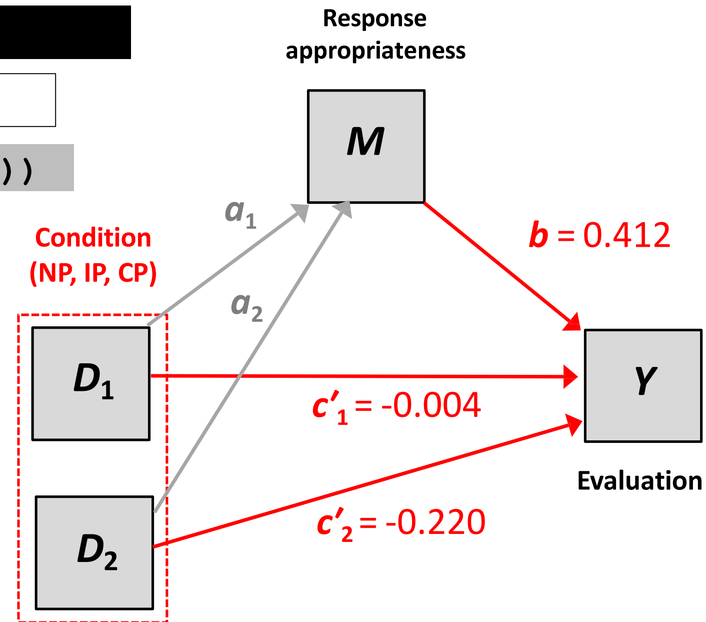
along with the effect of response appropriateness on evaluation (b path)

```
regression/dep = eval/method = enter approp d1 d2.
```

```
proc reg data=lawyer2;model eval=approp d1 d2;run;
```

```
summary(lm(eval~approp+d1+d2,data=lawyer2.dummy))
```

The code above assumes you already created d1 and d2 and they are available in the data file/data frame



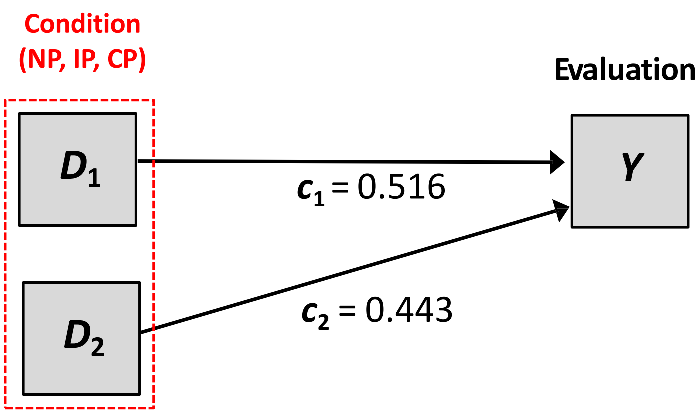
b path

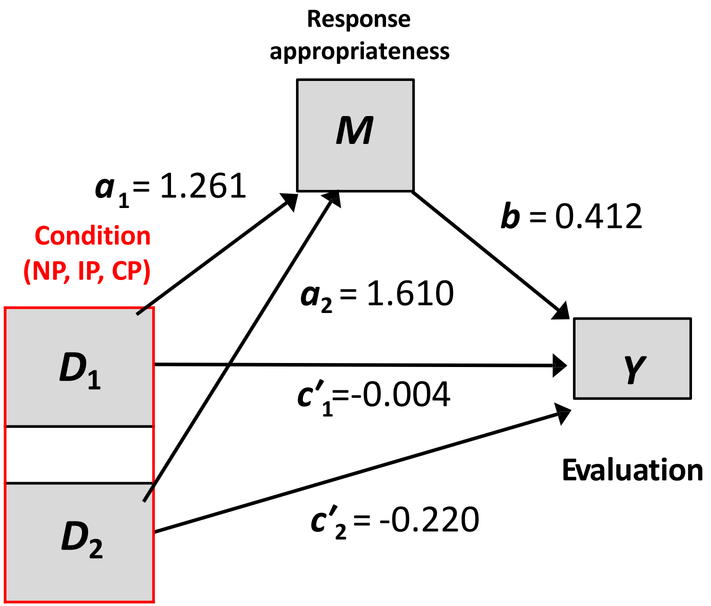
```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.710325   0.307372  12.071  < 2e-16 ***
approp       0.411910   0.070000   5.884 3.43e-08 ***
d1          -0.003699   0.218963  -0.017    0.987
d2          -0.220208   0.228004  -0.966    0.336
---
```

Relative direct effects

Controlling for perceived responses appropriateness, those told she individually protested did not evaluate her more positively, on average, than those told she did not protest (c'_1 = -0.004, p = .987). And those told she collectively protested did not evaluate her any more positively, 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 evaluated her more positively (b = 0.412).

(Relative) total, direct, and indirect effects





c_1 and c_2: *Relative* total effects of condition on evaluation ($c_1 = 0.516$, $c_2 = 0.443$).
c'_1 and c'_2: *Relative* direct effects of condition on evaluation ($c'_1 = -0.004$, $c'_2 = -0.220$).
a_1b and a_2b: *Relative* indirect effects of condition on evaluation 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 total, direct, and indirect effects

	Perceived Response Appropriateness (M)		Evaluation (Y)		
	M	SD	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

$a_1 = 1.261$ $c_1 = 0.516$ $c'_1 = -0.004$

$a_2 = 1.610$ $c_2 = 0.443$ $c'_2 = -0.220$

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

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

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

Adjusted means

$$\hat{Y}_i = 3.710 - 0.004D_{1i} - 0.220D_{2i} + 0.412M_i$$

	Perceived Response Appropriateness (M)		Liking (Y)		
	M	SD	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

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

No protest $\quad Y = 3.710 - 0.004(0) - 0.220(0) + 0.412(4.866) = 5.715 = Y^*_{NP}$

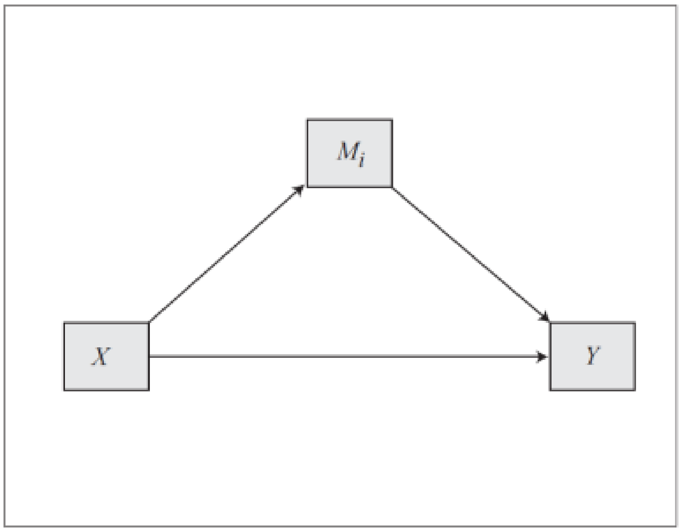
Individual protest $\quad Y = 3.710 - 0.004(1) - 0.220(0) + 0.412(4.866) = 5.711 = Y^*_{IP}$

Collective protest $\quad Y = 3.710 - 0.004(0) - 0.220(1) + 0.412(4.866) = 5.495 = Y^*_{CP}$

Estimation using PROCESS

PROCESS has an option for specifying *X* as a multicategorical variable with up to 9 categories. Four options are available for coding the groups.

Model 4



MCX=1 tells PROCESS that X is a multicategorical variable and to use dummy coding to represent the groups. Other coding options are available. See an excerpt from the PROCESS documentation in your course book.

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

```
process y=eval/m=approp/x=protest/mcx=1/model=4/total=1/seed=61235.
```

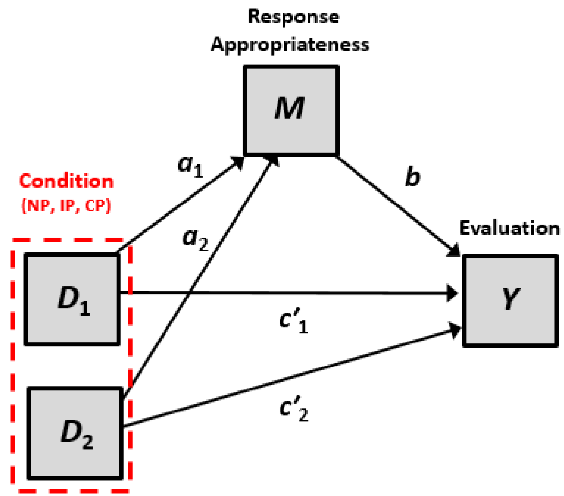
```
%process (data=lawyer2,y=eval,m=approp,x=protest,mcx=1,model=4,total=1,seed=61235)
```

```
process(data=lawyer2,y="eval",m="approp",x="protest",mcx=1,model=4,total=1,
  seed=61235)
```

PROCESS output

```
Model  : 4
    Y  : eval
    X  : protest
    M  : approp

Sample
Size:  129

Custom
Seed:  61235

Coding of categorical X variable for analysis:
 protest       X1       X2
     .000     .000     .000
    1.000    1.000     .000
    2.000     .000    1.000
```

X1 codes individual protest, X2 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 VARIABLE:
 approp
```

$$\widehat{M_i} = 3.884 + 1.261D_{1i} + 1.610D_{2i}$$

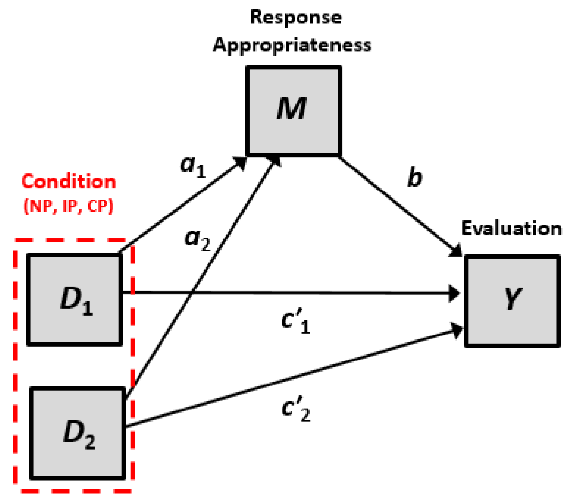
Model Summary

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	
X1	1.2612	.2550	4.9456	.0000	.7565	1.7659	a_1 path
X2	1.6103	.2522	6.3842	.0000	1.1111	2.1095	a_2 path

PROCESS output



$$\hat{Y}_i = 3.710 - 0.004D_{1i} - 0.220D_{2i} + 0.412M_i$$

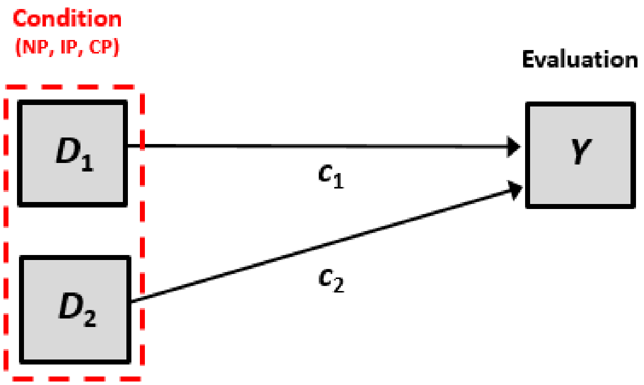
```
**************************************************************************
OUTCOME VARIABLE:
 eval

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
X1           -.0037      .2190     -.0169      .9865     -.4371      .4297
X2           -.2202      .2280     -.9658      .3360     -.6715      .2310
approp        .4119      .0700     5.8844      .0000      .2734      .5504
```

c'_1 path

c'_2 path

b path

PROCESS output




```
************************** TOTAL EFFECT MODEL ****************************
OUTCOME VARIABLE:
 eval

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
X1            .5158      .2255     2.2870      .0239      .0695      .9621
X2            .4431      .2231     1.9863      .0492      .0016      .8845
```

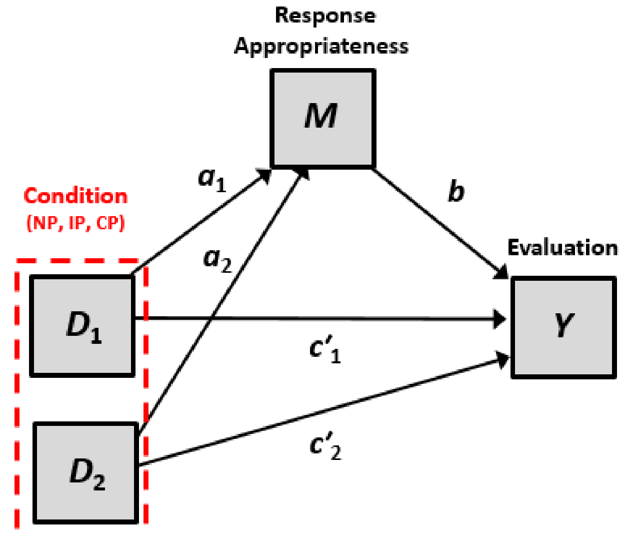
c_1 path

c_2 path

$$\widehat{Y}_i = 5.310 + 0.516D_{1i} + 0.443D_{2i}$$

PROCESS output

```
************* TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Relative total effects of X on Y:
         Effect         se          t          p       LLCI       ULCI
X1        .5158      .2255     2.2870      .0239      .0695      .9621
X2        .4431      .2231     1.9863      .0492      .0016      .8845

Omnibus test of total effect of X on Y:
  R2-chng          F        df1        df2          p
    .0463     3.0552     2.0000   126.0000      .0506

Relative direct effects of X on Y
         Effect         se          t          p       LLCI       ULCI
X1       -.0037      .2190     -.0169      .9865     -.4371      .4297
X2       -.2202      .2280     -.9658      .3360     -.6715      .2310

Omnibus test of direct effect of X on Y:
     R2-chng          F        df1        df2          p
       .0087      .7286     2.0000   125.0000      .4846

Relative indirect effects of X on Y

 protest    ->    approp    ->    eval

         Effect     BootSE   BootLLCI   BootULCI
X1        .5195      .1517      .2588      .8542
X2        .6633      .1684      .3682     1.0279
```



Indirect effect a_1b with bootstrap confidence interval

Indirect effect a_2b with bootstrap confidence interval

Those told she individually protested evaluated her more positively those told she did not protest because protesting was perceived as more appropriate than not, which in turn enhanced evaluation (a_1b = 0.520, 95% CI: 0.259 to 0.854). There is no direct effect of individually protesting on evaluation. Those told she collectively protested evaluated her positively than those told she did not protest because protesting was perceived as more appropriate than not, which in turn enhanced evaluation (a_2b = 0.663, 95% CI: 0.368 to 1.028). There is no direct effect of collectively protesting on evaluation.

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 OF X ON Y **************

Relative total effects of X on Y:
        Effect          se           t            p         LLCI        ULCI
X1       .5158       .2255      2.2870        .0239        .0695       .9621
X2       .4431       .2231      1.9863        .0492        .0016       .8845

Omnibus test of total effect of X on Y:
     R2-chng           F         df1          df2            p
       .0463      3.0552      2.0000     126.0000        .0506
----------------

Relative direct effects of X on Y
        Effect          se           t            p         LLCI        ULCI
X1      -.0037       .2190      -.0169        .9865       -.4371       .4297
X2      -.2202       .2280      -.9658        .3360       -.6715       .2310

Omnibus test of direct effect of X on Y:
     R2-chng           F         df1          df2            p
       .0087       .7286      2.0000     125.0000        .4846
----------------

Relative indirect effects of X on Y

protest      ->     approp     ->      eval

         Effect       BootSE  BootLLCI   BootULCI
     X1   .5195          .1517  .2588       .8542
     X2   .6633          .1684  .3682      1.0279
```

Test of the "omnibus" total effect.

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

Omnibus inference

PROCESS gives us tests of the $k - 1$ relative direct effects. It also provides a test of equality of the k group adjusted means on Y when the mediator is held constant---the "omnibus" direct effect. This is equivalent to a single-factor ANCOVA.

```
 ************* TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Relative total effects of X on Y:
          Effect         se          t           p        LLCI        ULCI
 X1        .5158      .2255     2.2870       .0239       .0695       .9621
 X2        .4431      .2231     1.9863       .0492       .0016       .8845

Omnibus test of total effect of X on Y:
  R2-chng          F        df1        df2          p
   .0463      3.0552     2.0000   126.0000      .0506

Relative direct effects of X on Y
          Effect         se          t           p        LLCI        ULCI
 X1       -.0037      .2190     -.0169       .9865      -.4371       .4297
 X2       -.2202      .2280     -.9658       .3360      -.6715       .2310

 Omnibus test of direct effect of X on Y:
     R2-chng          F        df1        df2          p
       .0087      .7286     2.0000   125.0000      .4846
 ----------

 Relative indirect effects of X on Y

 protest         -> approp      ->      eval

          Effect         BootSE  BootLLCI    BootULCI
     X1    .5195          .1517   .2588        .8542
     X2    .6633          .1684   .3682       1.0279
```

Test of the "omnibus" direct effect.

The three conditions do not differ on average in how positively they evaluated her after accounting for group differences in perceived response appropriateness, $\phi R^2 = 0.009$, $F(2,125) = 0.729$, $p = .485$. ϕR^2 is the change in R^2 when the $k - 1$ variables coding group are added to the model of Y that already contains the mediator.

Omnibus inference about the indirect effect

- The omnibus tests for the total and direct effect of *X* are not dependent on the system used for coding the groups, even though the relative direct and total effects are.
- Our rule that *X* indirectly affects *Y* if at least one relative indirect effect is different from zero means our conclusion will depend on the system used for coding groups, since the relative indirect effects are dependent on that choice.
- If all of the bootstrap confidence intervals for the relative indirect effects straddle zero, that does NOT mean *X* does not indirectly affect *Y*. It could be that a different coding choice produces a different outcome.
- Our rule can confirm that *X* indirect affects *Y* if at least one relative indirect effect is different from zero. But a failure to meet this criterion does not disconfirm the existence of an indirect effect of *X* on *Y* through *M*.
- Moral: Choose your coding system wisely, so that it produces relative indirect effects you care about and that are sensitive to the question you are trying to answer.
- There are omnibus tests of the indirect effect that are not sensitive to the coding choice. This must be done in SEM and requires problematic assumptions.

A different coding system

Other systems for coding groups can be used. For instance, we might instead choose to estimate the direct and indirect effects of protesting (regardless of form) relative to not, and the effects of collectively protesting relative to individually protesting.

Condition	H_1	H_2
No protest	-2/3	0
Individual	1/3	-1/2
Collective	1/3	1/2

You may recognize these as two *orthogonal contrasts*. It is also called "Helmert coding" when the multicategorical variable is ordinal.

Effects for H_1 will compare no protest to the average of the two protest conditions, and effects for H_2 will compare collective protest to individual protest.

Creating the codes in SPSS, SAS, and R

	Condition	H_1	H_2
protest = 0	**No protest**	-2/3	0
protest = 1	**Individual**	1/3	-1/2
protest = 2	**Collective**	1/3	1/2

```
if (protest = 0) h1 = -2/3.
if (protest > 0) h1 = 1/3.
if (protest = 0) h2 = 0.
if (protest = 1) h2 = -1/2.
if (protest = 2) h2 = 1/2.
```

```
data lawyer2;set lawyer2;
  if (protest = 0) then h1 = -2/3;
  if (protest > 0) then h1 = 1/3;
  if (protest = 0) then h2 = 0;
  if (protest = 1) then h2 = -1/2;
  if (protest = 2) then h2 = 1/2;
run;
```

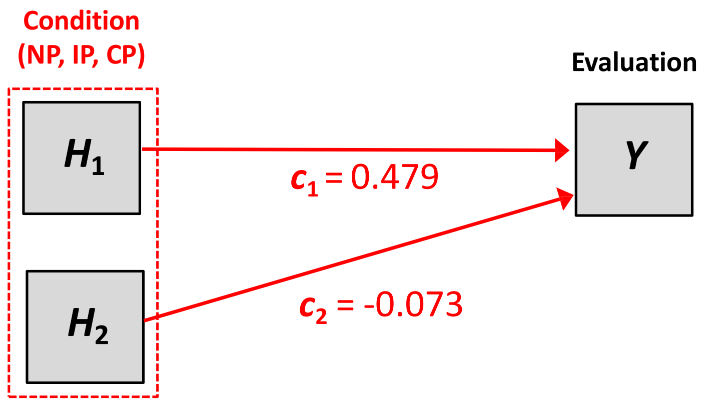
```
h1<-(lawyer2$protest==0)*(-2/3)+(lawyer2$protest>0)*(1/3)
h2<-(lawyer2$protest==1)*(-1/2)+(lawyer2$protest==2)*(1/2)
lawyer2.helmert<-data.frame(lawyer2,h1,h2)
```

The total effect of experimental condition on evaluation (c paths)

```
regression/dep = eval/method = enter h1 h2.
```

```
proc reg data=lawyer2;model eval=h1 h2;run;
```

```
summary(lm(eval~h1+h2,data=lawyer2.helmert))
```




```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.62987    0.09104  61.841   <2e-16 ***
h1           0.47945    0.19539   2.454   0.0155 *
h2          -0.07271    0.22034  -0.330   0.7419
```

Relative total effects

$$c_1 = \frac{(\bar{Y}_{IP} + \bar{Y}_{CP})}{2} - \bar{Y}_{NP} = \frac{(5.826 + 5.753)}{2} - 5.310 = 5.789 - 5.310 = 0.479$$

$$c_2 = \bar{Y}_{CP} - \bar{Y}_{IP} = 5.753 - 5.826 = -0.073$$

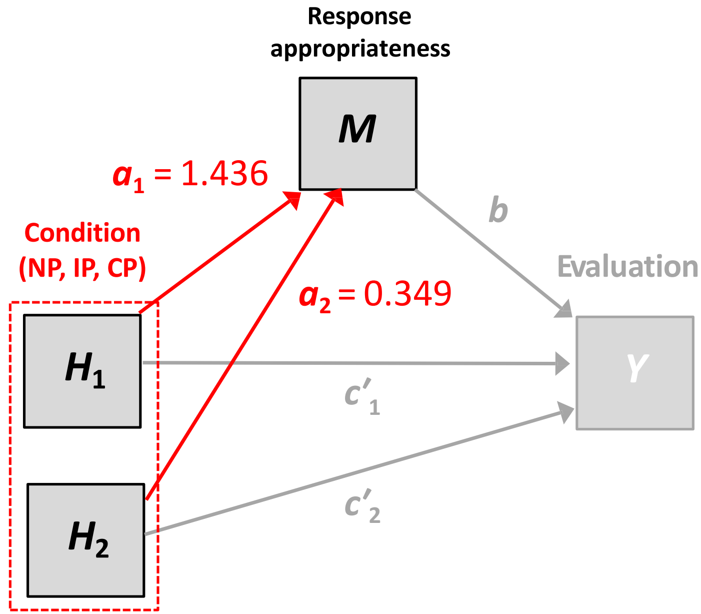
Relative to those told she did not protest, those told she protested evaluated her more positively on average (c_1 = 0.479, p = .016). Those told she collectively protested did not differ, on average, in how positively they evaluated her relative to those told she individually protested (c_2 = -0.073, p = .742).

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

```
regression/dep = approp/method = enter h1 h2.
```

```
proc reg data=lawyer2;model approp=h1 h2;run;
```

```
summary(lm(approp~h1+h2,data=lawyer2.helmert))
```




```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.8413     0.1029  47.032  < 2e-16 ***
h1            1.4358     0.2209   6.499 1.71e-09 ***
h2            0.3491     0.2491   1.401    0.164
```

$$a_1 = \frac{(\overline{M}_{IP} + \overline{M}_{CP})}{2} - \overline{M}_{NP} = \frac{(5.145 + 5.494)}{2} - 3.884 = 5.320 - 3.884 = 1.436$$

$$a_2 = \overline{M}_{CP} - \overline{M}_{IP} = 5.494 - 5.145 = 0.349$$

Relative to those told she did not protest, those told she protested felt this was a more appropriate response, on average (a_1 = 1.436, p < .001). Those told she collectively protested did not perceive this as any more or less appropriate, on average, relative to those told she individually protested (a_2 = 0.349, p = .164).

The direct effect of condition on evaluation (*c′* paths)

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

```
regression/dep = eval/method = enter approp h1 h2.
```

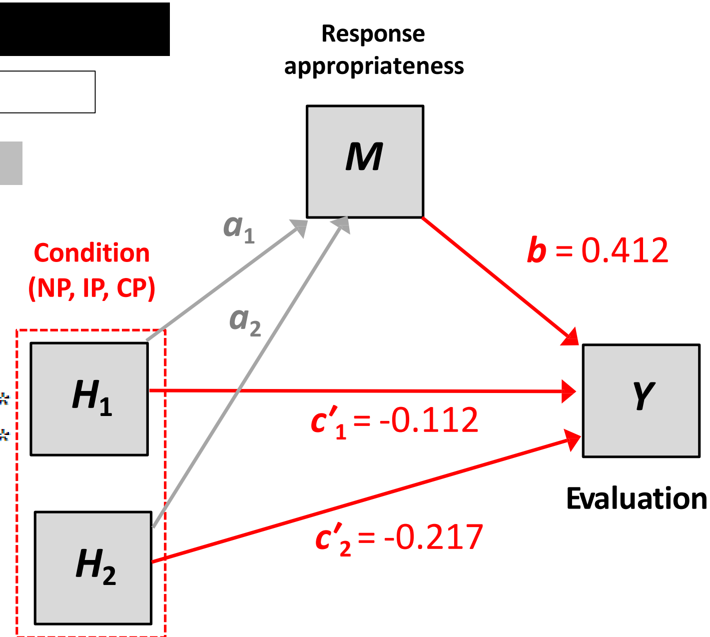
```
proc reg data=lawyer2;model eval=approp h1 h2;run;
```

```
summary(lm(eval~approp+h1+h2,data=lawyer2.helmert))
```

b path

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.6357     0.3484  10.435  < 2e-16 ***
approp        0.4119     0.0700   5.884 3.43e-08 ***
h1           -0.1119     0.2006  -0.558    0.578
h2           -0.2165     0.1973  -1.097    0.275
---
```

Relative direct effects

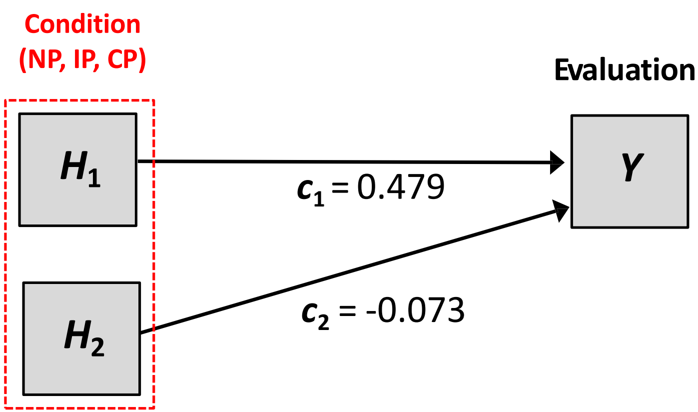


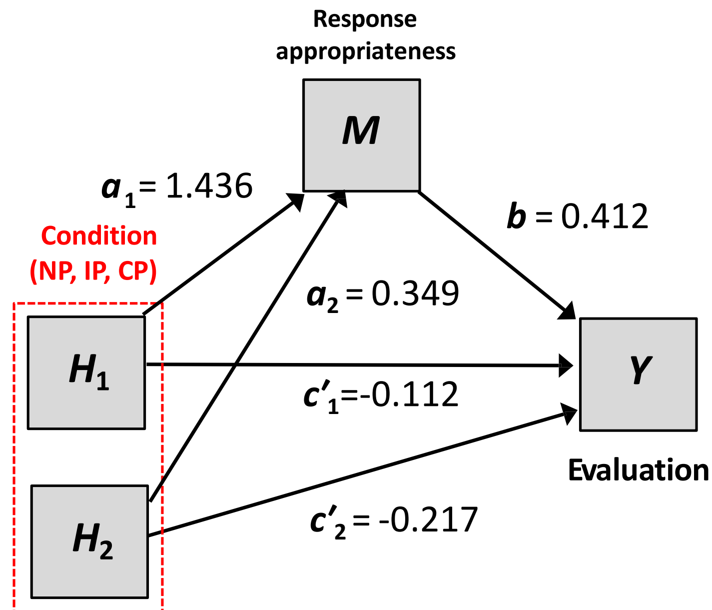
$$c'_1 = \frac{(\bar{Y}^*_{IP} + \bar{Y}^*_{CP})}{2} - \bar{Y}^*_{NP} = \frac{(5.711 + 5.495)}{2} - 5.715 = 5.603 - 5.715 = -0.112$$

$$c'_2 = \bar{Y}^*_{CP} - \bar{Y}^*_{IP} = 5.495 - 5.711 = -0.217$$

Controlling for perceived responses appropriateness, those told she protested did not evaluated her any more positively on average than those told she did not protest (c'_1 = -0.112, p = .578). And those told she collectively protested did not evaluated her more positively on average those told she individually protested (c'_2 = -0.217, p = .275). Holding condition constant, those who perceived her behavior as relatively more appropriate evaluated her more positively (b = 0.412).

(Relative) total, direct, and indirect effects



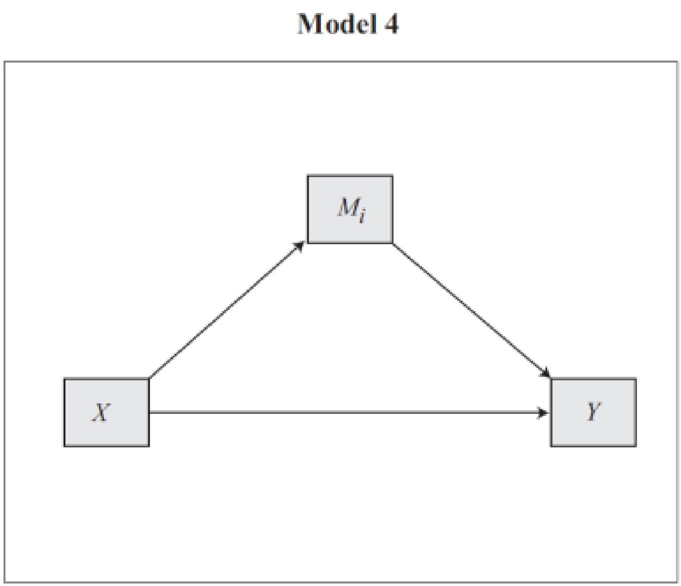


c_1 and c_2: *Relative* total effects of condition on evaluation ($c_1 = 0.479$, $c_2 = -0.073$).
c'_1 and c'_2: *Relative* direct effects of condition on evaluation ($c'_1 = -0.112$, $c'_2 = -0.217$).
a_1b and a_2b: *Relative* indirect effects of condition on evaluation through perceived response appropriateness, $a_1b = 1.436(0.412) = 0.591$, $a_2b = 0.349(0.412) = 0.144$

$$c_1 = c'_1 + a_1b: \quad 0.479 = -0.112 + 1.436(0.412) = -0.112 + 0.591$$
$$c_2 = c'_2 + a_2b: \quad -0.073 = -0.217 + 0.349(0.412) = -0.217 + 0.144$$

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.

Estimation using PROCESS



PROCESS has an option for specifying *X* as a multicategorical variable with up to 9 categories. Four options are available for coding the groups.

MCX=3 tells PROCESS that X is a multicategorical variable and to use “Helmert coding” to represent the groups. This is equivalent to the orthogonal contrasts we set up manually in this example.

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

```
process y=eval/m=approp/x=protest/mcx=3/model=4/total=1/seed=61235.
```

```
%process (data=lawyer2,y=eval,m=approp,x=protest,mcx=3,model=4,total=1,seed=61235)
```

```
process(data=lawyer2,y="eval",m="approp",x="protest",mcx=3,model=4,total=1,
   seed=61235)
```

PROCESS output

```
Model  : 4
    Y  : eval
    X  : protest
    M  : approp

Sample
Size:  129

Coding of categorical X variable for analysis:
protest      X1      X2
    .000   -.667    .000
   1.000    .333   -.500
   2.000    .333    .500
```

X1 codes protest versus no protest. X2 codes collective versus individual protest.

```
**************************************************************************
OUTCOME VARIABLE:
 approp
```

$$\widehat{M_{1i}} = 4.841 + 1.436H_{1i} + 0.349H_{2i}$$

```
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     4.8413      .1029    47.0321      .0000     4.6376     5.0450
X1           1.4358      .2209     6.4988      .0000      .9985     1.8730
X2            .3491      .2491     1.4012      .1636     -.1440      .8421
```

a_1 path

a_2 path

```
**************************************************************************
OUTCOME VARIABLE:
 eval
```

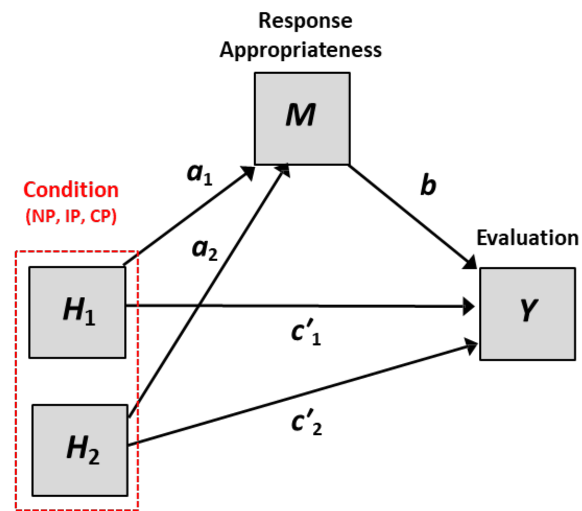
$$\widehat{Y_i} = 3.636 - 0.112H_{1i} - 0.217H_{2i} + 0.412M_i$$

```
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     3.6357      .3484    10.4351      .0000     2.9461     4.3252
X1           -.1120      .2006     -.5581      .5778     -.5089      .2850
X2           -.2165      .1973    -1.0974      .2746     -.6070      .1739
approp        .4119      .0700     5.8844      .0000      .2734      .5504
```

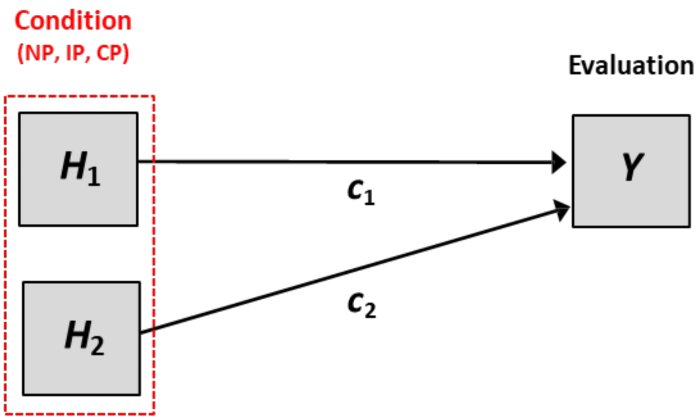
c'_1 path

c'_2 path

b path



PROCESS output




```
************************* TOTAL EFFECT MODEL ***************************
OUTCOME VARIABLE:
 eval

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     5.6299      .0910    61.8414      .0000     5.4497     5.8100
X1            .4794      .1954     2.4538      .0155      .0928      .8661
X2           -.0727      .2203     -.3300      .7419     -.5088      .3633
```

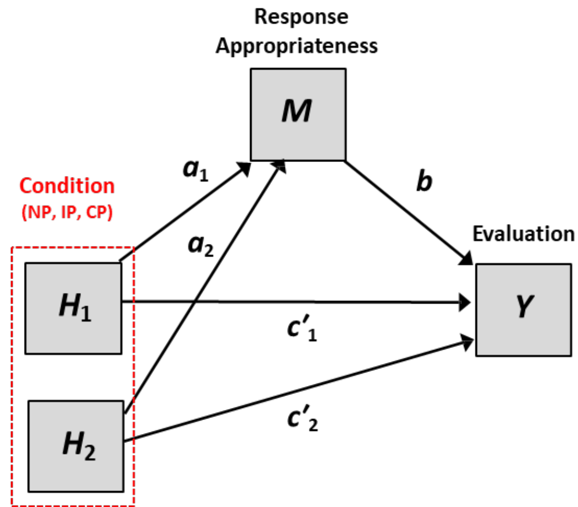
c_1 path

c_2 path

$$\widehat{Y}_i = 5.630 + 0.479H_{1i} - 0.073H_{2i}$$

PROCESS output

```
************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Relative total effects of X on Y:
          Effect         se          t          p       LLCI       ULCI
X1         .4794      .1954     2.4538      .0155      .0928      .8661
X2        -.0727      .2203     -.3300      .7419     -.5088      .3633

Omnibus test of total effect of X on Y:
  R2-chng           F        df1        df2          p
   .0463       3.0552     2.0000   126.0000      .0506
----------

Relative direct effects of X on Y
          Effect         se          t          p       LLCI       ULCI
X1        -.1120      .2006     -.5581      .5778     -.5089      .2850
X2        -.2165      .1973    -1.0974      .2746     -.6070      .1739

Omnibus test of direct effect of X on Y:
     R2-chng          F        df1      df2          p
      .0087       .7286     2.0000  125.0000      .4846
----------

Relative indirect effects of X on Y

 protest       ->    approp    ->      eval

         Effect     BootSE   BootLLCI   BootULCI
X1        .5914      .1531      .3272      .9330
X2        .1438      .0948     -.0283      .3477
```



Indirect effect a_1b with bootstrap confidence interval

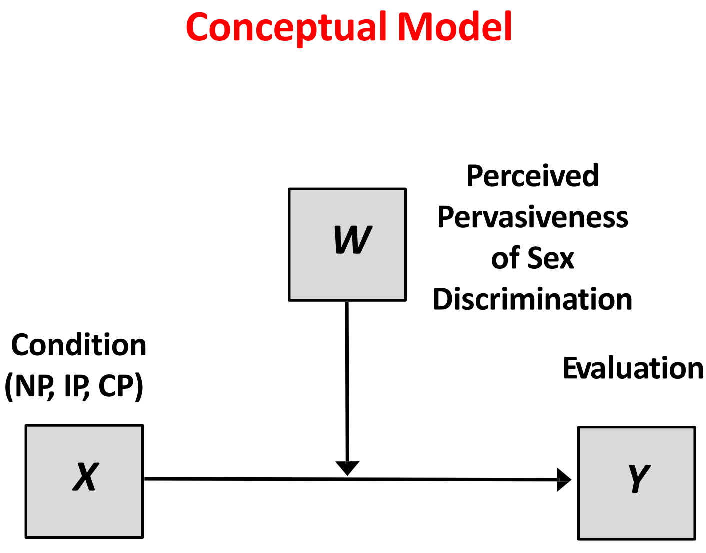
Indirect effect a_2b with bootstrap confidence interval

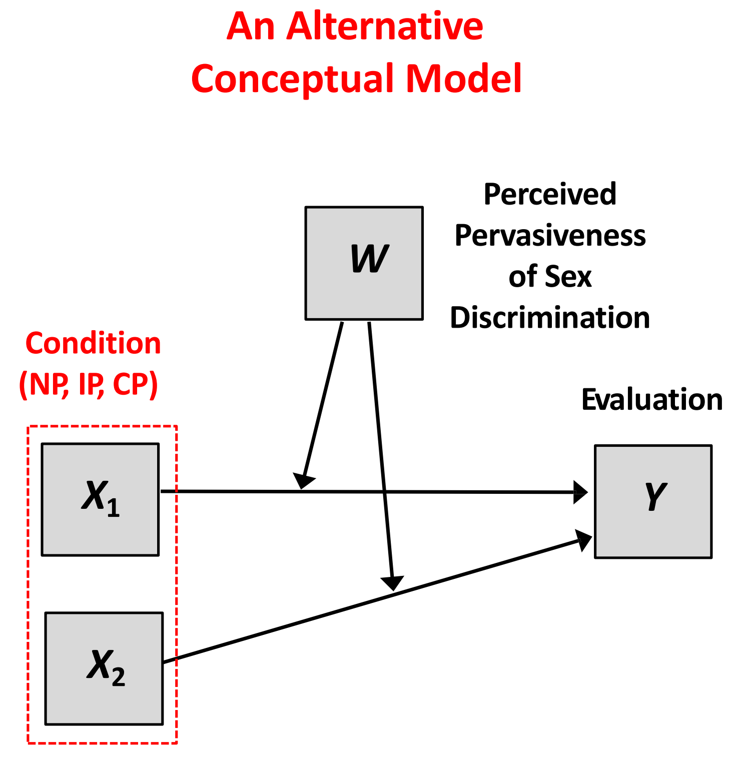
Those told she protested liked her more than those told she did not protest because protesting was perceived as more appropriate than not, which in turn enhanced evaluation (relative indirect effect = 0.591, 95% CI: 0.327 to 0.933). The was no direct effect of protesting relative to not on evaluation (relative direct effect = -0.112, p = .578). The relative direct and indirect effects of collectively protesting relative to individually protesting were not significantly different from zero (relative indirect effect = 0.144, 95% CI = -0.028 to 0.348; relative direct effect = -0.217, p = .274)

Multicategorical Variables in Moderation

Moderation with a multicategorical focal predictor

Does the effect of the lawyer's behavior on how positively she is evaluated depend on the perceiver's belief about how prevalent sex discrimination is in society?





Testing a hypothesis about moderation that includes a multicategorical predictor defined by *k* groups requires an integration of lessons about moderation and the use of a group coding system.

Moderation involving a multicategorical focal predictor

Testing a hypothesis about moderation that includes a multicategorical predictor *X* defined by *k* groups requires the use of a group coding system. In its most general form, for a dichotomous or continuous moderator *W*, estimating such a moderation model requires the addition of *k* – 1 variables coding group and *k* – 1 product variable to capture the moderation by *W*:

$$\hat{Y}_i = b_0 + \underbrace{\boxed{\sum_{k=1}^{K-1} b_k X_{ki}}}_{X} + \underbrace{\boxed{b_K W_i}}_{W} + \underbrace{\boxed{\sum_{k=1}^{K-1} b_{K+k} X_{ki} W_i}}_{XW}$$

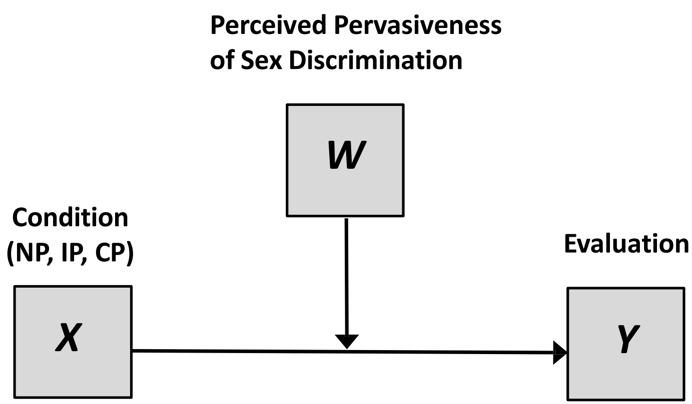
Example

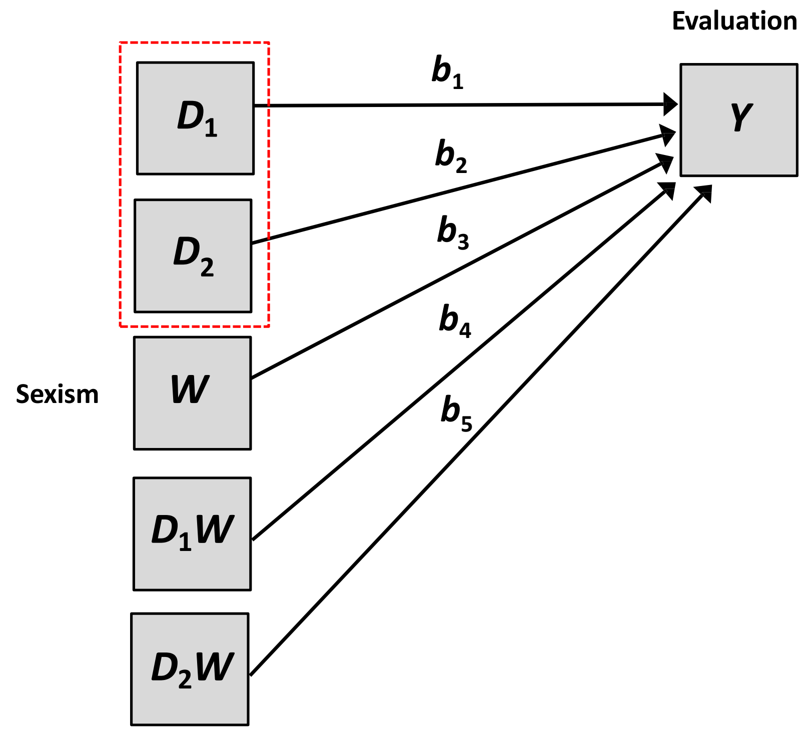
We've seen that whether the attorney protested individually, collectively, or not at all influenced how positively she was perceived. We now ask whether this effect depends on the perceiver's beliefs about how pervasive sex discrimination is in society (*W*). We will use the same indicator coding system, with the no protest group as the reference

$$\hat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 W_i + b_4 D_{1i} W_i + b_5 D_{2i} W_i$$

The conceptual and statistical models

$$\widehat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 W + b_4 D_{1i} W_i + b_5 D_{2i} W_i$$





Before estimating this, let's look at what a model without the product terms looks like.

An unconditional effects model (no moderation)

```
compute d1 = (protest = 1).
compute d2 = (protest = 2).
regression/dep = eval/method = enter d1 d2 sexism.
```

```
data lawyer2;set lawyer2;
d1=(protest=1);d2=(protest=2);run;
proc reg;model eval = d1 d2 sexism;run;
```

```
d1<-as.numeric(lawyer2$protest==1)
d2<-as.numeric(lawyer2$protest==2)
lawyer2.dummy<-data.frame(lawyer2,d1,d2)
summary(lm(eval~d1+d2+sexism,data=lawyer2.dummy))
```

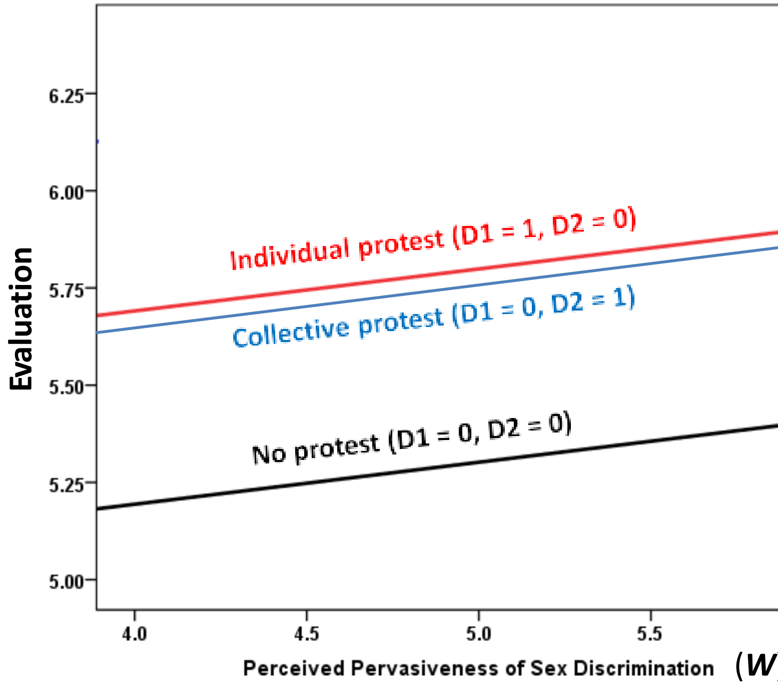
$$\widehat{Y}_i = 4.762 + 0.497D_{1i} + 0.447D_{2i} + 0.108W_i$$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.7617     0.6168   7.720 3.24e-12 ***
d1            0.4970     0.2266   2.193   0.0301 *
d2            0.4468     0.2232   2.001   0.0475 *
sexism        0.1082     0.1174   0.921   0.3586
```

Visualizing the model

$$\widehat{Y}_i = 4.762 + 0.497D_{1i} + 0.447D_{2i} + 0.108W_i$$

D_1	D_2	W	$\hat{Y}$
0	0	4	5.19
0	0	5	5.30
0	0	6	5.41
1	0	4	5.69
1	0	5	5.80
1	0	6	5.91
0	1	4	5.64
0	1	5	5.75
0	1	6	5.86



This model constrains the difference between the three groups in their evaluation of the attorney to be constant across values of perceived pervasiveness of sex discrimination in society.

Releasing this constraint on the model

Suppose we let each comparison with the reference group be a function of W. For a more versatile model, we allow those functions to differ from each other.

$$\widehat{Y}_i = b_0 + f_1(W_i)D_{1i} + f_2(W_i)D_{2i} + b_3 W_i$$

For instance, let $f_1(W)$ be a linear function of W, $b_1 + b_4 W$, and let $f_2(W)$ be a different linear function of W, $b_2 + b_5 W$. Thus

$$\widehat{Y}_i = b_0 + (b_1 + b_4 W_i)D_{1i} + (b_2 + b_5 W_i)D_{2i} + b_3 W_i$$

which can be rewritten as

$$\widehat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 W_i + b_4 W_i D_{1i} + b_5 W_i D_{2i}$$

X's effect (a 3-category focal predictor) as a function of W

$$\hat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 W_i + b_4 W_i D_{1i} + b_5 W_i D_{2i}$$

can be rewritten as

$$\hat{Y}_i = b_0 + (b_1 + b_4 W_i) D_{1i} + (b_2 + b_5 W_i) D_{2i} + b_3 W_i$$

So the difference in Y between the group coded with D_1 and the reference group depends on W. And the difference in Y in between the group coded with D_2 and the reference group depends on W.

Let $\theta_{D_1 \rightarrow Y} = b_1 + b_4 W$ = be a **relative** conditional effect of X (for D_1 relative to the reference group)

Let $\theta_{D_2 \rightarrow Y} = b_2 + b_5 W$ = be a **relative** conditional effect of X (for D_2 relative to the reference group)

$$\hat{Y}_i = b_0 + \theta_{D_1 \rightarrow Y} D_{1i} + \theta_{D_2 \rightarrow Y} D_{2i} + b_3 W_i$$

A conditional effects model (*X*'s effect moderated by *W*)

```
compute d1w = d1*sexism.
compute d2w = d2*sexism.
regression/dep = eval/method = enter d1 d2 sexism d1w d2w.
```

```
data lawyer2;set lawyer2;
d1=(protest=1);d2=(protest=2);d1w=d1*sexism;d2w=d2*sexism;run;
proc reg data=lawyer2;model eval = d1 d2 sexism d1w d2w;run;
```

```
d1<-as.numeric(lawyer2$protest==1)
d2<-as.numeric(lawyer2$protest==2)
lawyer2.dummy<-data.frame(lawyer2,d1,d2)
summary(lm(eval~d1*sexism+d2*sexism,data=lawyer2.dummy))
```

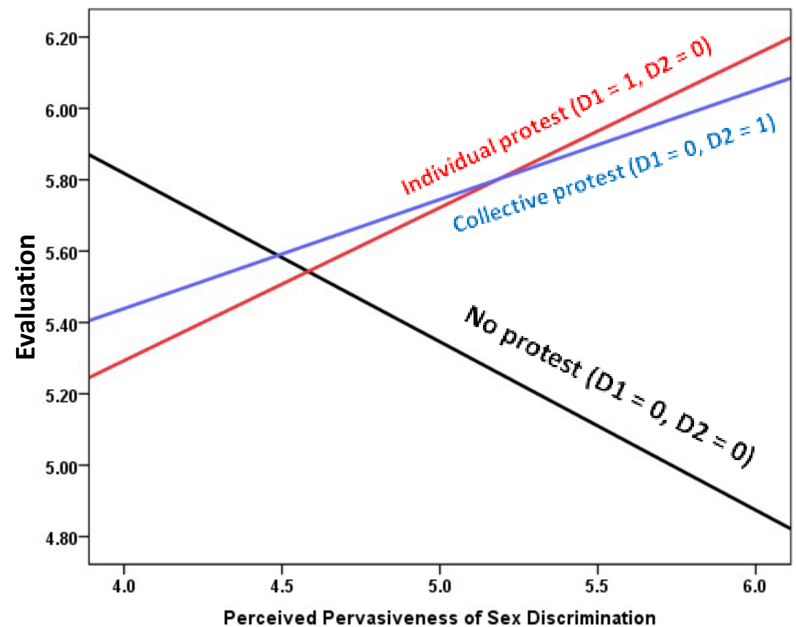
$$\hat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.7062     1.0525   7.322 2.79e-11 ***
d1           -4.1288     1.4985  -2.755  0.00676 **
d2           -3.4908     1.4078  -2.480  0.01451 *
sexism       -0.4725     0.2053  -2.302  0.02303 *
d1:sexism     0.9012     0.2875   3.135  0.00215 **
d2:sexism     0.7778     0.2752   2.827  0.00549 **
```

Constructing a visual representation of the model (SPSS)

```
data list free/d1 d2 w protest.
begin data.
0 0 4 0
0 0 5 0
0 0 6 0
1 0 4 1
1 0 5 1
1 0 6 1
0 1 4 2
0 1 5 2
0 1 6 2
end data.
compute yhat = 7.706-4.129*d1-3.491*d2-0.472*w+
               0.901*d1*w+0.778*d2*w.
graph/scatterplot = w with yhat by protest.
```

$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$

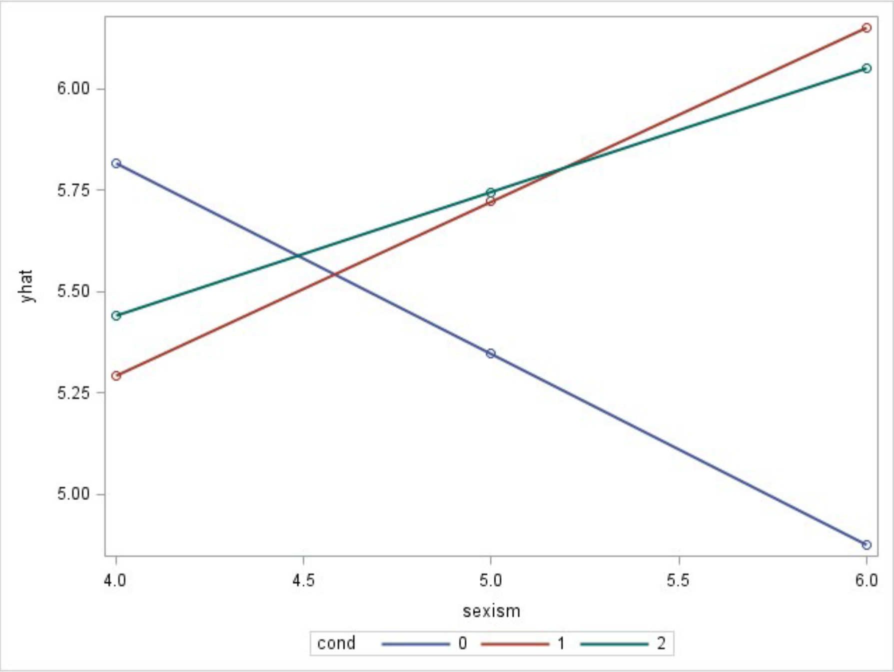


D_1	D_2	W	$\hat{Y}$
0	0	4	5.82
0	0	5	5.35
0	0	6	4.87
1	0	4	5.29
1	0	5	5.72
1	0	6	6.15
0	1	4	5.44
0	1	5	5.75
0	1	6	6.05

This model allows the difference between the three groups to depend on perceived pervasiveness of sex discrimination in society.

Constructing a visual representation of the model (SAS)

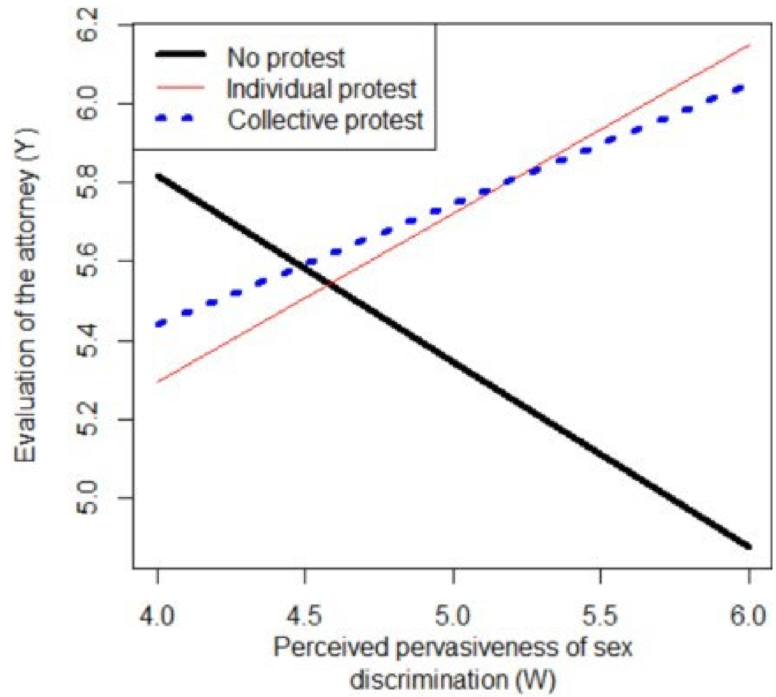
No protest (D1=0,D2=0)
Individual protest (D1=1,D2=0)
Collective protest (D1=0,D2=1)




```
data;
input d1 d2 sexism protest;
yhat = 7.706-4.129*d1-3.491*d2-0.472*sexism+
    0.901*d1*sexism+0.778*d2*sexism;
cards;
 0 0 4 0
 0 0 5 0
 0 0 6 0
 1 0 4 1
 1 0 5 1
 1 0 6 1
 0 1 4 2
 0 1 5 2
 0 1 6 2
run;
proc sgplot;reg x=sexism y=yhat/group=protest;run;
```

Constructing a visual representation of the model (R)

```
x<-c(0,1,2,0,1,2,0,1,2)
w<-c(4,4,4,5,5,5,6,6,6)
d1<-(x==1);
d2<-(x==2);
y<-7.706-4.129*d1-3.491*d2-0.472*w+0.901*d1*w+0.778*d2*w
plot(y=y,x=w,pch=15,col="white",xlab="Perceived pervasiveness of sex
discrimination (W)",ylab="Evaluation of the attorney (Y)")
legend.txt<-c("No protest","Individual protest","Collective protest")
legend("topleft",legend=legend.txt,lty=c(1,1,3),lwd=c(4,1,4),
col=c("black","red","blue"))
lines(w[x==0],y[x==0],lwd=4,lty=1,col="black")
lines(w[x==1],y[x==1],lwd=1,lty=1,col="red")
lines(w[x==2],y[x==2],lwd=4,lty=3,col="blue")
```



Interpretation of the coefficients

$$\widehat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 W_i + b_4 W_i D_{1i} + b_5 W_i D_{2i}$$

can be rewritten as

$$\widehat{Y}_i = b_0 + (b_1 + b_4 W_i) D_{1i} + (b_2 + b_5 W_i) D_{2i} + b_3 W_i$$

b_1 estimates the estimated difference in Y between the group coded $D_1 = 1$ (i.e., individual protest) and the reference group (i.e., no protest), conditioned on W being zero.

b_2 estimates the estimated difference in Y between the group coded $D_2 = 1$ (i.e., collective protest) and the reference group (i.e., no protest), conditioned on W being zero.

b_3 estimates the effect of W in the reference group (i.e., D_1 and $D_2 = 0$, no protest).

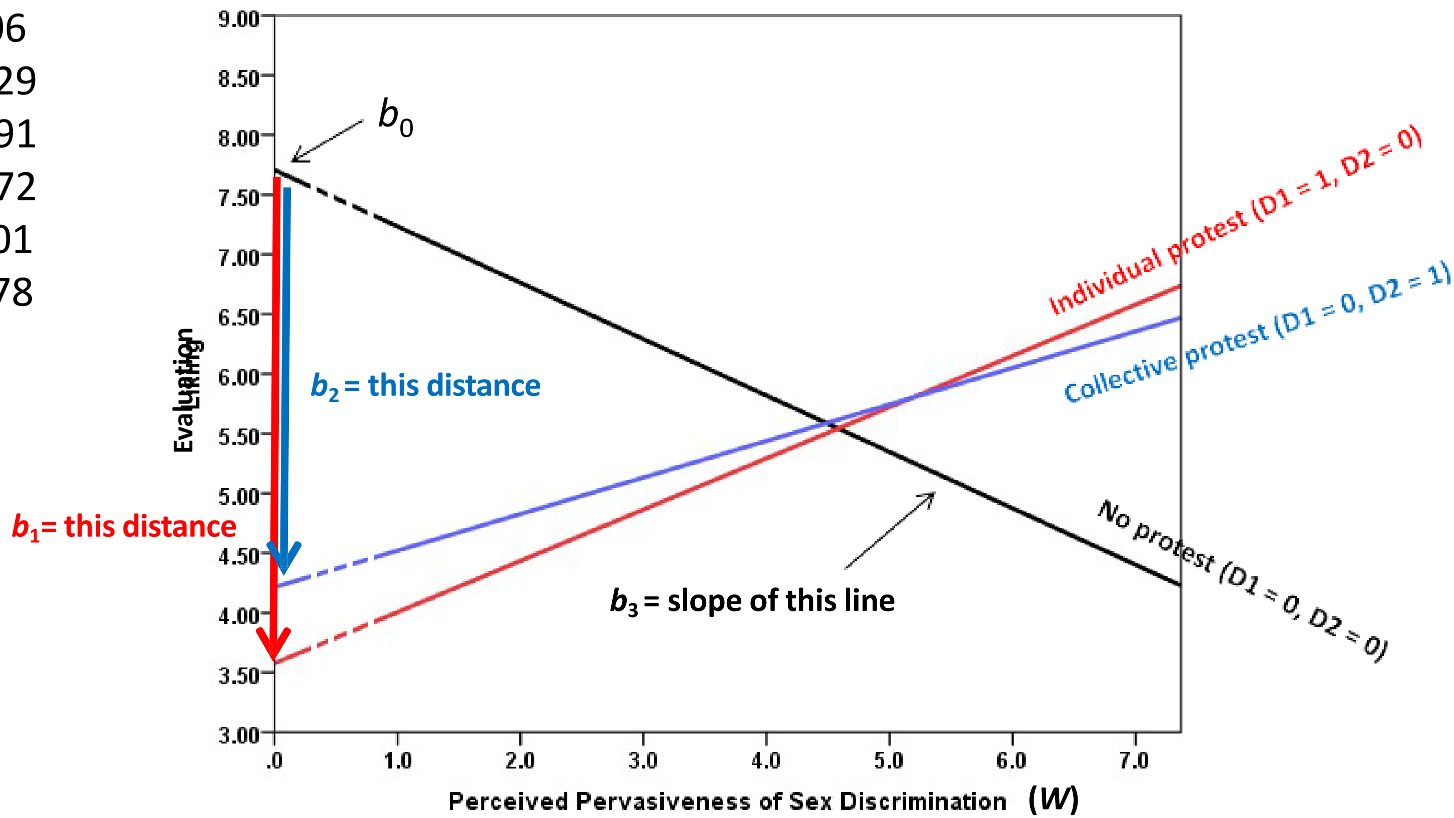
b_4 estimates how much the difference in Y between the group coded $D_1 = 1$ (i.e., individual protest) and the reference group (i.e., no protest) differs for each one unit difference in W.

b_5 estimates how much the difference in Y between the group coded $D_2 = 1$ (i.e., collective protest) and the reference group (i.e., no protest) differs for each one unit difference in W.

Interpretation of the coefficients

$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$

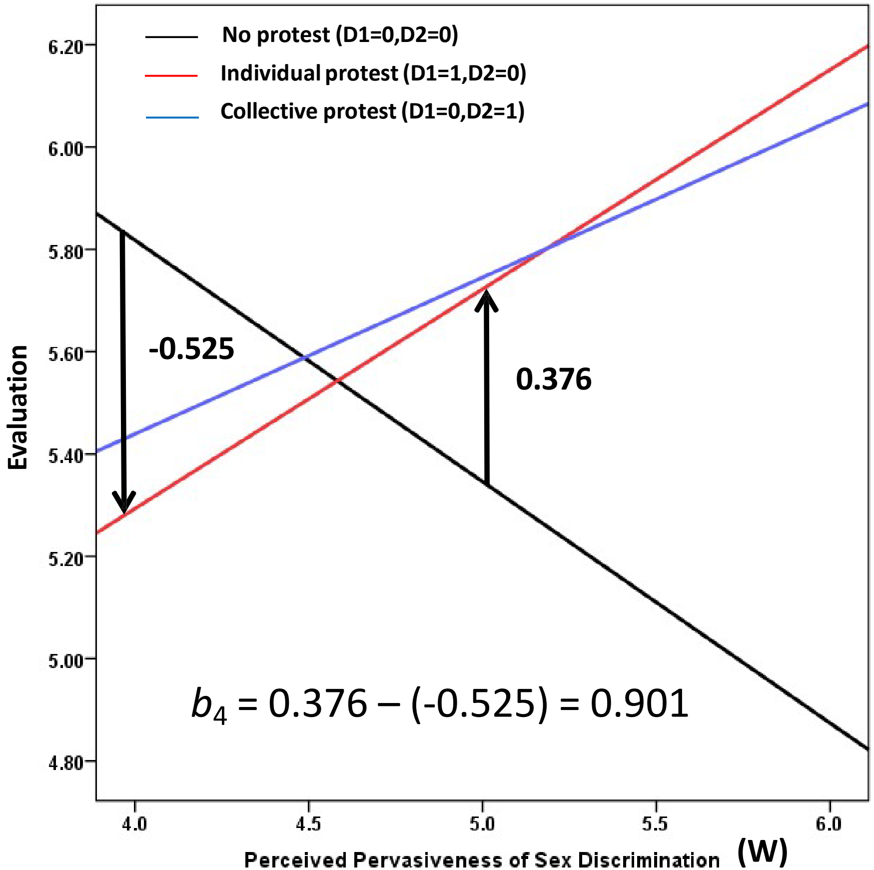
$b_0 =$ 7.706
$b_1 =$ -4.129
$b_2 =$ -3.491
$b_3 =$ -0.472
$b_4 =$ 0.901
$b_5 =$ 0.778



b_1 and b_2 are substantively meaningless. The moderator is scaled from 1 to 7. An estimate of group differences when W = 0 has no interpretation. Even if W = 0 were meaningful, the estimate of Y for the no protest group when W = 0 is off the top of the measurement scale of Y.

Interpretation of the coefficients

$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$



$\theta_{D_1 \to Y} = b_1 + b_4 W$

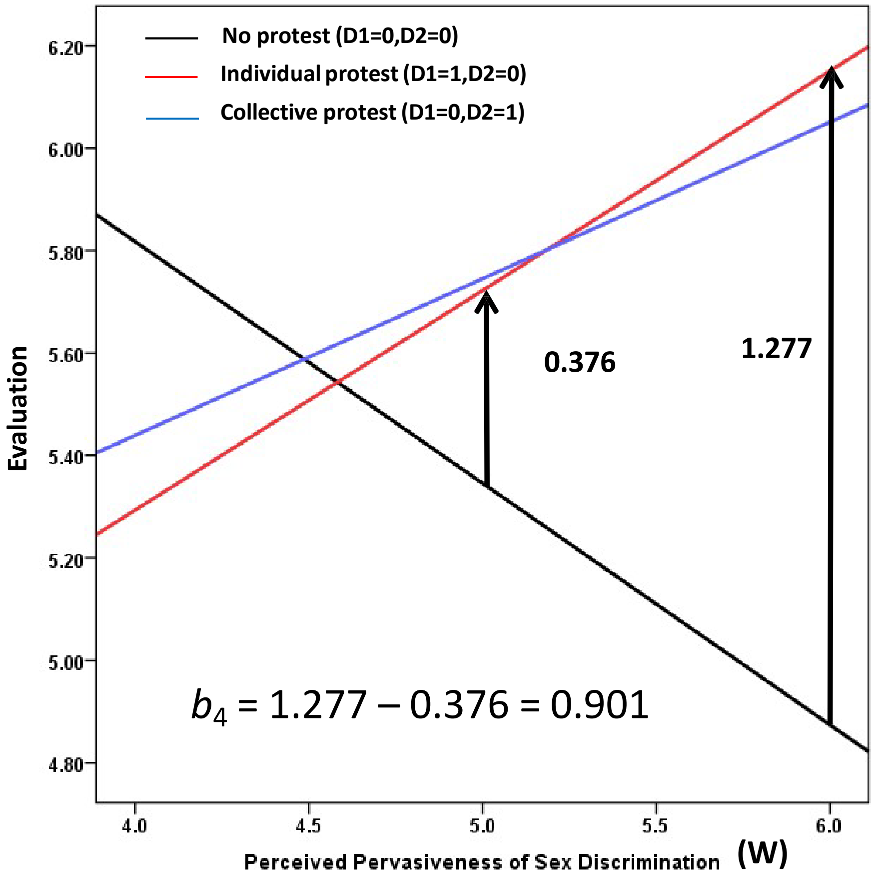
$\theta_{D_1 \to Y} = -4.129 + 0.901W$

SEXISM (W)	$\theta_{D_1 \to Y}$
4.00	-0.525
5.00	0.376
6.00	1.277

$$b_4 = (\theta_{D_1 \to Y} \mid W = \mathrm{A}+1) - (\theta_{D_1 \to Y} \mid W = \mathrm{A}) \text{ for all A}$$

Interpretation of the coefficients

$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$



$\theta_{D_1 \rightarrow Y} = b_1 + b_4W$

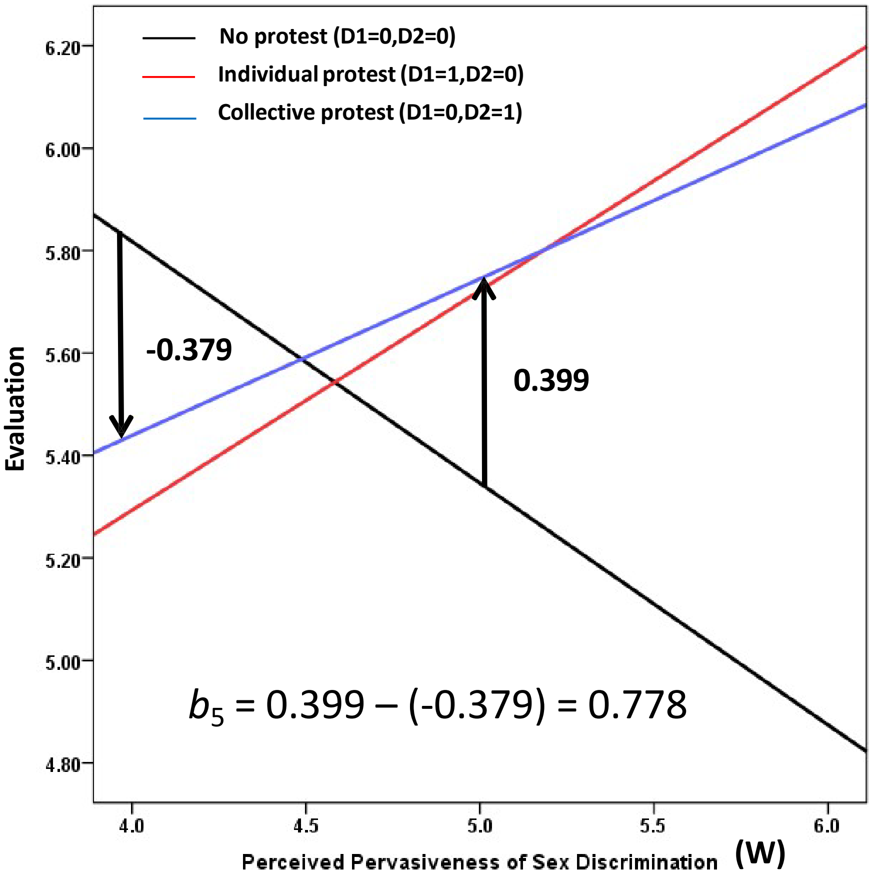
$\theta_{D_1 \rightarrow Y} = -4.129 + 0.901W$

SEXISM (W)	$\theta_{D_1 \rightarrow Y}$
4.00	-0.525
5.00	0.376
6.00	1.277

$$b_4 = (\theta_{D_1 \rightarrow Y} \mid W = \mathrm{A}+1) - (\theta_{D_1 \rightarrow Y} \mid W = \mathrm{A}) \text{ for all A}$$

Interpretation of the coefficients

$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$



$b_5 = 0.399 - (-0.379) = 0.778$

$\theta_{D_2 \rightarrow Y} = b_2 + b_5W$

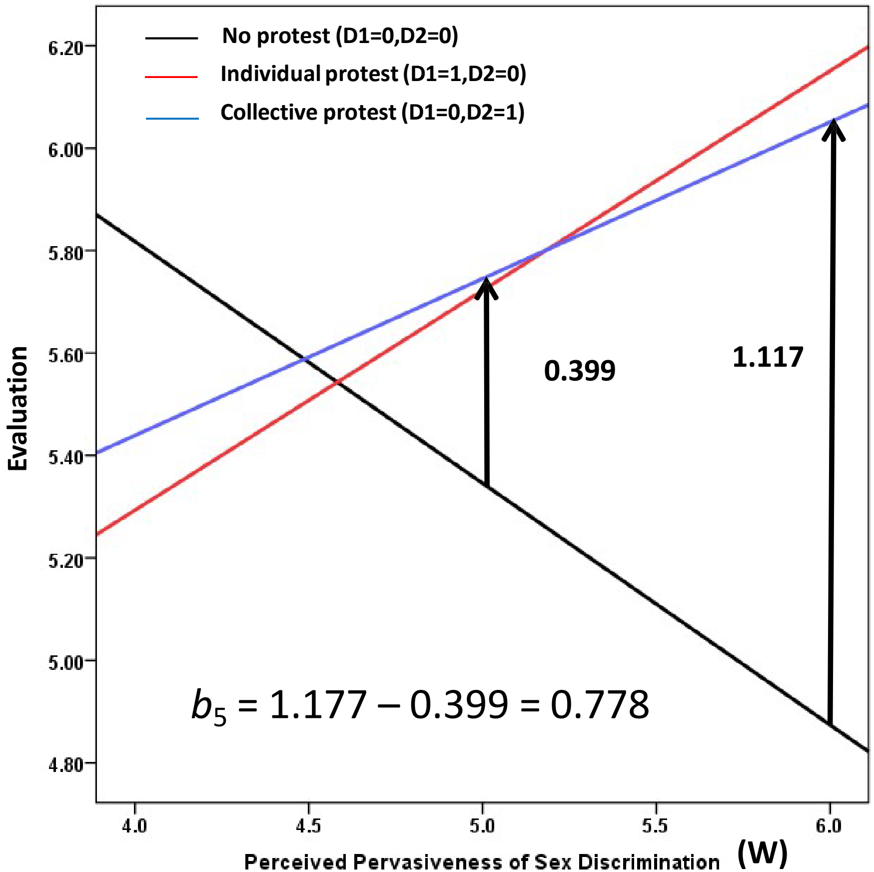
$\theta_{D_2 \rightarrow Y} = -3.491 + 0.778W$

SEXISM (W)	$\theta_{D_2 \rightarrow Y}$
4.00	-0.379
5.00	0.399
6.00	1.177

$$b_5 = (\theta_{D_2 \rightarrow Y} \mid W = \mathrm{A}+1) - (\theta_{D_2 \rightarrow Y} \mid W = \mathrm{A}) \text{ for all A}$$

Interpretation of the coefficients

$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$



$\theta_{D_2 \to Y} = b_2 + b_5W$

$\theta_{D_2 \to Y} = -3.491 + 0.778W$

SEXISM (W)	$\boldsymbol{\theta_{D_2 \to Y}}$
4.00	-1.157
5.00	0.399
6.00	1.177

$$b_5 = (\theta_{D_2 \to Y} \mid W = \mathrm{A}+1) - (\theta_{D_2 \to Y} \mid W = \mathrm{A}) \text{ for all A}$$

Statistical inference

Suppose X is a multicategorical focal predictor variable with 3 categories, represented with $k - 1 = 2$ variables D (e.g., indicator codes), and W is a quantitative (or dichotomous) moderator. Of interest is whether the effect of X is moderated by W.

Model 1: $\hat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 W_i$

Model 2: $\hat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 W_i + b_4 W_i D_{1i} + b_5 W_i D_{2i}$

Model 1 is nested under Model 2, so the increase in R^2 that results when the 2 products D_1W and D_2W are added to model 1 can be converted to an F-ratio and a p-value derived from the $F(2, df_{residual2})$ distribution, where $df_{residual2}$ is the residual degrees of freedom for model 2. Rejection of the null hypothesis that the change in R^2 equals zero implies that the effect of the multicategorical focal predictor on Y depends on W.

This test is akin to testing the composite null hypothesis that the regression coefficients for both D_1W and D_2W are both equal to zero.

A visual representation of nested models

Model 1: $\widehat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 W_i$

Model 2: $\widehat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 W_i + b_4 W_i D_{1i} + b_5 W_i D_{2i}$

Y = Evaluation

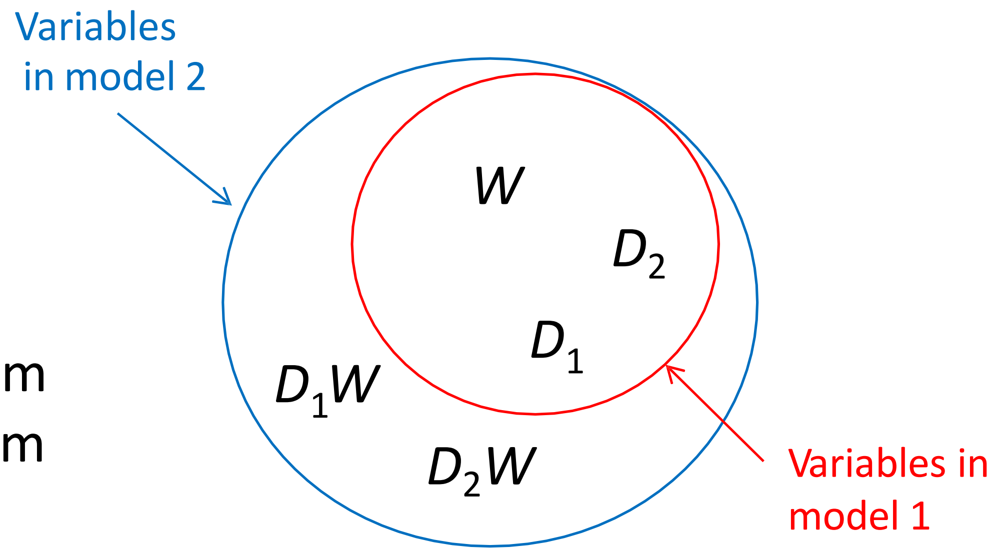
D_1 = Individual protest

D_2 = Collective protest

W = Sexism

D_1W = Individual × Sexism

D_2W = Collective × Sexism



Given model 1 is nested within 2, then a known mathematical function of their difference in R^2 = R^2(model2) – R^2(model 1) = ΔR^2 = is distributed as $F(2, df_{residual2})$ under the null hypothesis that the effect of her choice on how she is perceived does not depend on perceived pervasiveness of sex discrimination.

Implementation

Model 1: $\widehat{Y}_i = 4.76 + 0.497D_{1i} + 0.457D_{2i} + 0.108W_i$

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.230[a]	.053	.030	1.03386

a. Predictors: (Constant), SEXISM: perceived pervasiveness of sex discrimination, d2, d1

$R^2 = 0.053$

Model 2: $\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.367[a]	.135	.100	.99596

a. Predictors: (Constant), d2w, SEXISM: perceived pervasiveness of sex discrimination, d1, d2, d1w

$R^2 = 0.135$

$df_{residual2}$ = residual degrees of freedom from model 2.

$\Delta R^2 = 0.135 - 0.053 = 0.082$

$$F(2, df_{residual2}) = \frac{df_{residual2}\Delta R^2}{2(1 - R_2^2)}$$

$$F(2{,}123) = \frac{123\,(0.082)}{2(1 - 0.135)} = 5.847$$

Implementation

```
regression/statistics defaults change/dep = eval/method = enter d1 d2 sexism
                                          /method = enter d1w d2w.
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.230[a]	.053	.030	1.03386	.053	2.317	3	125	.079
2	.367[b]	.135	.100	.99596	.082	5.847	2	123	.004

a. Predictors: (Constant), SEXISM: perceived pervasiveness of sex discrimination, d2, d1

b. Predictors: (Constant), SEXISM: perceived pervasiveness of sex discrimination, d2, d1, d2w, d1w

The effect of her behavior depends on beliefs about the pervasiveness of sex discrimination in society.

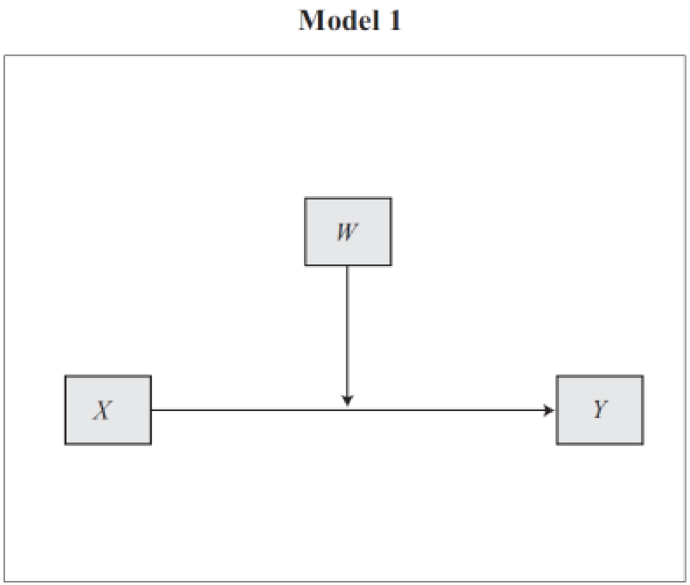
```
data lawyer2;set lawyer2;
d1=(protest=1);d2=(protest=2);
d1w=d1*sexism;d2w=d2*sexism;run;
proc reg data=lawyer2;
model eval = d1 d2 d1w d2w sexism;
test d1w=0,d2w=0;
run;
```

```
                The REG Procedure
                 Model: MODEL1

      Test 1 Results for Dependent Variable eval

                                    Mean
Source              DF            Square    F Value    Pr > F

Numerator            2           5.80021       5.85    0.0038
Denominator        123           0.99193
```

```
model1<-lm(eval~d1+d2+sexism,data=lawyer2.dummy)
model2<-lm(eval~d1*sexism+d2*sexism,data=lawyer2.dummy)
anova(model1,model2)
```

```
Model 1: eval ~ d1 + d2 + sexism
Model 2: eval ~ d1 + d2 + sexism + d1 * sexism + d2 * sexism
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1    125 133.61
2    123 122.01  2      11.6 5.8474 0.003751 **
```

Implementation in PROCESS



PROCESS has an option 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 X is a multicategorical variable and to use dummy coding to represent the groups. Other coding options are available. See an excerpt from the PROCESS documentation in your course book.

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

```
process y=eval/w=sexism/x=protest/model=1/mcx=1/plot=1.
```

```
%process (data=lawyer2,y=eval,w=sexism,x=protest,model=1,mcx=1,plot=1);
```

```
process(data=lawyer2,y="eval",w="sexism",x="protest",model=1,mcx=1,
  plot=1)
```

PROCESS Output

```
Model  : 1
    Y  : eval
    X  : protest
    W  : sexism

Sample
Size:  129

Coding of categorical X variable for analysis:
protest    X1     X2
  .000    .000    .000
 1.000   1.000    .000
 2.000    .000   1.000
```

X1 codes individual protest, X2 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 VARIABLE:
 eval
```

$$Y = 7.706 - 4.129D_1 - 3.491D_2 - 0.472W + 0.901D_1W + 0.778D_2W$$

Model Summary

R	R-sq	MSE	F	df1	df2	p
.3673	.1349	.9919	3.8372	5.0000	123.0000	.0029

Model

	coeff	se	t	p	LLCI	ULCI
constant	7.7062	1.0525	7.3218	.0000	5.6229	9.7896
X1	-4.1288	1.4985	-2.7553	.0068	-7.0949	-1.1626
X2	-3.4908	1.4078	-2.4796	.0145	-6.2775	-.7041
sexism	-.4725	.2053	-2.3017	.0230	-.8788	-.0662
Int_1	.9012	.2875	3.1346	.0022	.3321	1.4703
Int_2	.7778	.2752	2.8266	.0055	.2331	1.3225

b_1 = -4.129
b_2 = -3.491
b_3 = -0.473
b_4 = 0.901
b_5 = 0.778

Product terms generated for interaction

```
Product terms key:
 Int_1    :    X1    x    sexism
 Int_2    :    X2    x    sexism
```

Test(s) of highest order unconditional interaction(s):

	R2-chng	F	df1	df2	p
X*W	.0822	5.8474	2.0000	123.0000	.0038

Test of moderation

Generating a graph from PROCESS "PLOT" option output

```
DATA LIST FREE/
   protest     sexism      eval          .
  BEGIN DATA.
          .0000      4.2500       5.6981
         1.0000      4.2500       5.3996
         2.0000      4.2500       5.5131
          .0000      5.1200       5.2871
         1.0000      5.1200       5.7726
         2.0000      5.1200       5.7787
          .0000      5.8960       4.9204
         1.0000      5.8960       6.1053
         2.0000      5.8960       6.0156
END DATA.
GRAPH/SCATTERPLOT=
 sexism   WITH      eval      BY        protest  .
```

PROCESS for SPSS writes the code for you. Cut and paste as syntax into SPSS and run to generate an editable plot of the model.

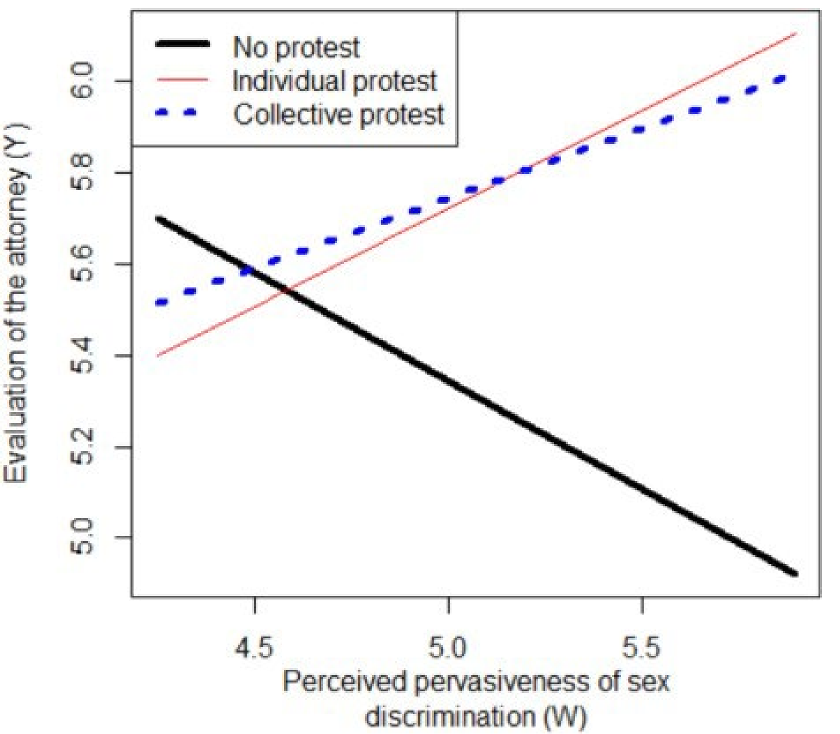
```
data;
input protest sexism esteval;
datalines;
         .0000   4.2500      5.6981
        1.0000   4.2500      5.3996
        2.0000   4.2500      5.5131
         .0000   5.1200      5.2871
        1.0000   5.1200      5.7726
        2.0000   5.1200      5.7787
         .0000   5.8960      4.9204
        1.0000   5.8960      6.1053
        2.0000   5.8960      6.0156
run;
proc sgplot;reg x=sexism y=esteval/group=protest;run;
```

PROCESS for SAS only gives you this. The rest of the code you have to enter yourself.

Constructing a visual representation of the model (R)

From PROCESS plot output

```
x<-c(0,1,2,0,1,2,0,1,2)
w<-c(4.25,4.25,4.25,5.12,5.12,5.12,5.896,5.896,5.896)
y<-c(5.698,5.400,5.513,5.287,5.773,5.779,4.920,6.105,6.016)
plot(y=y,x=w,pch=15,col="white",xlab="Perceived pervasiveness of sex
discrimination (W)",ylab="Evaluation of the attorney (Y)")
legend.txt<-c("No protest","Individual protest","Collective protest")
legend("topleft",legend=legend.txt,lty=c(1,1,3),lwd=c(4,1,4),
col=c("black","red","blue"))
lines(w[x==0],y[x==0],lwd=4,lty=1,col="black")
lines(w[x==1],y[x==1],lwd=1,lty=1,col="red")
lines(w[x==2],y[x==2],lwd=4,lty=3,col="blue")
```



Probing the interaction

Having established that differences between groups in how they perceived the attorney depends on how pervasive a person perceives sex discrimination to be in society, it would be common to probe this moderation by estimating the effect of her behavior at various values of perceived pervasiveness of sex discrimination.

We can use the pick-a-point approach, implemented through the regression-centering method.

Insight: In this model

$$\widehat{Y_i} = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 W_i + b_4 W_i D_{1i} + b_5 W_i D_{2i}$$

b_1 and b_2 estimate the effect of X (for a given group relative to the reference group) when $W = 0$. We can do a simultaneous test that both of these regression coefficients are equal to zero using hierarchical entry, but after first centering W around values of the moderator of interest.

Review: Pick-a-point through regression centering

$$\widehat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3X_iW_i$$

In the above model, b_1 estimates the conditional effect of X when W = 0. If we desire the conditional effect of X when W equals some value λ we can produce a new variable W' that is W centered around λ, such that W' = 0 when $W = \lambda$. Then substitute W' for W in the model above. That is, we will estimate

$$\widehat{Y}_i = b'_0 + b'_1X_i + b_2(W_i - \lambda) + b_3X_i(W_i - \lambda)$$

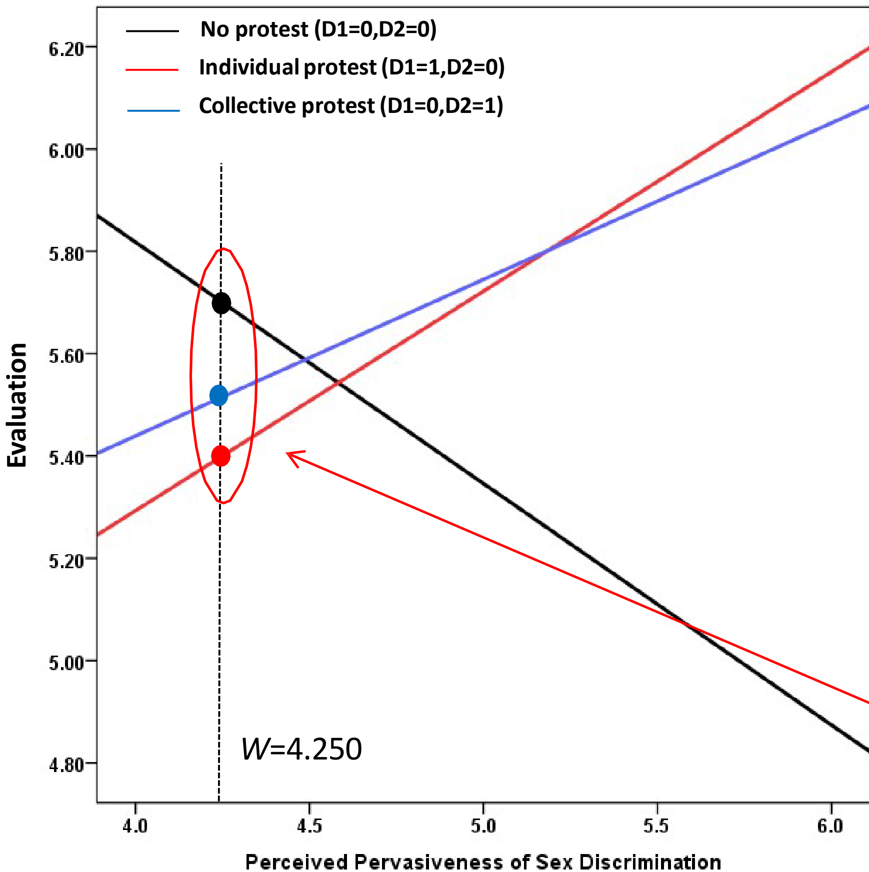
as

$$\widehat{Y}_i = b'_0 + b'_1X_i + b_2W'_i + b_3X_iW'_i \text{ where } W'_i = W_i - \lambda$$

In this model, b_1 is the conditional effect of X when W' = 0. But W' = 0 when $W = \lambda$. So b_1 estimates the conditional effect of X when $W = \lambda$. Your regression program will give you a standard error, **t-** and *p*-values, and a confidence interval.

Example

Did the attorney's behavior affect how she was perceived among those who see sexual discrimination as relatively less prevalent? We'll define such people as those at the **16th percentile** of the distribution as defined by the distribution in the sample.




```
frequencies variables = sexism
/statistics mean stddev
/percentiles = 16 50 84.
```

```
proc univariate data=lawyer2;var sexism;
output out=outpdata pctlpts=16 50 84
pctlpre=p;
proc print data=outpdata;run;
```

```
quantile(lawyer2$sexism, c(.16,.50,.84))
```

Statistics

SEXISM: perceived pervasiveness of s

N	Valid	129
	Missing	0
Mean		5.1170
Std. Deviation		.78376
Percentiles	16	4.2500
	50	5.1200
	84	5.8960

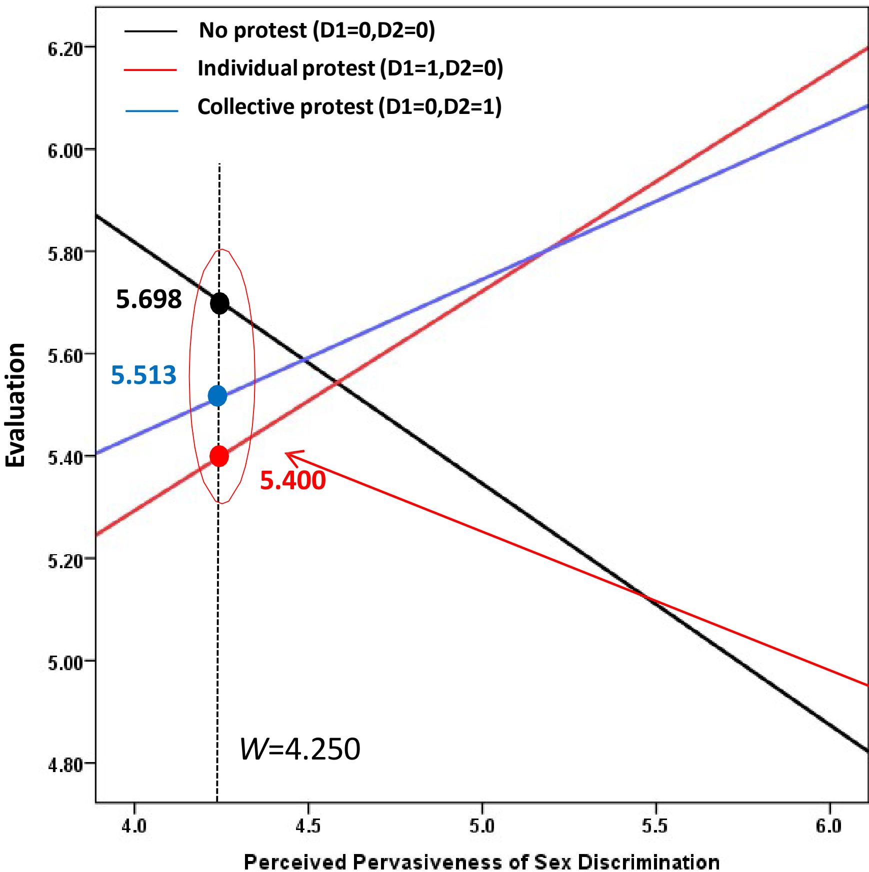
The 16th percentile is 4.250

We are interested in testing the equality of the estimates of evaluation among those "relatively low" on the moderator.

Calculation of the estimated *Y*

We don't need anything fancy to generate the estimated evaluation for three groups conditioned on being "relatively low" in their perceptions of sex discrimination prevalence. Just use the model.

$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$



D_1	D_2	W	$\hat{Y}$
0	0	4.250	5.698
1	0	4.250	5.400
0	1	4.250	5.513

We seek a formal test of the difference between these estimated values.

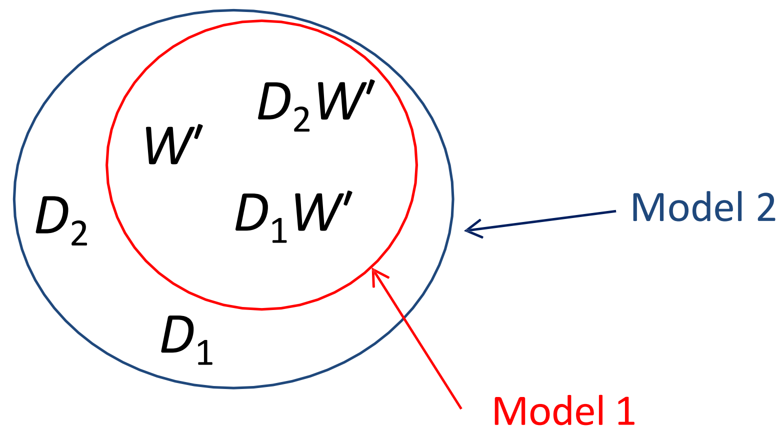
Example

Define $W' = W - \lambda$. In this model:

$$Y = b_0 + b_1 D_1 + b_2 D_2 + b_3 W' + b_4 D_1 W' + b_5 D_2 W'$$

b_1 estimates the difference in Y between the group coded with D_1 and the reference group, conditioned on $W = \lambda$.

b_2 estimates the difference in Y between the group coded with D_2 and the reference group, conditioned on $W = \lambda$.



Model 1 is nested within Model 2

Given model 1 is nested within 2, then a certain function of their difference in R^2 is distributed as $F(2, df_{residual2})$ under the null hypothesis that the regression coefficients for both D_1 and D_2 = 0.

Implementation

```
compute sexism_c = sexism-4.25.
compute d1w_c = d1*sexism_c.
compute d2w_c = d2*sexism_c.
regression/dep=eval/method=enter sexism_c d1w_c d2w_c.
```

```
data lawyer2;set lawyer2;
d1=(protest=1);d2=(protest=2);sexism_c=sexism-4.25;d1w_c=d1*sexism_c;d2w_c=d2*sexism_c;run;
proc reg data=lawyer2;model eval = d1w_c d2w_c sexism_c;run;
```

```
sexismp<-lawyer2.dummy$sexism-4.25
sexpd1<-sexismp*lawyer2.dummy$d1;sexpd2<-sexismp*lawyer2.dummy$d2
lawyer2.dummy<-data.frame(lawyer2.dummy,sexismp,sexpd1,sexpd2)
summary(lm(eval~sexismp+sexpd1+sexpd2,data=lawyer2.dummy))
```

```
Residual standard error: 0.9912 on 125 degrees of freedom
Multiple R-squared:  0.1293,    Adjusted R-squared:  0.1084
F-statistic: 6.187 on 3 and 125 DF,  p-value: 0.0005923
```

$R^2 = 0.129$

```
regression/dep=eval/method=enter d1 d2 sexism_c d1w_c d2w_c.
```

```
proc reg data=lawyer2;model eval = d1 d2 d1w_c d2w_c sexism_c;run;
```

```
summary(lm(eval~sexismp+d1+d2+sexpd1+sexpd2,data=lawyer2.dummy))
```

```
Residual standard error: 0.996 on 123 degrees of freedom
Multiple R-squared:  0.1349,    Adjusted R-squared:  0.09977
F-statistic: 3.837 on 5 and 123 DF,  p-value: 0.002893
```

$$F(2, df_{residual2}) = \frac{df_{residual2}\Delta R^2}{2(1-R_2^2)}$$

$R^2 = 0.135$

$\Delta R^2 = 0.006$

$df_{residual2} = 123$

$$F(2,123) = \frac{123(0.006)}{2(1-0.135)} = 0.402$$

Implementation

```
regression/statistics defaults change/dep = eval/method = enter sexism_c d1w_c d2w_c
                                              /method = enter d1 d2.
```

```
proc reg data=lawyer2;model eval = d1 d2 d1w_c d2w_c sexism_c;test d1=0,d2=0;run;
```

```
model1<-lm(eval~sexismp+sexpd1+sexpd2,data=lawyer2.dummy)
model2<-lm(eval~sexismp+d1+d2+sexpd1+sexpd2,data=lawyer2.dummy)
anova(model1,model2)
```

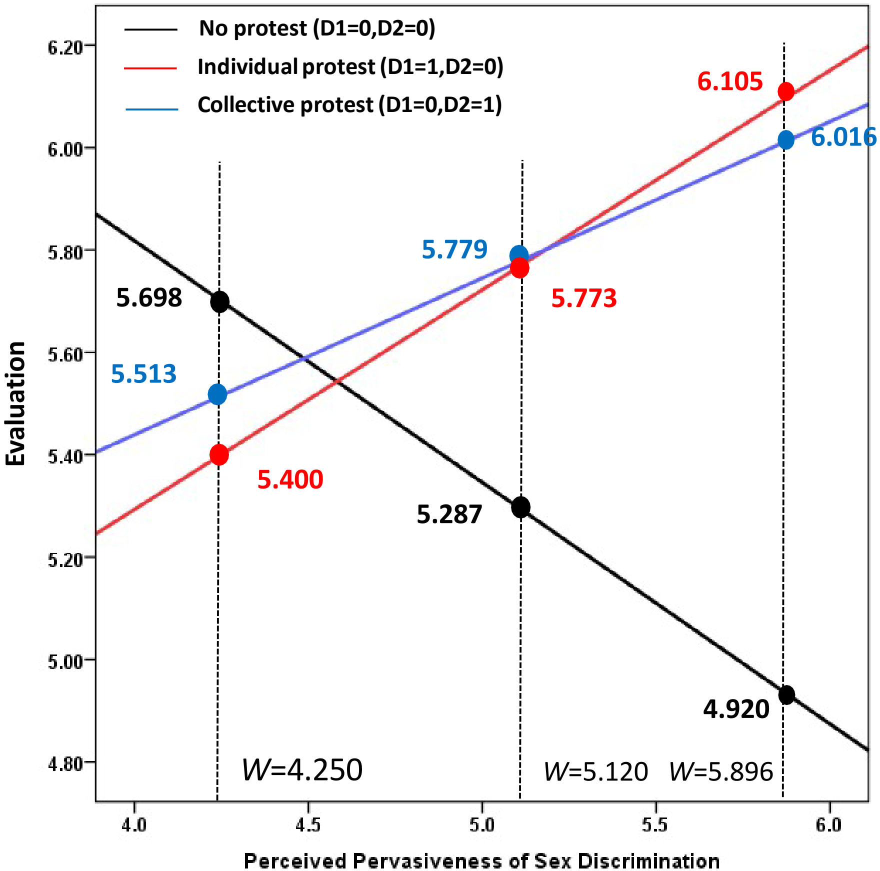
```
Analysis of Variance Table

Model 1: eval ~ sexismp + sexpd1 + sexpd2
Model 2: eval ~ sexismp + d1 + d2 + sexpd1 + sexpd2
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    125 122.81
2    123 122.01  2   0.79765 0.4021 0.6698
```

No difference between conditions in evaluation among those relatively low in perceptions of the prevalence of sex discrimination in society. $F(2,123) = 0.402$, $p = 0.670$.

Repeat for other values of the moderator

Repeat this procedure for any value of the moderator you choose. Here I use the 50th percentile (W = 5.120) for "moderate" and the 84th percentile (W = 5.896) for "high"

$$\hat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$



	W		
Condition	Low	Mod.	High
NP	5.698	5.287	4.920
IP	5.400	5.773	6.105
CP	5.513	5.779	6.016

Low: $F(2,123) = 0.402$, $p = .670$
Mod: $F(2,123) = 3.338$, $p = .039$
High: $F(2,123) = 8.818$, $p < .001$

PROCESS does this for you

```
Focal predict: protest  (X)
      Mod var: sexism   (W)

Conditional effects of the focal predictor at values of the moderator(s):

Moderator value(s):
sexism       4.2500

         Effect        se         t         p      LLCI      ULCI
X1       -.2985     .3402    -.8775     .3819    -.9720     .3749
X2       -.1851     .3089    -.5991     .5502    -.7966     .4264

Test of equality of conditional means
        F       df1        df2         p
    .4021    2.0000   123.0000     .6698

Estimated conditional means being compared:
    protest      eval
      .0000    5.6981
     1.0000    5.3996
     2.0000    5.5131
```

16th percentile of *W*

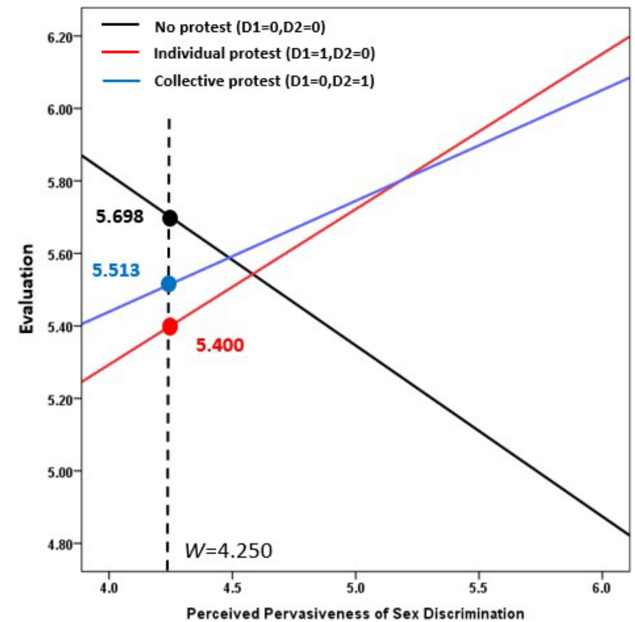
$\theta_{D_1 \to Y} = b_1 + b_4 W = -4.129 + 0.901(4.250)$

$\theta_{D_2 \to Y} = b_2 + b_5 W = -3.491 + 0.778(4.250)$

Difference in evaluation between IP and NP condition ($\theta_{D_1 \to Y}$) and between CP and NP condition ($\theta_{D_2 \to Y}$) among people "relatively low" in perceived pervasiveness of sex discrimination.

Test of null that both of these differences are equal to zero.

$$Y = 7.706 - 4.129D_1 - 3.491D_2 - 0.472W + 0.901D_1W + 0.778D_2W$$
$$= 7.706 - 4.129D_1 - 3.491D_2 - 0.472(4.25) + 0.901D_1(4.25) + 0.778D_2(4.25)$$



PROCESS probes the moderation using the pick-a-point approach setting the moderator to the 16th, 50th, and 84th percentiles.

PROCESS Output

50th percentile of *W*

```
Moderator value(s):
sexism     5.1200
```

$\theta_{D1\rightarrow Y} = b_1 + b_4W = -4.129 + 0.901(5.120)$

$\theta_{D2\rightarrow Y} = b_2 + b_5W = -3.491 + 0.778(5.120)$

Difference in evaluation between IP and NP condition ($\theta_{D1\rightarrow Y}$) and between CP and NP condition ($\theta_{D2\rightarrow Y}$) among people "relatively moderate" in perceived pervasiveness of sex discrimination.

	Effect	se	t	p	LLCI	ULCI
X1	.4855	.2191	2.2163	.0285	.0519	.9192
X2	.4916	.2158	2.2782	.0244	.0645	.9188

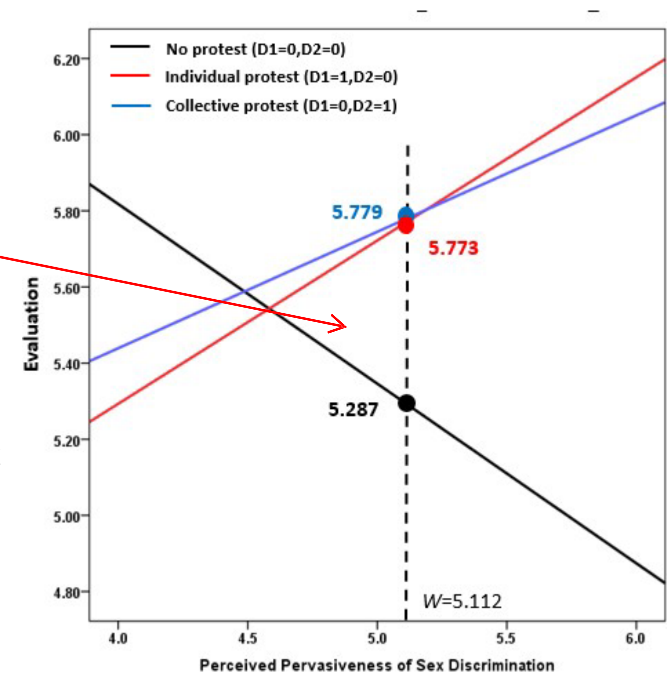
Test of equality of conditional means

F	df1	df2	p
3.3377	2.0000	123.0000	.0388

Test of null that both of these differences are equal to zero.

Estimated conditional means being compared:

protest	eval
.0000	5.2871
1.0000	5.7726
2.0000	5.7787



$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$

$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47(5.12) + 0.90(5.12)D_{1i} + 0.78(5.12)D_{2i}$$

PROCESS Output

84th percentile of W

```
Moderator value(s):
sexism      5.8960

          Effect        se          t          p       LLCI       ULCI
X1        1.1849     .3052     3.8825      .0002      .5808     1.7890
X2        1.0952     .3158     3.4684      .0007      .4702     1.7203

Test of equality of conditional means
         F         df1         df2           p
    8.8180      2.0000    123.0000       .0003

Estimated conditional means being compared:
   protest        eval
     .0000      4.9204
    1.0000      6.1053
    2.0000      6.0156
```

$\theta_{D_1 \rightarrow Y} = b_1 + b_4 W = -4.129 + 0.901(5.896)$

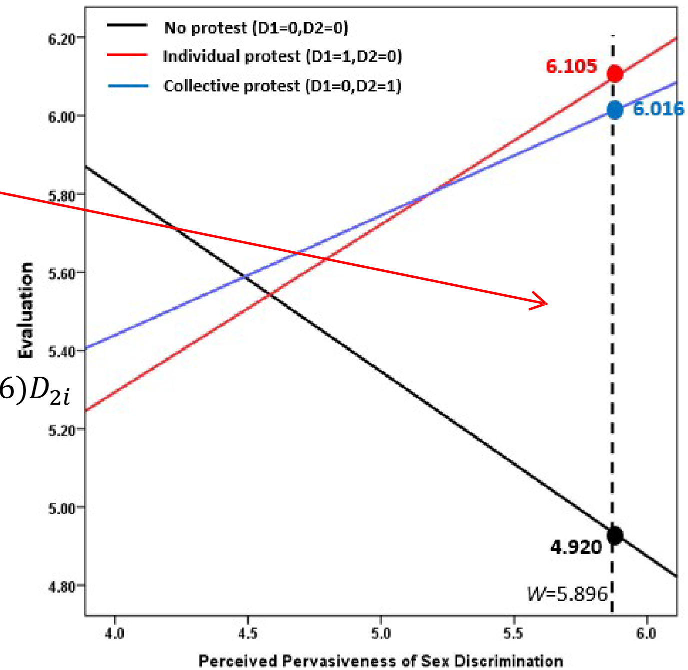
$\theta_{D_2 \rightarrow Y} = b_2 + b_5 W = -3.491 + 0.778(5.896)$

Difference in evaluation between IP and NP condition ($\theta_{D_1 \rightarrow Y}$) and between CP and NP condition ($\theta_{D_2 \rightarrow Y}$) among people "relatively high" in perceived pervasiveness of sex discrimination.

Test of null that both of these differences are equal to zero.

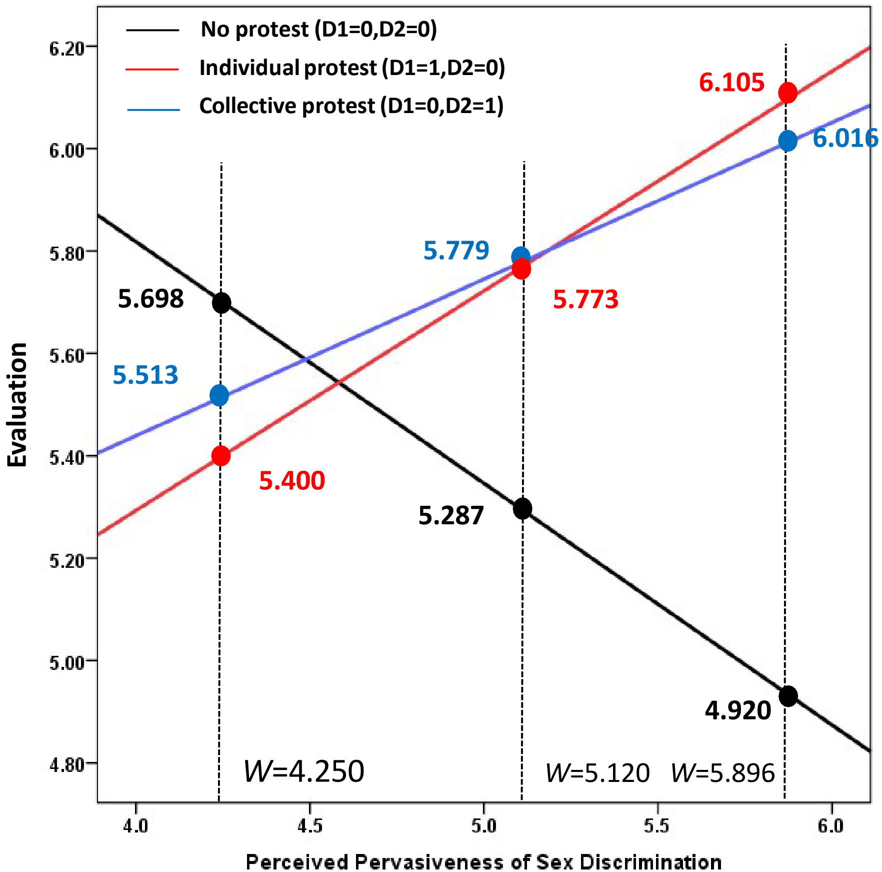
$$\hat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47W_i + 0.90W_iD_{1i} + 0.78W_iD_{2i}$$

$$\hat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47(5.896) + 0.90(5.896)D_{1i} + 0.78(5.896)D_{2i}$$



Use the **moments** or **wmodval** options if you would prefer to use other values of W.

The Johnson-Neyman technique



	W		
	4.250	5.120	5.896
Condition	Low	Mod.	High
NP	5.698	5.287	4.920
IP	5.400	5.773	6.105
CP	5.513	5.779	6.016

Low: $F(2,123) = 0.402$, $p = .670$
Mod: $F(2,123) = 3.338$, $p = .039$
High: $F(2,123) = 8.818$, $p < .001$

The Johnson-Neyman technique

Hayes, A. F., & Montoya, A. K. (2017). A tutorial for testing, visualizing, and probing an interaction involving a multicategorical variable in linear regression analysis. *Communication Methods and Measures*, *11*, 1-30.

- When *X* is three groups, analytically finding the regions of significance is a tedious mathematical task, but it is possible. With four or more groups, it is not mathematically possible to solve this problem analytically.

- Montoya (2016, Master's thesis) wrote the **OGRS** macro for SPSS and SAS that finds region(s) of significance iteratively through a systematic hunt across the moderator space for the points of transition between a significant and nonsignificant group difference. Download OGRS from *www.akmontoya.com.*

In OGRS, the moderator is *M*

```
ogrs vars=eval protest sexism/y=eval/x=protest/m=sexism.
```

```
%ogrs (data=lawyer2,vars=eval protest sexism,y=eval,x=protest,m=sexism)
```

OGRS is not available for R

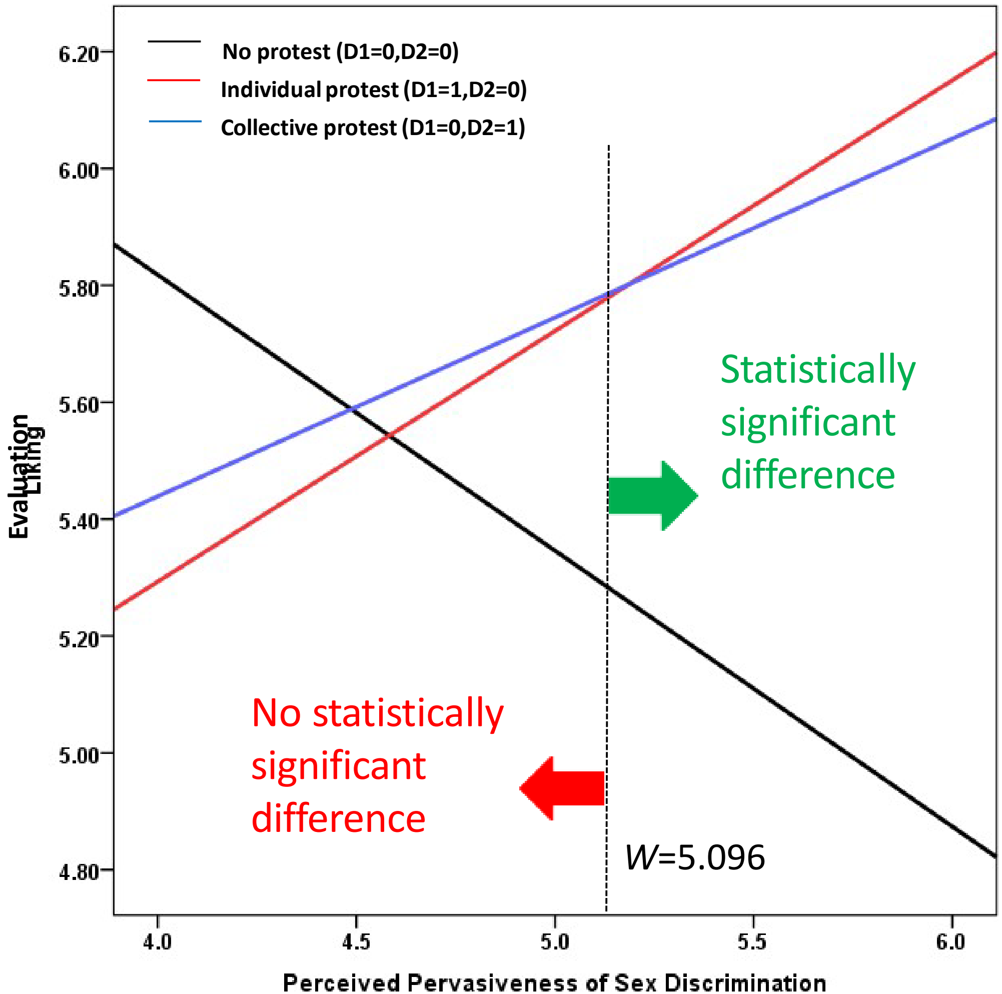
OGRS output

```
Variables:
X = protest
M = sexism
Y = eval

Dummy Variable Coding Scheme:
protest   D1   D2
      0    1    0
      1    0    1
      2    0    0

Sample size:
  129

*****************************************************************************
Outcome: eval

Model Summary
          R       R-sq         F        df1        df2          p
     .3673        .1349     3.8372     5.0000   123.0000      .0029

Model
              coeff         SE          t          p        LLCI        ULCI
constant     4.2154      .9350     4.5087      .0000      2.3647      6.0662
D1           3.4908     1.4078     2.4796      .0145       .7041      6.2775
D2           -.6380     1.4184     -.4498      .6537     -3.4456      2.1697
sexism        .3053      .1833     1.6660      .0983      -.0574       .6681
Int1         -.7778      .2752    -2.8266      .0055     -1.3225      -.2331
Int2          .1234      .2722      .4533      .6511      -.4155       .6623

Interactions:
 Int1     =        D1      X      sexism
 Int2     =        D2      X      sexism

R-square increase due to interaction(s):
     R2-chng          F             df1   df2         p
       .0822     5.8474       2.0000  123.0000     .0038
```

OGRS output

```
************************* JOHNSON-NEYMAN TECHNIQUE **************************

Moderator value(s) defining Johnson-Neyman boundaries of significance:
    5.0957

Conditional effect of X on Y at values of the moderator (M)
      sexism     R2-chng          F          p
      2.8700       .0413     2.9379      .0567
      3.0765       .0377     2.6809      .0725
      3.2830       .0335     2.3831      .0965
      3.4895       .0287     2.0376      .1347
      3.6960       .0231     1.6396      .1983
      3.9025       .0168     1.1910      .3074
      4.1090       .0100      .7132      .4921
      4.3155       .0039      .2748      .7602
      4.5220       .0006      .0428      .9581
      4.7285       .0047      .3331      .7174
      4.9350       .0213     1.5132      .2243
      5.0957       .0432     3.0699      .0500
      5.1415       .0503     3.5781      .0309
      5.3480       .0823     5.8487      .0037
      5.5545       .1063     7.5603      .0008
      5.7610       .1197     8.5132      .0003
      5.9675       .1253     8.9082      .0002
      6.1740       .1264     8.9884      .0002
      6.3805       .1254     8.9145      .0002
      6.5870       .1234     8.7740      .0003
      6.7935       .1211     8.6106      .0003
      7.0000       .1188     8.4451      .0004
```

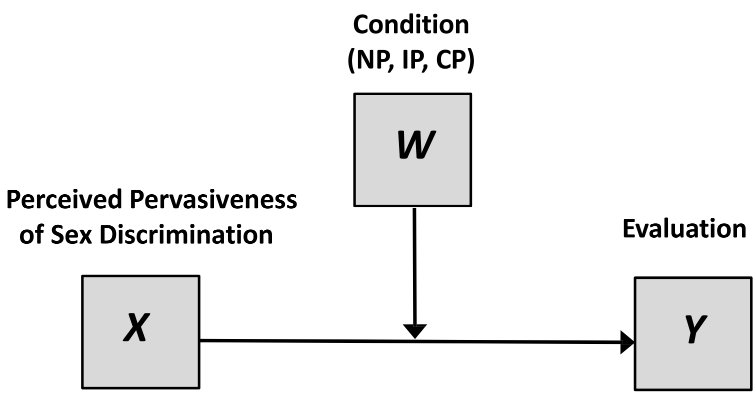
Visual representation



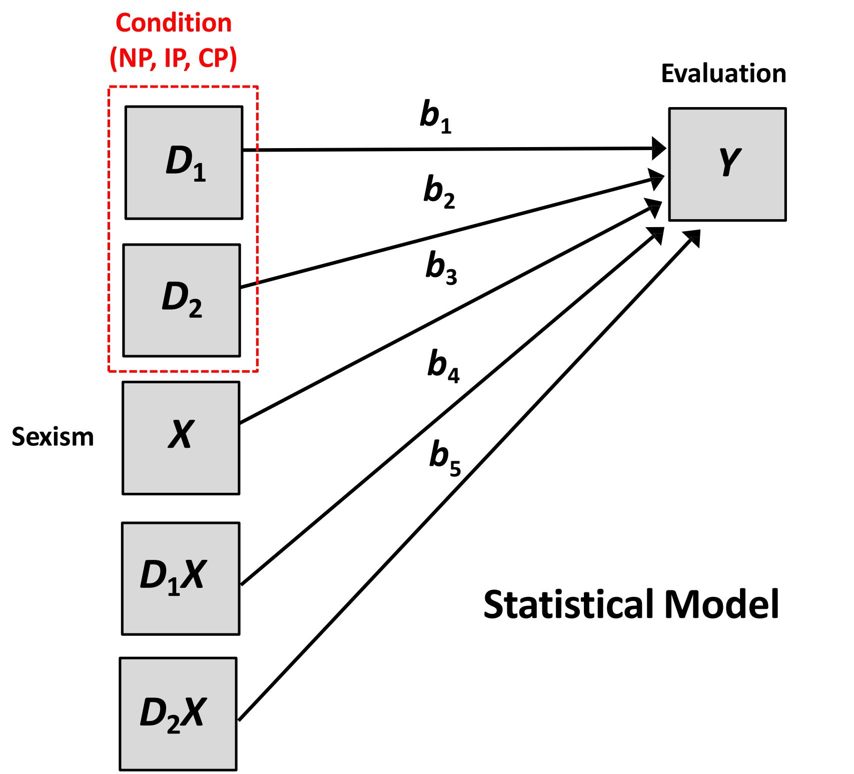
A multicategorical variable as moderator

The same model can be used to estimate the moderation of the effect of a continuous variable by a multicategorical variable. Here, we reverse the roles of perceived pervasiveness of sex discrimination and experimental condition and ask how the effect of such perceptions on how the attorney was evaluated depends on her choice of action. We'll dummy code as before with the no protest condition as the reference group.

$$\hat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 X_i + b_4 X_i D_{1i} + b_5 X_i D_{2i}$$



Conceptual Model



Statistical Model

A conditional effects model (*X* moderated by *W*)

```
compute d1 = (protest = 1).
compute d2 = (protest = 2).
compute d1x = d1*sexism.
compute d2x = d2*sexism.
regression/dep = eval/method = enter d1 d2 sexism d1x d2x.
```

```
data lawyer2;set lawyer2;d1=(protest=1);d2=(protest=2);d1x=d1*sexism;d2x=d2*sexism;run;
proc reg data=lawyer2;model eval = d1 d2 sexism d1x d2x;run;
```

```
d1<-as.numeric(lawyer2$protest==1)
d2<-as.numeric(lawyer2$protest==2)
lawyer2.dummy<-data.frame(lawyer2,d1,d2)
summary(lm(eval~d1+d2+sexism+d1*sexism+d2*sexism,data=lawyer2.dummy))
```

$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47X_i + 0.90X_iD_{1i} + 0.78X_iD_{2i}$$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.7062     1.0525   7.322 2.79e-11 ***
d1           -4.1288     1.4985  -2.755  0.00676 **
d2           -3.4908     1.4078  -2.480  0.01451 *
sexism       -0.4725     0.2053  -2.302  0.02303 *
d1:sexism     0.9012     0.2875   3.135  0.00215 **
d2:sexism     0.7778     0.2752   2.827  0.00549 **
```

Of course, this is the same model as when perceived pervasiveness of sex discrimination was the moderator and condition was the focal predictor. The coefficients are identical, and the test of the interaction is the same

X's effect as a function of W (3 category moderator)

$$\hat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 X_i + b_4 X_i D_{1i} + b_5 X_i D_{2i}$$

can be rewritten as

$$\hat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + (b_3 + b_4 D_{1i} + b_5 D_{2i}) X_i$$

So the effect of a one unit difference in X on Y depends on W, with W represented with indicator codes D_1 and D_2. Defining the conditional effect of X as

$$\theta_{X \to Y} = (b_3 + b_4 D_1 + b_5 D_2)$$

The model can be represented in a simple form:

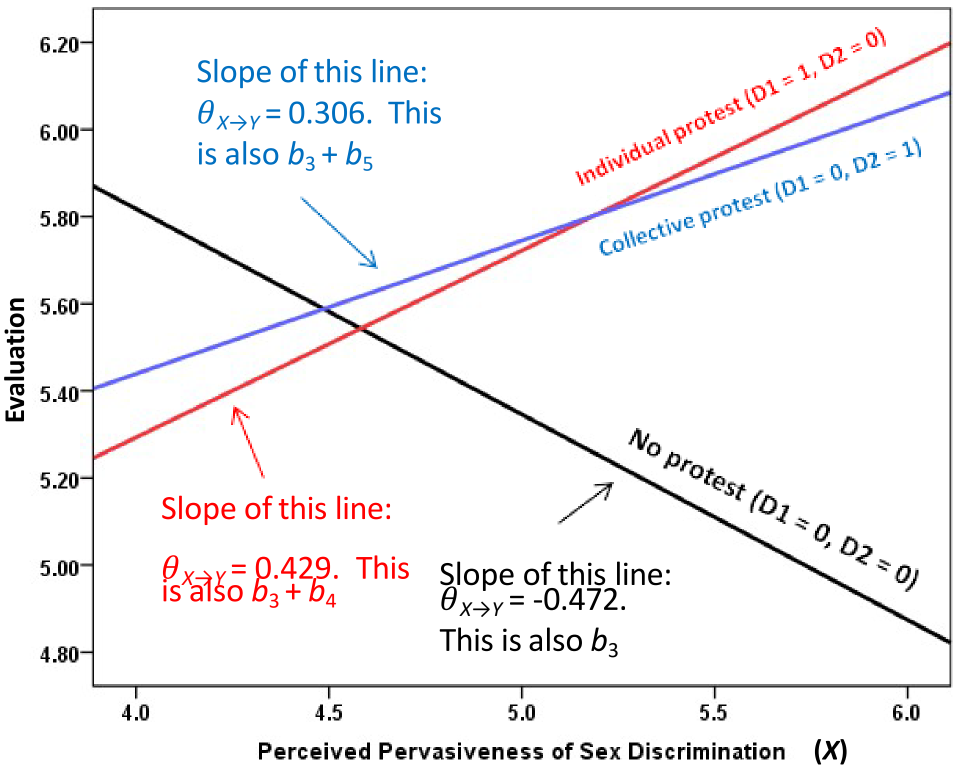
$$\hat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + \theta_{X \to Y} X_i$$

A graphical depiction of the model

$$\hat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47X_i + 0.90X_iD_{1i} + 0.78X_iD_{2i}$$

or, equivalently,

$$\hat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} + (-0.47 + 0.90D_{1i} + 0.78D_{2i})X_i$$



The conditional effect of perceived pervasiveness of sex discrimination ($\theta_{X\rightarrow Y}$) is defined by the function

$$\theta_{X\rightarrow Y} = b_3 + b_4D_1 + b_5D_2$$
$$= -0.472 + 0.901D_1 + 0.778D_2$$

D_1	D_2	$\theta_{X\rightarrow Y}$
0	0	-0.472
1	0	0.429
0	1	0.306

Interpretation of the coefficients

$$\widehat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + b_3 X_i + b_4 X_i D_{1i} + b_5 X_i D_{2i}$$

can be rewritten as

$$\widehat{Y}_i = b_0 + b_1 D_{1i} + b_2 D_{2i} + (b_3 + b_4 D_{1i} + b_5 D_{2i}) X_i$$

b_1 estimates the estimated difference in Y between the group coded $D_1 = 1$ and the reference group, conditioned on X being zero.

b_2 estimates the estimated difference in Y between the group coded $D_2 = 1$ and the reference group, conditioned on X being zero.

b_3 estimates the effect of X in the reference group (i.e., D_1 and $D_2 = 0$).

b_4 estimates the difference in the effect of X on Y between the group coded with D_1 and the reference group.

b_5 estimates the difference in the effect of X on Y between the group coded with D_2 and the reference group.

Comparisons between conditional effects of *X*

Returning to the original model, with the no protest condition as the reference group, D_1 coding individual protest, and D_2 coding collective protest:

$$\hat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47X_i + 0.90X_iD_{1i} + 0.78X_iD_{2i}$$

The conditional effect of perceived pervasiveness of sex discrimination ($\theta_{X\rightarrow Y}$) is defined by the function

$$\theta_{X\rightarrow Y} = b_3 + b_4D_1 + b_5D_2$$
$$= -0.472 + 0.901D_1 + 0.778D_2$$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.7062     1.0525   7.322 2.79e-11 ***
d1           -4.1288     1.4985  -2.755  0.00676 **
d2           -3.4908     1.4078  -2.480  0.01451 *
sexism       -0.4725     0.2053  -2.302  0.02303 *
d1:sexism     0.9012     0.2875   3.135  0.00215 **
d2:sexism     0.7778     0.2752   2.827  0.00549 **
```

D_1	D_2	$\theta_{X\rightarrow Y}$
0	0	-0.472
1	0	0.429
0	1	0.306

$b_4 = 0.901$
$= \theta_{X\rightarrow Y}$(individual protest) – $\theta_{X\rightarrow Y}$(no protest)
$= 0.429 - (-0.472)$
$= 0.901$

So the effect of perceived pervasiveness of sex discrimination on evaluation differs between those told she individually protested and those told she did not protest, $b_4 = 0.901$, $t(123) = 3.135$, $p = .002$

Comparisons between conditional effects of *X*

Returning to the original model, with the no protest condition as the reference group, D_1 coding individual protest, and D_2 coding collective protest:

$$\widehat{Y}_i = 7.71 - 4.13D_{1i} - 3.49D_{2i} - 0.47X_i + 0.90X_iD_{1i} + 0.78X_iD_{2i}$$

The conditional effect of perceived pervasiveness of sex discrimination ($\theta_{X \to Y}$) is defined by the function

$$\theta_{X \to Y} = b_3 + b_4D_1 + b_5D_2$$
$$= -0.472 + 0.901D_1 + 0.778D_2$$

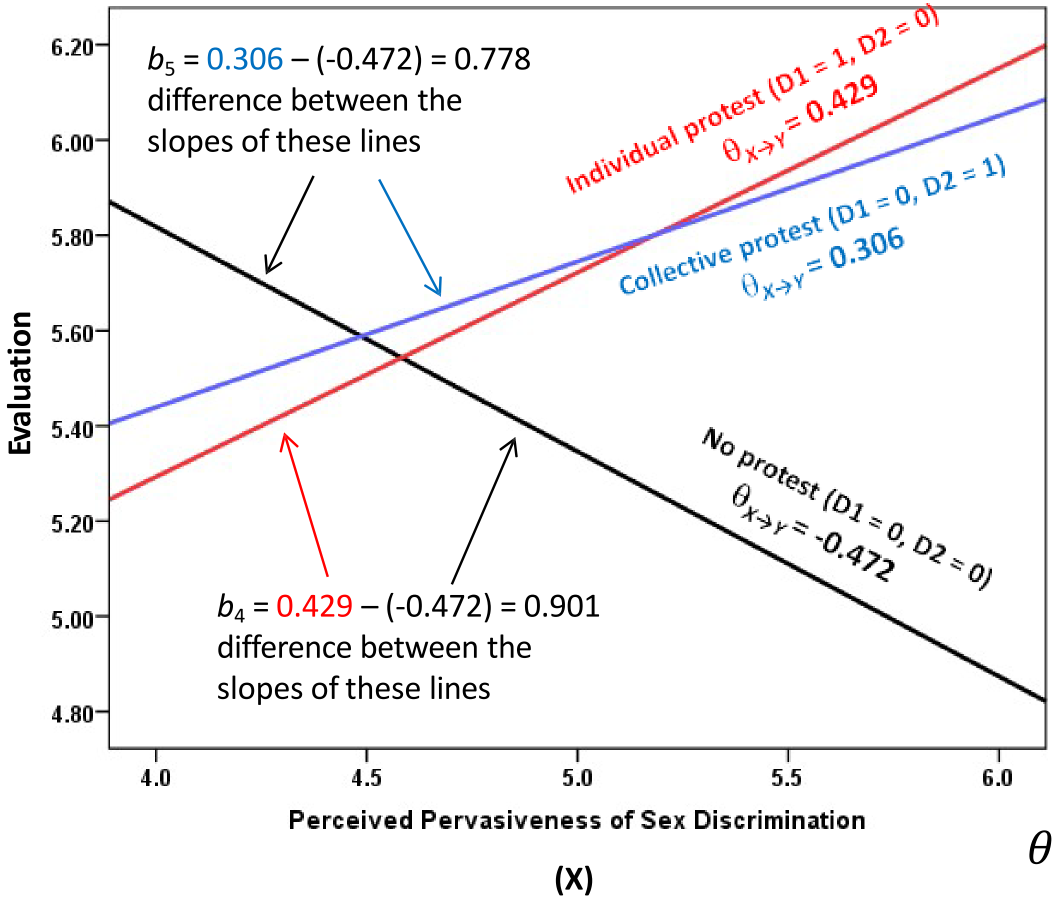
D_1	D_2	$\theta_{X \to Y}$
0	0	-0.472
1	0	0.429
0	1	0.306

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.7062     1.0525   7.322 2.79e-11 ***
d1            -4.1288     1.4985  -2.755  0.00676 **
d2            -3.4908     1.4078  -2.480  0.01451 *
sexism        -0.4725     0.2053  -2.302  0.02303 *
d1:sexism      0.9012     0.2875   3.135  0.00215 **
d2:sexism      0.7778     0.2752   2.827  0.00549 **
```

$b_5 = 0.778$
$= \theta_{X \to Y}(\text{collective protest}) - \theta_{X \to Y}(\text{no protest})$
$= 0.306 - (-0.472)$
$= 0.778$

So the effect of perceived pervasiveness of sex discrimination on evaluation differs between those told she collectively protested and those told she did not protest, $b_5 = 0.778$, $t(123) = 2.827$, $p = .005$

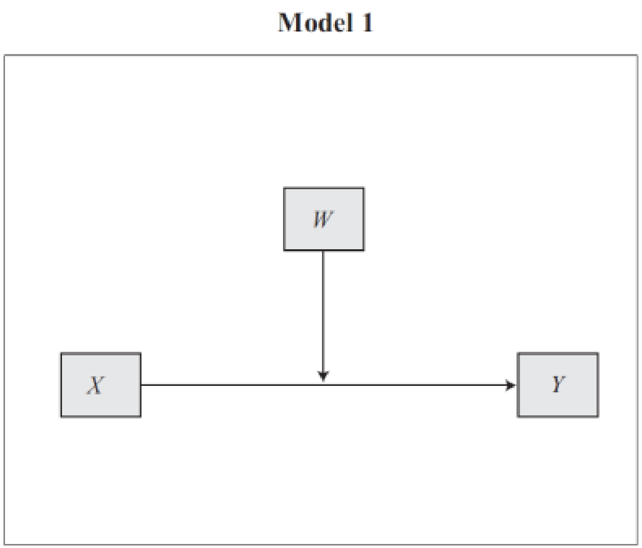
A visual representation



D_1	D_2	$\theta_{X\to Y}$
0	0	-0.472
1	0	0.429
0	1	0.306

$$\theta_{X\to Y} = b_3 + b_4 D_1 + b_5 D_2$$
$$= -0.472 + 0.901 D_1 + 0.778 D_2$$

Implementation in PROCESS



PROCESS has an option for specifying *W* as a multicategorical variable with up to 9 categories.

MCW=1 tells PROCESS that the moderator *W* is a multicategorical variable and to use dummy coding to represent the groups. Other coding options are available.

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

```
process y=eval/x=sexism/w=protest/model=1/mcw=1/plot=1.
```

```
%process (data=lawyer2,y=eval,x=sexism,w=protest,model=1,mcw=1,plot=1);
```

```
process(data=lawyer2,y="eval",x="sexism",w="protest",model=1,mcw=1,plot=1)
```

PROCESS Output

```
Model  : 1
    Y   :  eval
    X  : sexism
    W  : protest

Sample
Size:  129

Coding of categorical W variable for analysis:
 protest     W1      W2
   .000    .000    .000
  1.000   1.000    .000
  2.000    .000   1.000
```

W1 codes individual protest, W2 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 VARIABLE:
 eval
```

$$Y_i = 7.706 - 4.129D_{1i} - 3.491D_{2i} - 0.472X_i + 0.901D_{1i}X_i + 0.778D_{2i}X_i$$

Model Summary

R	R-sq	MSE	F	df1	df2	p
.3673	.1349	.9919	3.8372	5.0000	123.0000	.0029

Model

	coeff	se	t	p	LLCI	ULCI
constant	7.7062	1.0525	7.3218	.0000	5.6229	9.7896
sexism	-.4725	.2053	-2.3017	.0230	-.8788	-.0662
W1	-4.1288	1.4985	-2.7553	.0068	-7.0949	-1.1626
W2	-3.4908	1.4078	-2.4796	.0145	-6.2775	-.7041
Int_1	.9012	.2875	3.1346	.0022	.3321	1.4703
Int_2	.7778	.2752	2.8266	.0055	.2331	1.3225

Product terms generated for interaction

```
Product terms key:
 Int_1:    sexism   x        W1
 Int_2:    sexism   x        W2
```

Test(s) of highest order unconditional interaction(s):

	R2-chng	F	df1	df2	p
X*W	.0822	5.8474	2.0000	123.0000	.0038

Test of moderation

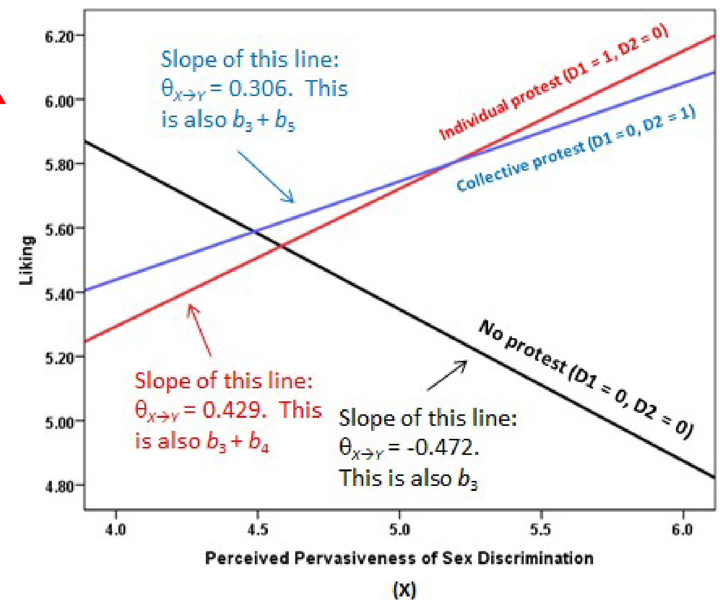
PROCESS Output

PROCESS estimates the effect of in the three groups defined by *W*, along with standard errors, *t* and *p*-values, and confidence intervals.

$$\theta_{X\rightarrow Y} = b_3 + b_4D_1 + b_5D_2$$
$$= -0.472 + 0.901D_1 + 0.778D_2$$

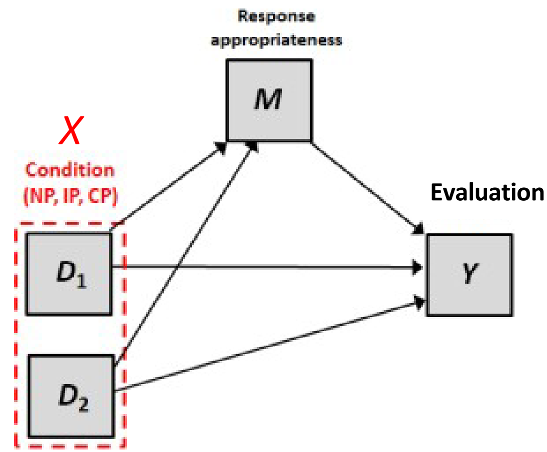
```
     Focal predict: sexism (X)
           Mod var: protest  (W)

Conditional effects of the focal predictor at values of the moderator(s):

     protest     Effect          se           t          p       LLCI       ULCI
       .0000     -.4725       .2053     -2.3017      .0230     -.8788     -.0662
      1.0000      .4287       .2013      2.1298      .0352      .0303      .8272
      2.0000      .3053       .1833      1.6660      .0983     -.0574      .6681

Data for visualizing the conditional effect of the focal predictor:
Paste text below into a SPSS syntax window and execute to produce plot.

DATA LIST FREE/
  sexism       protest     eval    .
 BEGIN DATA.
      4.2500       .0000    5.6981
      5.1200       .0000    5.2871
      5.8960       .0000    4.9204
      4.2500      1.0000    5.3996
      5.1200      1.0000    5.7726
      5.8960      1.0000    6.1053
      4.2500      2.0000    5.5131
      5.1200      2.0000    5.7787
      5.8960      2.0000    6.0156

END DATA.
GRAPH/SCATTERPLOT=
 sexism    WITH      eval      BY        protest  .
```

slopes of
these lines

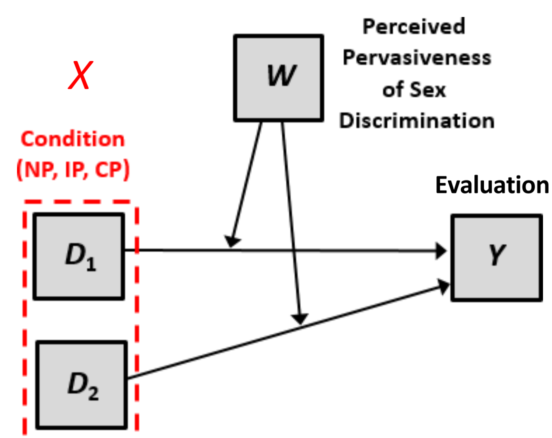


Integrating mediation with a multicategorical *X* with moderation

We've seen how to estimate direct and indirect effects in a mediation model when the causal antecedent *X* is a multicategorical variable.



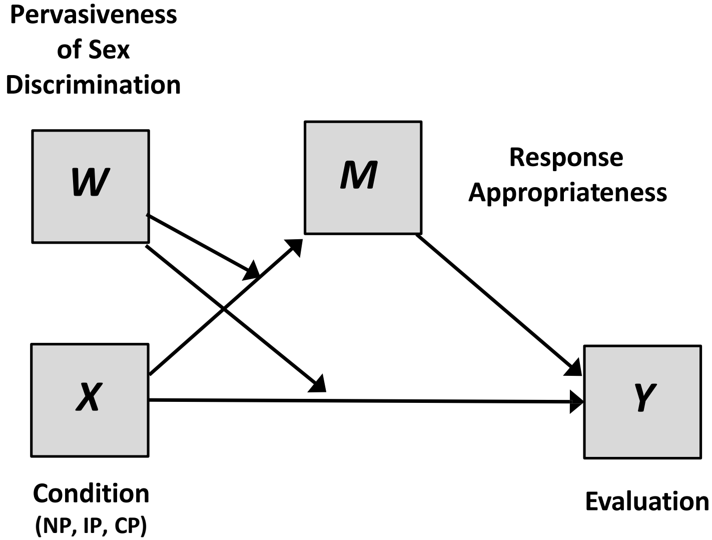
We've seen how to estimate the moderation of the effect of a multicategorical variable.



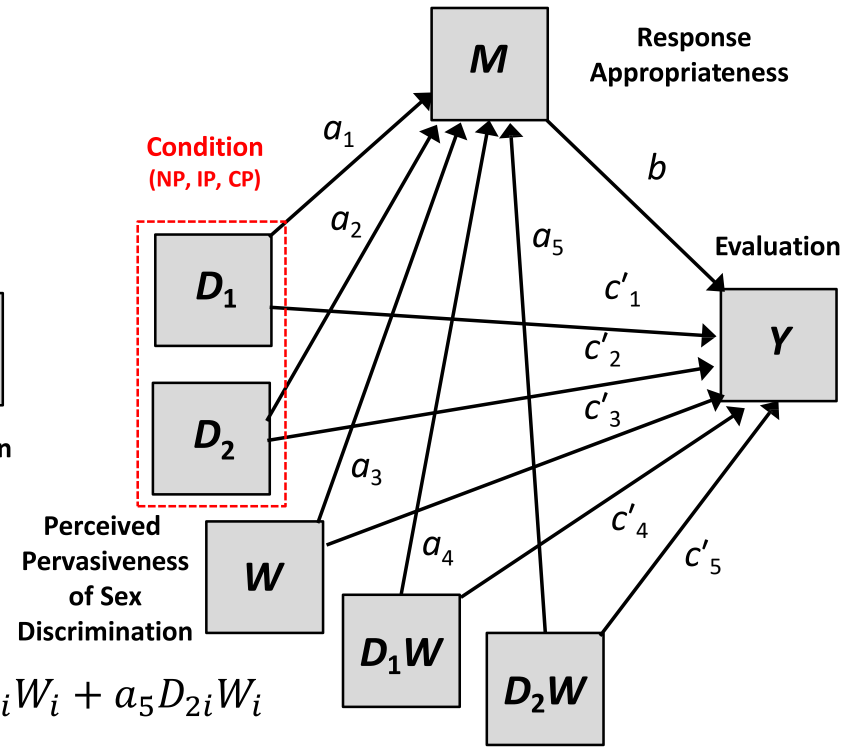
We now integrate these by estimating a conditional process model with a multicategorical *X*, moderation in the first stage by *W*, as well as moderation of the direct effect of *X*.

A first stage moderated mediation model with a multicategorical independent variable with three levels

The Conceptual Model



The Statistical Model



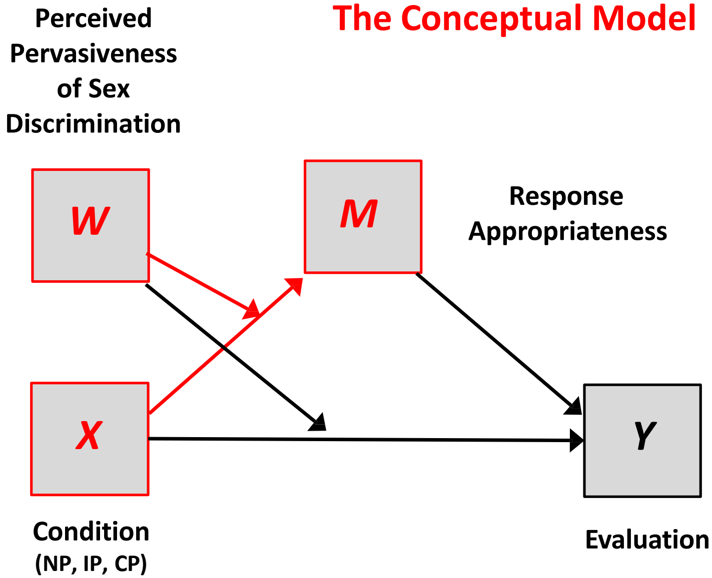
$$\widehat{M_i} = a_0 + a_1 D_{1i} + a_2 D_{2i} + a_3 W_i + a_4 D_{1i} W_i + a_5 D_{2i} W_i$$

$$\widehat{Y_i} = c'_0 + c'_1 D_{1i} + c'_2 D_{2i} + c'_3 W_i + c'_4 D_{1i} W_i + c'_5 D_{2i} W_i + b M_i$$

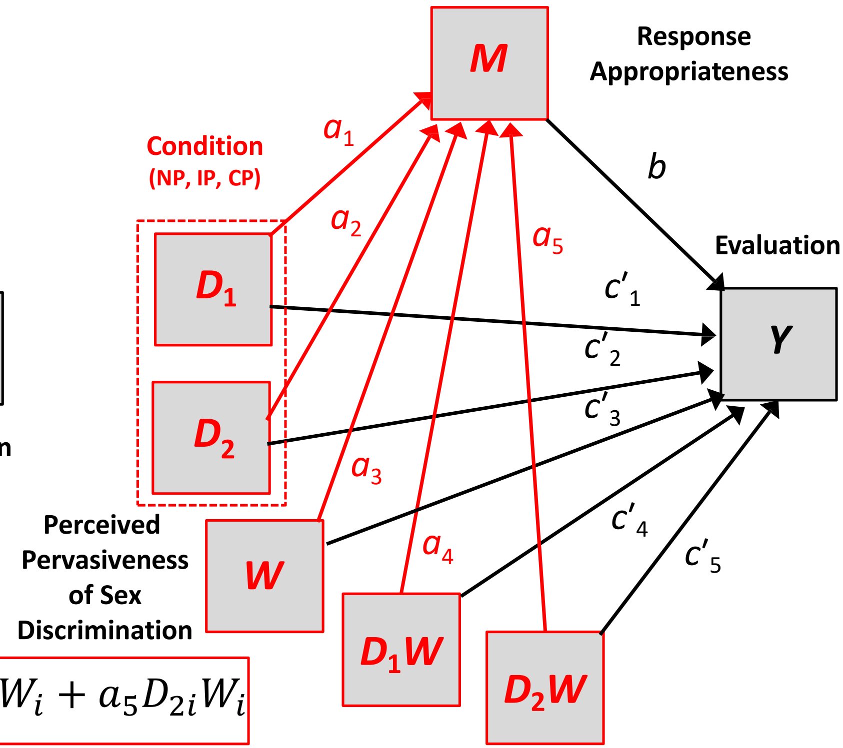
This model allows for moderation of the indirect and direct effects of experimental condition on evaluation by perceptions of the pervasiveness of sex discrimination in society.

A first stage moderated mediation model with a multicategorical independent variable with three levels

The Conceptual Model



The Statistical Model

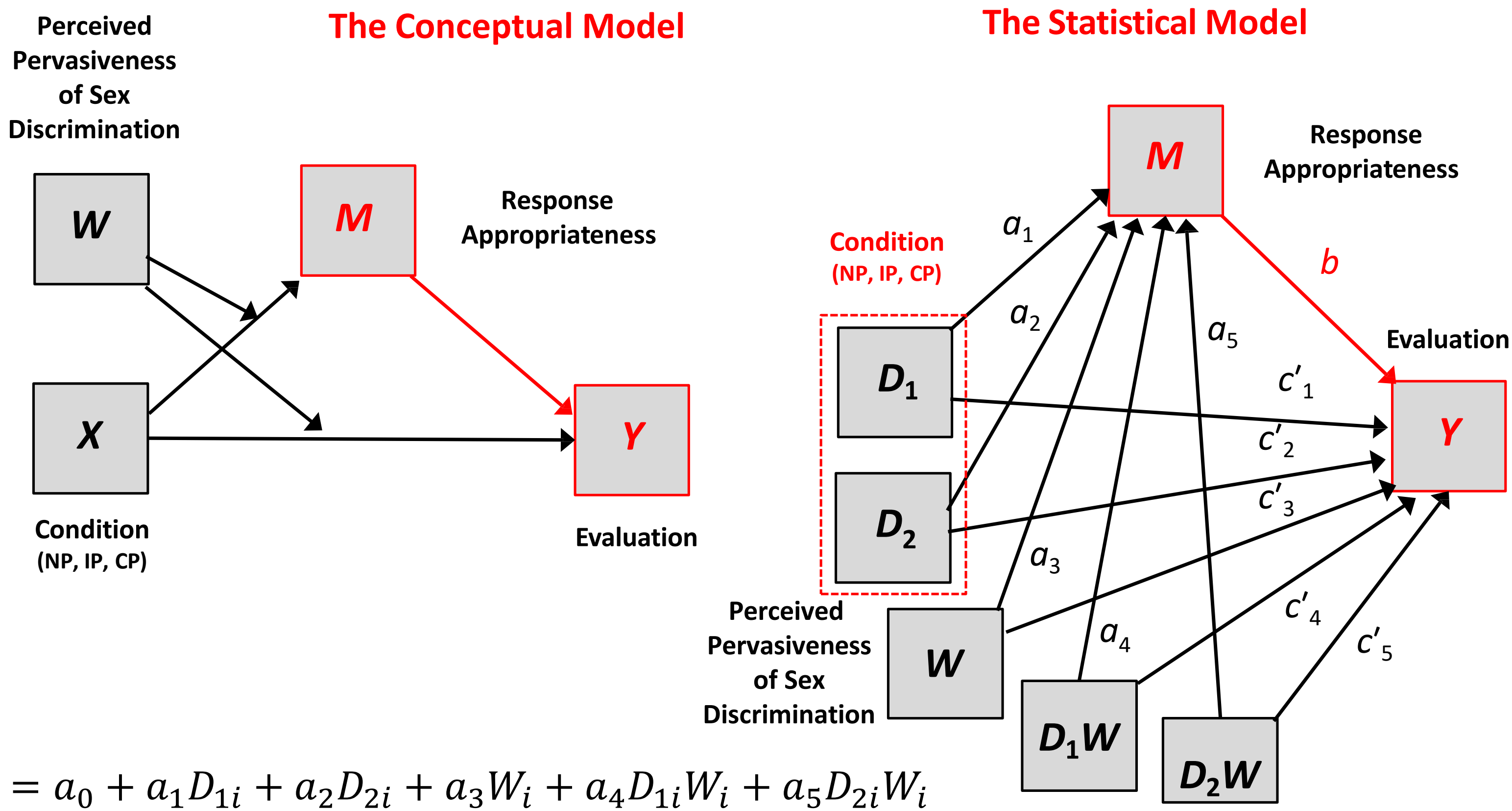


$$\widehat{M}_i = a_0 + \boxed{a_1 D_{1i} + a_2 D_{2i} + a_3 W_i + a_4 D_{1i} W_i + a_5 D_{2i} W_i}$$

$$\widehat{Y}_i = c'_0 + c'_1 D_{1i} + c'_2 D_{2i} + c'_3 W_i + c'_4 D_{1i} W_i + c'_5 D_{2i} W_i + b M_i$$

The moderation of the effect of experimental condition on perceived response appropriateness by beliefs about the pervasiveness of sex discrimination in society.

A first stage moderated mediation model with a multicategorical independent variable with three levels

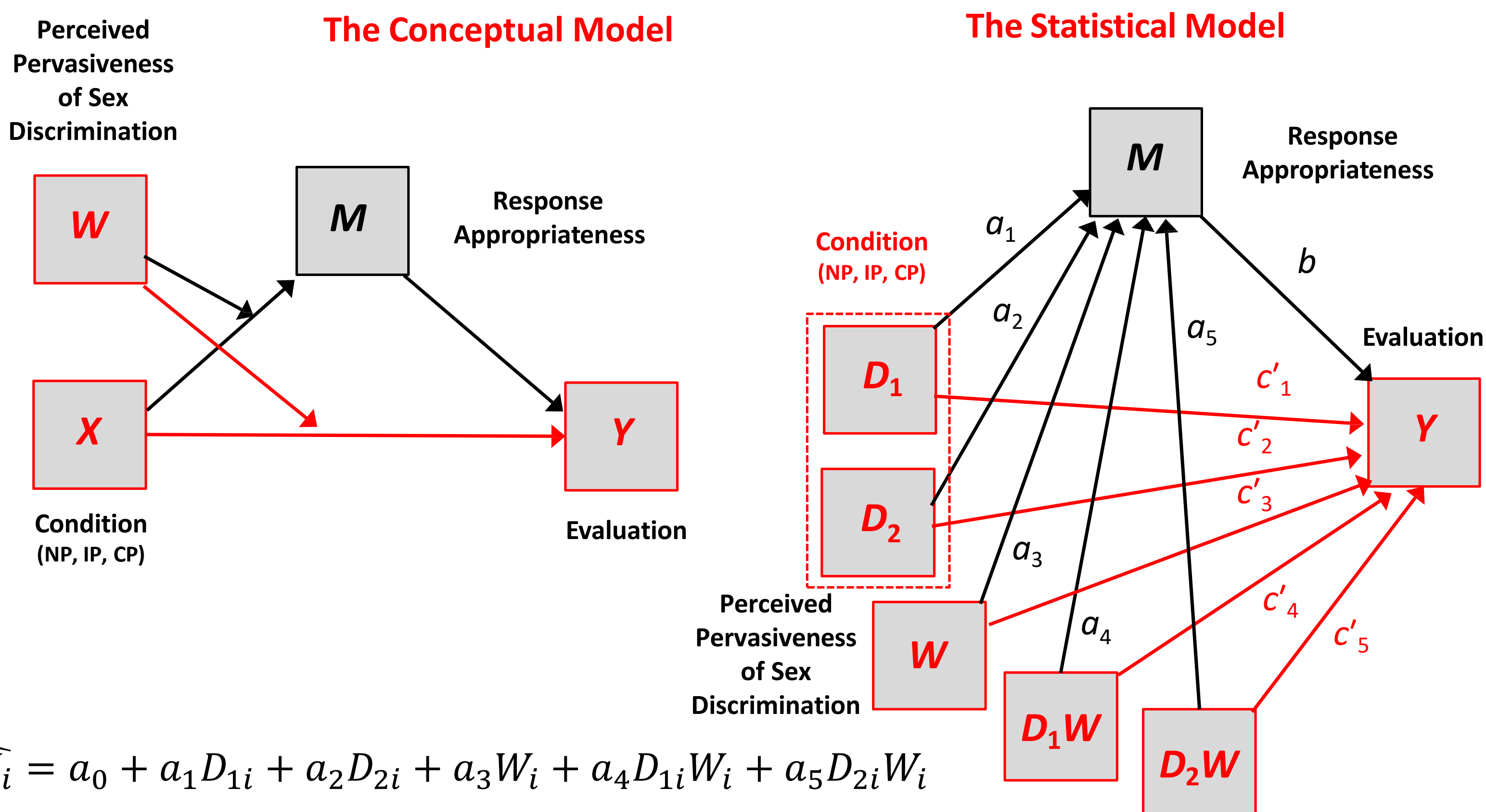


$$\widehat{M}_i = a_0 + a_1 D_{1i} + a_2 D_{2i} + a_3 W_i + a_4 D_{1i} W_i + a_5 D_{2i} W_i$$

$$\widehat{Y}_i = c'_0 + c'_1 D_{1i} + c'_2 D_{2i} + c'_3 W_i + c'_4 D_{1i} W_i + c'_5 D_{2i} W_i \boxed{+ bM_i}$$

The effect of perceived response appropriateness on evaluation of the attorney.

A first stage moderated mediation model with a multicategorical independent variable with three levels

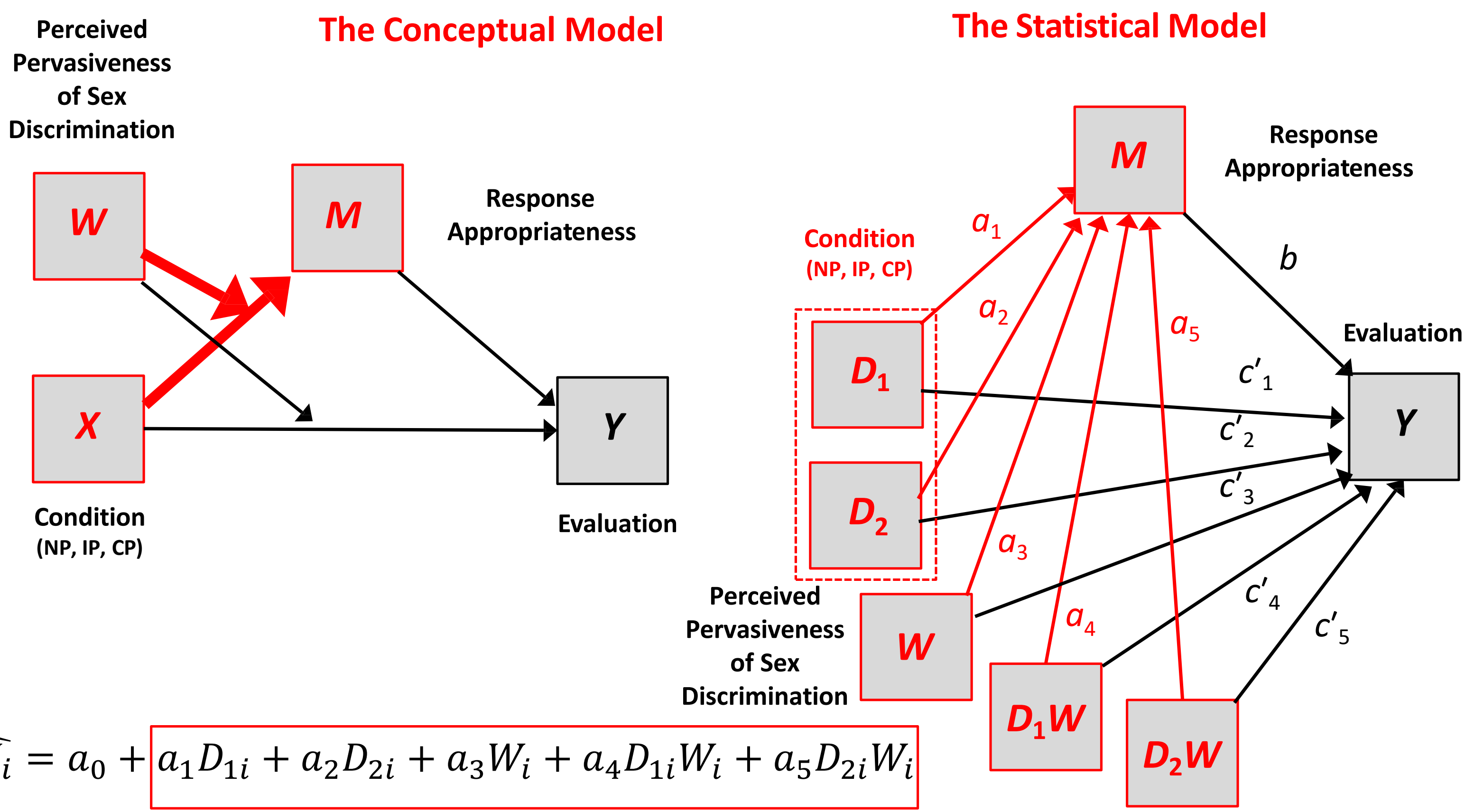


$$\widehat{M}_i = a_0 + a_1 D_{1i} + a_2 D_{2i} + a_3 W_i + a_4 D_{1i} W_i + a_5 D_{2i} W_i$$

$$\widehat{Y}_i = c'_0 + \boxed{c'_1 D_{1i} + c'_2 D_{2i} + c'_3 W_i + c'_4 D_{1i} W_i + c'_5 D_{2i} W_i} + bM_i$$

The moderation of the direct effect of experimental condition on evaluation of the attorney by perceived pervasiveness of sex discrimination in society.

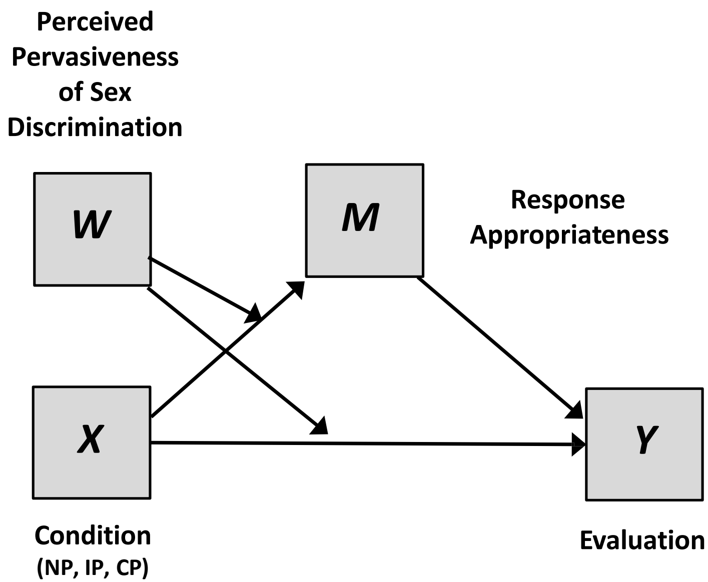
Estimating the moderation of the first stage effect

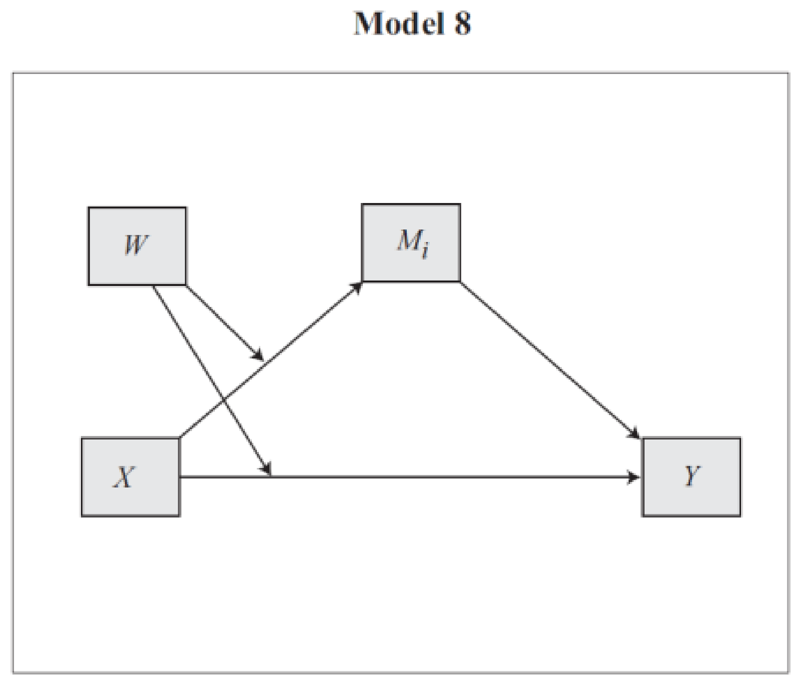


$$\widehat{M_i} = a_0 + \boxed{a_1 D_{1i} + a_2 D_{2i} + a_3 W_i + a_4 D_{1i} W_i + a_5 D_{2i} W_i}$$

$$\widehat{Y_i} = c'_0 + c'_1 D_{1i} + c'_2 D_{2i} + c'_3 W_i + c'_4 D_{1i} W_i + c'_5 D_{2i} W_i + b M_i$$

In PROCESS






```
process y=eval/x=protest/m=approp/w=sexism/model=8/center=1/
plot=1/mcx=1/ seed=193456.
```

```
%process (data=lawyer2,y=eval,x=protest,m=approp,w=sexism,model=8,
  center = 1,plot=1,mcx=1,seed=193456)
```

```
process(data=lawyer2,y="eval",x="protest",m="approp",w="sexism",
  model=8,center = 1,plot=1,mcx=1,seed=193456)
```

PROCESS Output

$a_1 = 1.2282$
$a_2 = 1.6510$
$a_4 = 0.978$
$a_5 = 0.734$

```
Model: 8
    Y: eval
    X: protest
    M: approp
    W: sexism

Sample size: 129
Custom seed: 193456
Coding of categorical X variable for analysis:
   protest       X1        X2
    0.0000    0.0000    0.0000
    1.0000    1.0000    0.0000
    2.0000    0.0000    1.0000

Outcome Variable: approp

Model Summary:
        R      R-sq       MSE         F       df1       df2         p
    0.5618    0.3156    1.2945   11.3424    5.0000  123.0000    0.0000

Model:
              coeff        se         t         p      LLCI      ULCI
constant     3.8598    0.1780   21.6827    0.0000    3.5074    4.2122
X1           1.2282    0.2503    4.9069    0.0000    0.7328    1.7237
X2           1.6510    0.2464    6.6994    0.0000    1.1632    2.1389
sexism      -0.5290    0.2345   -2.2559    0.0258   -0.9932   -0.0648
int_1        0.9778    0.3284    2.9771    0.0035    0.3277    1.6279
int_2        0.7339    0.3144    2.3347    0.0212    0.1117    1.3562

Product terms key:
int_1   :  X1   x   sexism
int_2   :  X2   x   sexism

Test(s) of highest order unconditional interaction(s):
      R2-chng         F       df1       df2         p
X*W    0.0537    4.8233    2.0000  123.0000    0.0096
```

a_1 → X1, a_2 → X2, a_3 → sexism, a_4 → int_1, a_5 → int_2

PROCESS creates the two products automatically

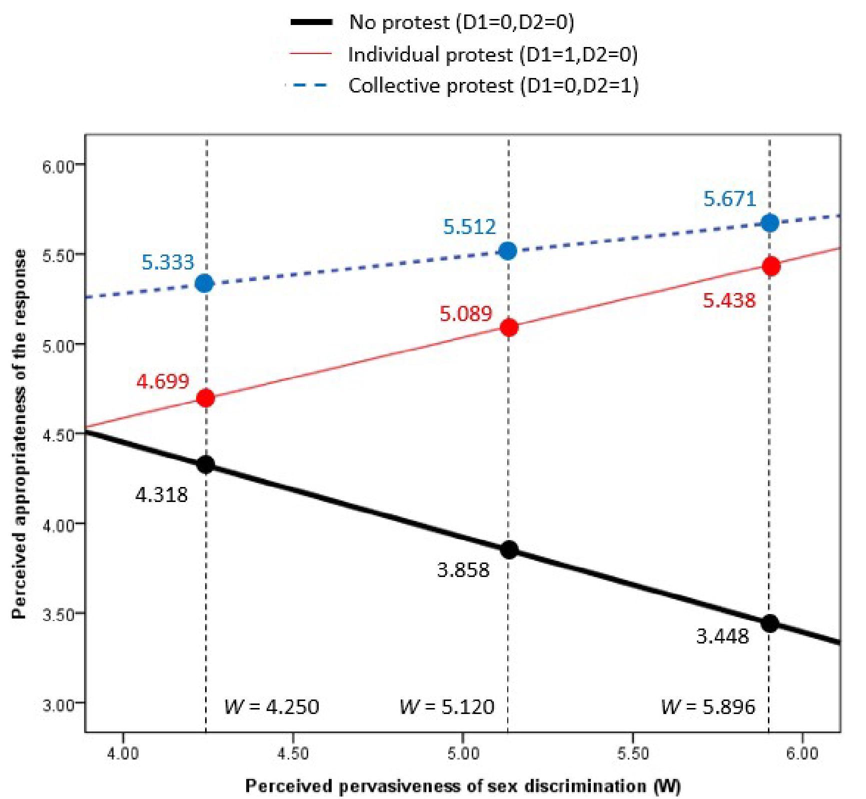
Test of interaction Between *X* and *W*

A visual representation of the model of *M*

```
DATA LIST FREE/
   protest     sexism      approp      .
BEGIN DATA.
     .0000      4.2500      4.3184
    1.0000      4.2500      4.6990
    2.0000      4.2500      5.3332
     .0000      5.1200      3.8582
    1.0000      5.1200      5.0894
    2.0000      5.1200      5.5115
     .0000      5.8960      3.4477
    1.0000      5.8960      5.4377
    2.0000      5.8960      5.6705
END DATA.
GRAPH/SCATTERPLOT=
 sexism   WITH      approp   BY  protest   .
```

PROCESS for SPSS writes the code to visualize this interaction. Cut and paste into SPSS syntax and execute. Edit to make it look pretty.

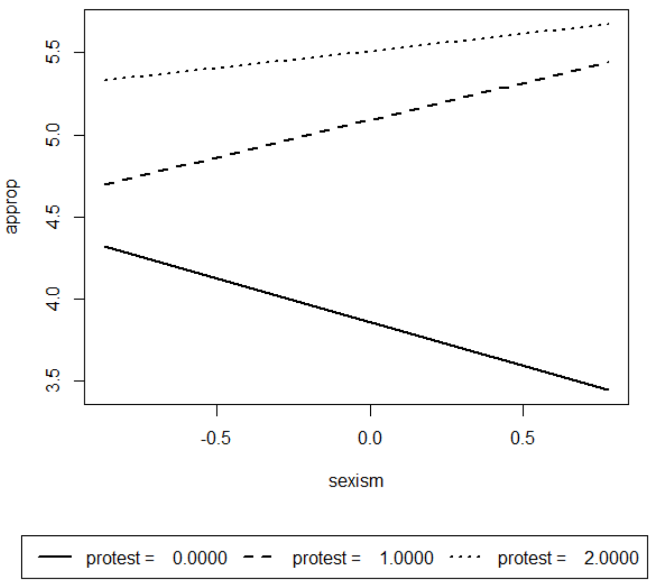
```
data;
input protest sexism approp;
datalines;
      .0000     4.2500      4.3184
     1.0000     4.2500      4.6990
     2.0000     4.2500      5.3332
      .0000     5.1200      3.8582
     1.0000     5.1200      5.0894
     2.0000     5.1200      5.5115
      .0000     5.8960      3.4477
     1.0000     5.8960      5.4377
     2.0000     5.8960      5.6705
run;
proc sgplot;reg x=sexism y=approp/group=protest;run;
```

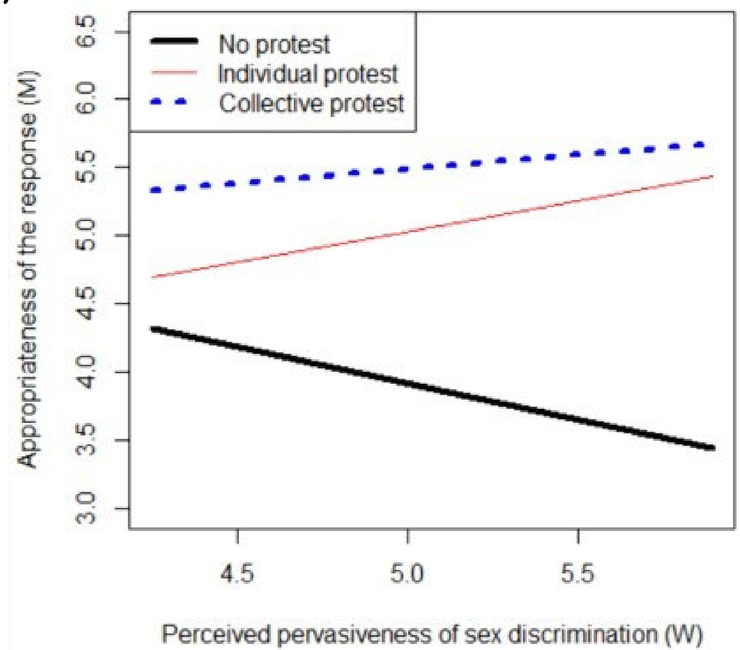


A visual representation of the model of *M* (in R)

```
x<-c(0,1,2,0,1,2,0,1,2)
w<-c(4.25,4.25,4.25,5.12,5.12,5.12,5.896,5.896,5.896)
m<-c(4.318,4.699,5.333,3.858,5.089,5.512,3.448,5.438,5.671)
plot(y=m,x=w,pch=15,col="white",ylim=c(3,6.5),
xlab="Perceived pervasiveness of sex discrimination (W)",
ylab="Appropriateness of the response (M)")
legend.txt<-c("No protest","Individual protest","Collective protest")
legend("topleft",legend=legend.txt,lty=c(1,1,3),lwd=c(4,1,4),
col=c("black","red","blue"))
lines(w[x==0],m[x==0],lwd=4,lty=1,col="black")
lines(w[x==1],m[x==1],lwd=1,lty=1,col="red")
lines(w[x==2],m[x==2],lwd=4,lty=3,col="blue")
```

From PROCESS plot output





Relative conditional effects of protesting on perceived appropriateness of the response

Model:

	coeff	se	t	p	LLCI	ULCI
constant	3.8598	0.1780	21.6827	0.0000	3.5074	4.2122
X1	1.2282	0.2503	4.9069	0.0000	0.7328	1.7237
X2	1.6510	0.2464	6.6994	0.0000	1.1632	2.1389
sexism	-0.5290	0.2345	-2.2559	0.0258	-0.9932	-0.0648
int_1	0.9778	0.3284	2.9771	0.0035	0.3277	1.6279
int_2	0.7339	0.3144	2.3347	0.0212	0.1117	1.3562

$$\widehat{M}_i = 3.86 + 1.23D_{1i} + 1.65D_{2i} - 0.529W_i + 0.978D_{1i}W_i + 0.734D_{2i}W_i$$

which can be written equivalently as

$$\widehat{M}_i = 3.86 + (\underset{a_1}{1.23} + \underset{a_4}{0.978}W_i)D_{1i} + (\underset{a_2}{1.65} + \underset{a_5}{0.734}W_i)D_{2i} - 0.529W_i$$

or

$$\widehat{M}_i = 3.86 + \theta_{D_1\rightarrow M}D_{1i} + \theta_{D_2\rightarrow M}D_{2i} - 0.529W_i$$

where

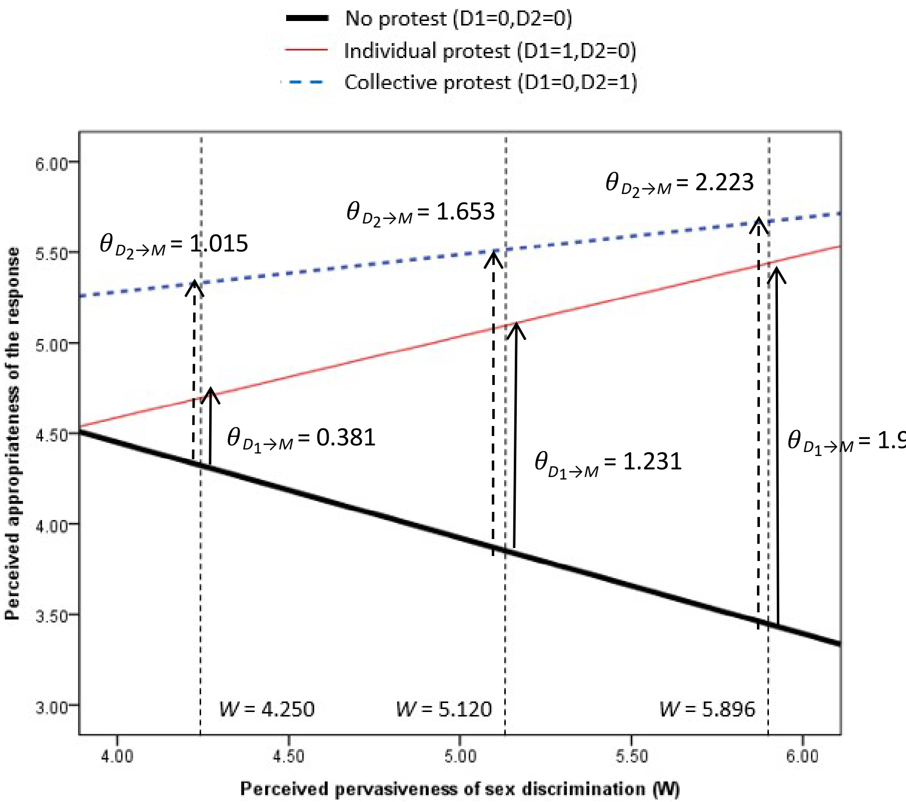
$$\theta_{D_1\rightarrow M} = a_1 + a_4W = 1.23 + 0.978W$$

$$\theta_{D_2\rightarrow M} = a_2 + a_5W = 1.65 + 0.734W$$

$\theta_{D_1\rightarrow M}$ is the conditional effect of individual protest (relative to not protesting) and $\theta_{D_2\rightarrow M}$ is the conditional effect of collective protest (relative to not protesting). They are both linear functions of perceived pervasiveness of sex discrimination.

Conditional effects of protesting on perceived appropriateness of the response

$$\widehat{M_i} = 6.567 + (-3.775 + 0.978W_i)D_{1i} + (-2.104 + 0.734W_i)D_{2i} - 0.529W_i$$



$$\theta_{D_1 \to M} = -3.775 + 0.978W$$

$$\theta_{D_2 \to M} = -2.104 + 0.734W$$

SEXISM (W)	$\theta_{D_1 \to M}$	$\theta_{D_2 \to M}$
4.250	0.381	1.015
5.120	1.231	1.653
5.896	1.990	2.223

16th, 50th, and 84th percentiles

Protesting is generally seen as more appropriate than not, except for among those relatively low in perceived pervasiveness of sex discrimination. For such people, individual protest is not seen as any more or less appropriate than not protesting. The effect of either kind of protest is larger among those who see sex discrimination as more pervasive. The difference in the effect of protest form is larger among those who see sex discrimination as less pervasive.

The relative conditional effect of X on M when $W' = -0.867$, $W = 4.250$

```
Moderator value(s):
sexism   -0.8670
```

16th percentile

$\theta_{D_1 \to M} = -3.775 + 0.978W = -3.775 + 0.978(4.250) = 0.381$

$\theta_{D_2 \to M} = -2.104 + 0.734W = -2.104 + 0.734(4.250) = 1.015$

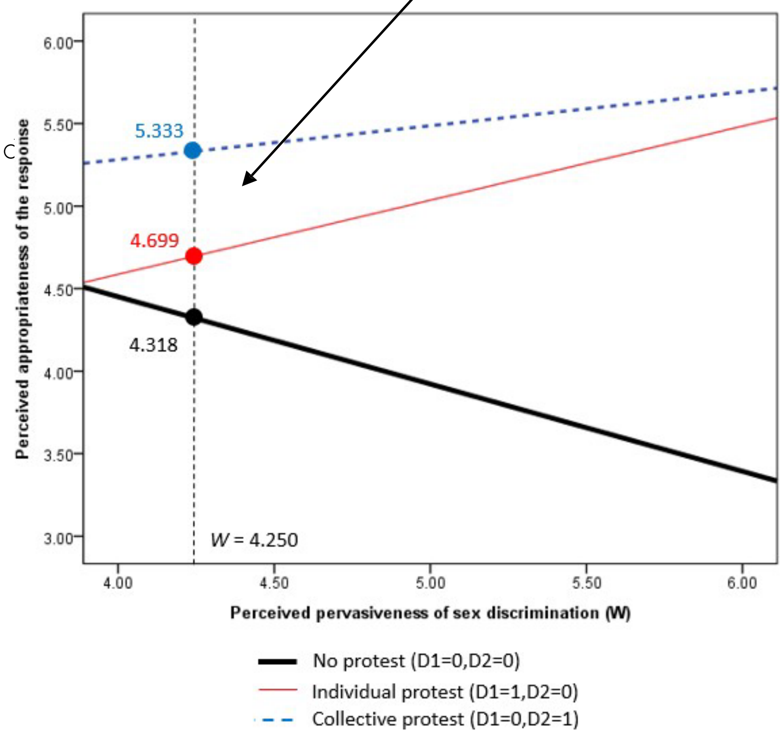
	effect	se	t	p	LLCI	ULCI
X1	0.3805	0.3887	0.9790	0.3295	-0.3888	1.1498
X2	1.0148	0.3529	2.8754	0.0048	0.3162	1.7133

Test of equality of conditional means

F	df1	df2	p
4.2751	2.0000	123.0000	0.0160

Estimated conditional means being compared

protest	approp
0.0000	**4.3184**
1.0000	**4.6990**
2.0000	**5.3332**



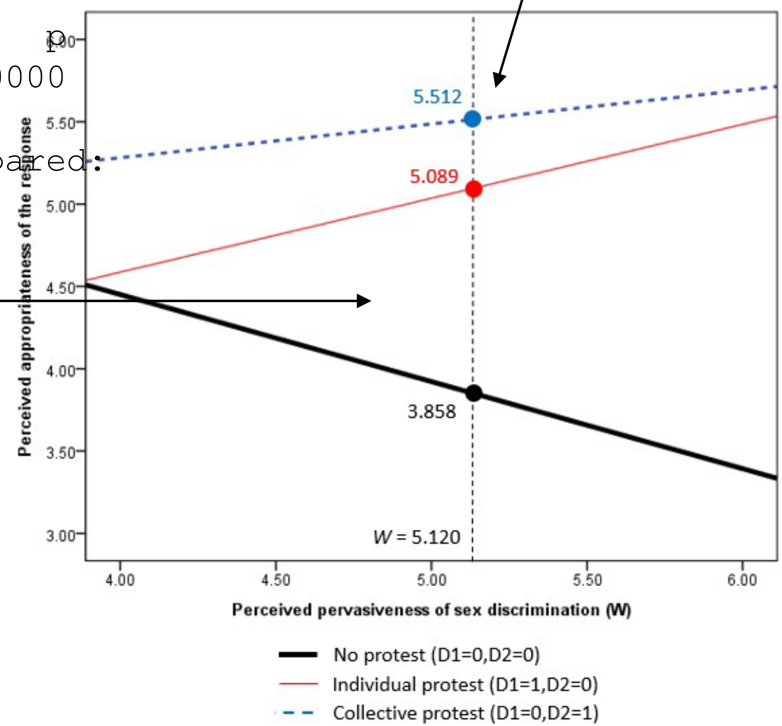
The relative conditional effect of *X* on *M* when *W'* = -0.003, *W* = 4.250

```
Moderator value(s):
```
50th percentile
```
sexism    0.0030
```

$\theta_{D1 \to M} = -3.775 + 0.978W = -3.775 + 0.978(5.120) = 1.231$

$\theta_{D2 \to M} = -2.104 + 0.734W = -2.104 + 0.734(5.120) = 1.653$

```
      effect       se          t          p       LLCI       ULCI
X1    1.2312   0.2503     4.9197     0.0000     0.7358     1.7266
X2    1.6533   0.2465     6.7063     0.0000     1.1653     2.1412

Test of equality of conditional means
         F        df1        df2          p
   23.9623     2.0000   123.0000     0.0000

Estimated conditional means being compared:
   protest     approp
    0.0000     3.8582
    1.0000     5.0894
    2.0000     5.5115
```



The relative conditional effect of *X* on *M* when *W* = 5.896

```
Moderator value(s):
```

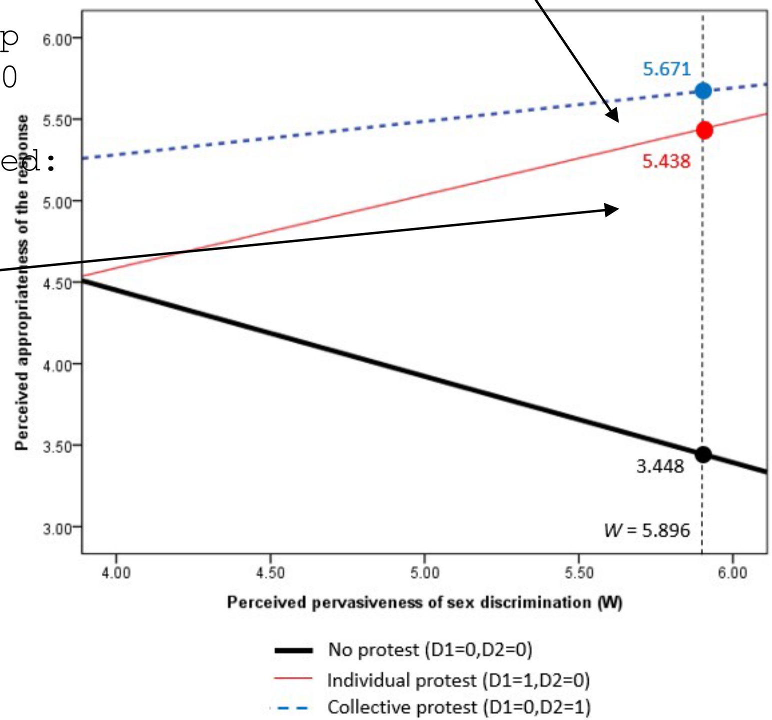
84th percentile

```
sexism    0.7790
```

$\theta_{D_1 \rightarrow M}$ = -3.775 + 0.978*W* = -3.775 + 0.978(5.896) = 1.990

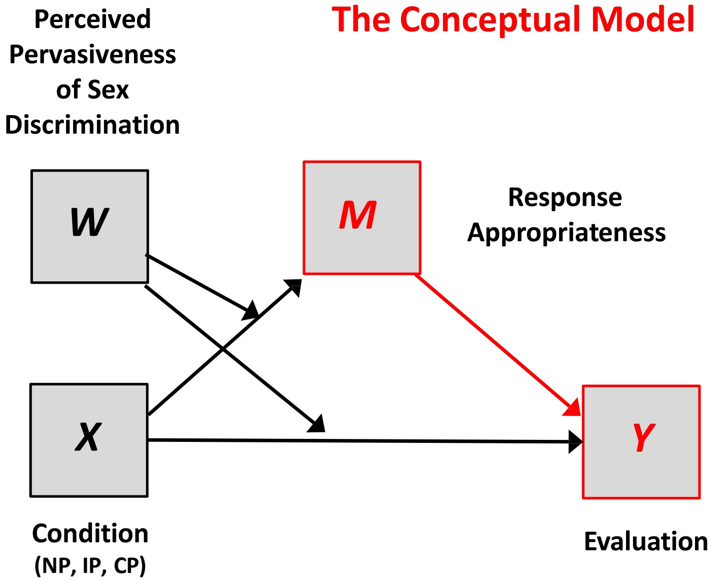
$\theta_{D_2 \rightarrow M}$ = -2.104 + 0.734*W* = -2.104 + 0.734(5.896) = 2.223

```
      effect        se         t         p      LLCI      ULCI
X1    1.9900    0.3486    5.7080    0.0000    1.2999    2.6801
X2    2.2228    0.3607    6.1621    0.0000    1.5088    2.9368

Test of equality of conditional means
         F       df1       df2         p
   22.8591    2.0000  123.0000    0.0000

Estimated conditional means being compared:
   protest    approp
    0.0000    3.4477
    1.0000    5.4377
    2.0000    5.6705
```

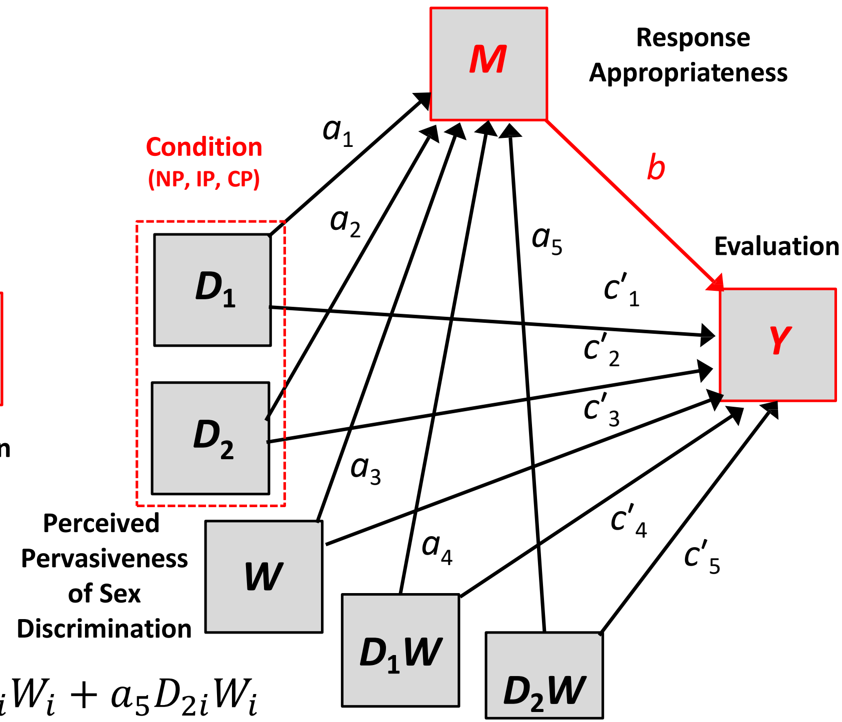


Estimating the effect of *M* on *Y* (path *b*)

The Conceptual Model



The Statistical Model



$$\widehat{M}_i = a_0 + a_1 D_{1i} + a_2 D_{2i} + a_3 W_i + a_4 D_{1i} W_i + a_5 D_{2i} W_i$$

$$\widehat{Y}_i = c_0' + c_1' D_{1i} + c_2' D_{2i} + c_3' W_i + c_4' D_{1i} W_i + c_5' D_{2i} W_i \boxed{+ bM_i}$$

The effect of perceived response appropriateness on evaluation of the attorney.

PROCESS Output

Outcome Variable: eval

Model Summary:

R	R-sq	MSE	F	df1	df2	p
0.5355	0.2868	0.8245	8.1767	6.0000	122.0000	0.0000

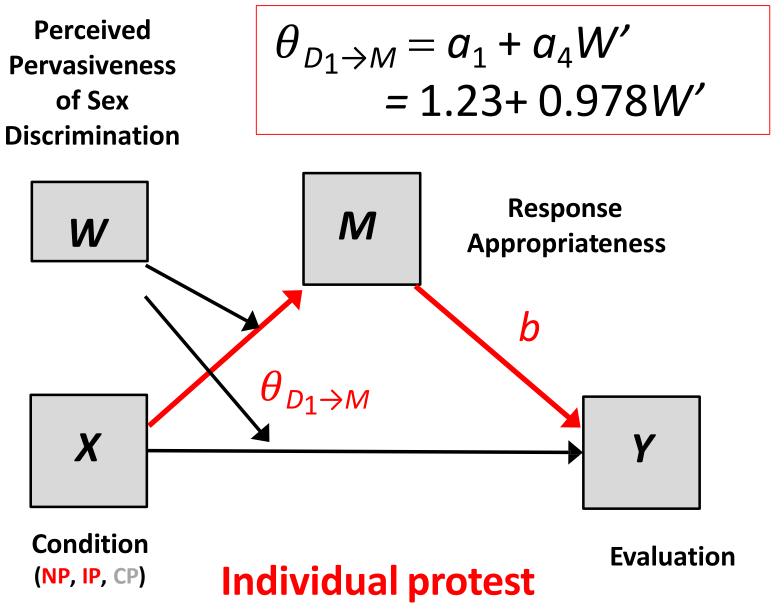
Model:

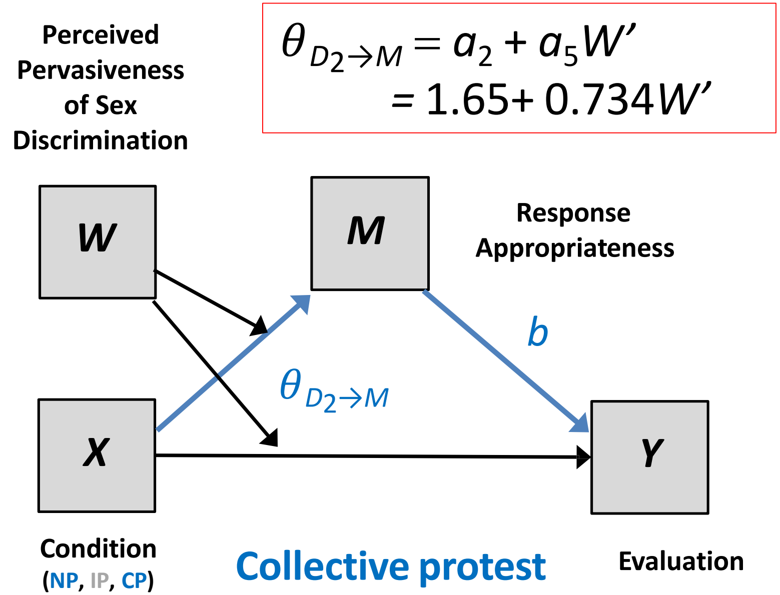
	coeff	se	t	p	LLCI	ULCI
constant	3.8728	0.3120	12.4137	0.0000	3.2552	4.4904
X1	0.0323	0.2184	0.1479	0.8826	-0.4001	0.4648
X2	-0.1163	0.2298	-0.5061	0.6137	-0.5712	0.3386
approp	0.3668	0.0720	5.0969	0.0000	0.2243	0.5092
sexism	-0.2785	0.1910	-1.4581	0.1474	-0.6565	0.0996
int_1	0.5426	0.2714	1.9992	0.0478	0.0053	1.0799
int_2	0.5086	0.2564	1.9839	0.0495	0.0011	1.0162

$b = 0.367$

For two participants in the same condition and with the same level of belief in the pervasiveness of sexism, but who differ by one unit in perceived appropriateness of the lawyers response, the person with the higher perceived appropriateness is expected to have a .367 unit higher evaluation of the lawyer.

The relative conditional indirect effect

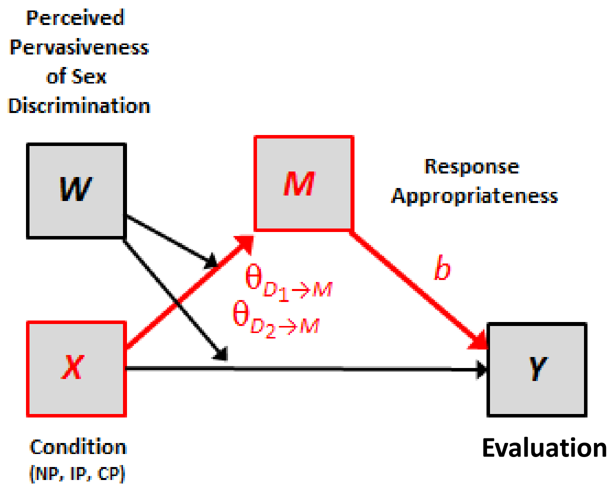




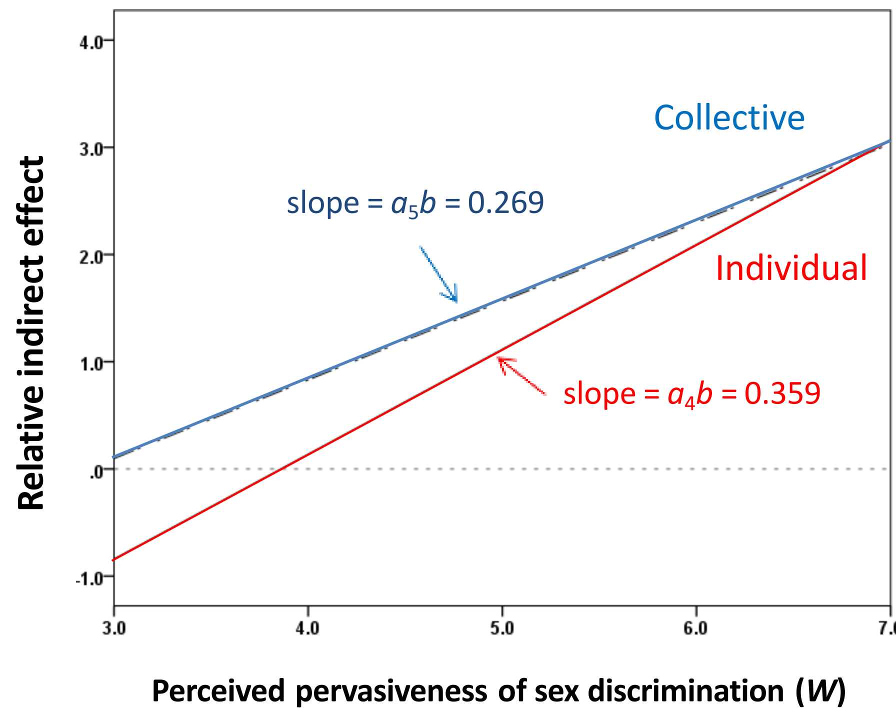
The relative conditional indirect effect of **individual protest** (relative to no protest) on evaluation (Y) through perceived response appropriateness (M) is the product of the relative condtional effect of individual protest on response appropriateness ($\theta_{D_1 \to M}$) and the effect of response appropriateness on evaluation (b): $(\theta_{D_1 \to M})b = (a_1 + a_4 W')b = (-3.775 + 0.978W')0.367$

The relative conditional indirect effect of **collective protest** (relative to no protest) on evaluation (Y) through perceived response appropriateness (M) is the product of the relative condtional effect of collective protest on response appropriateness ($\theta_{D_2 \to M}$) and the effect of response appropriateness on evaluation (b): $(\theta_{D_2 \to M})b = (a_2 + a_5 W')b = (-2.104 + 0.734W')0.367$

A visual representation



$$\theta_{D_1 \to M}\, b = (a_1 + a_4 W)b = a_1 b + a_4 b W$$
$$= -1.385 + 0.359W$$
$$\theta_{D_2 \to M}\, b = (a_2 + a_5 W)b = a_2 b + a_5 b W$$
$$= -0.772 + 0.269W$$



Both relative indirect effects are increasingly positive as perceived pervasiveness of sex discrimination increases. The "**indices of moderated mediation**" are $a_4b = 0.359$ for individual protest and $a_5b = 0.269$ for collective protest.

A hypothesis test that the slope ---the index of moderated mediation---is equal to 0 is a formal test of moderated mediation....moderation of the *relative* indirect effect.

PROCESS Output: Indices of moderated mediation

```
INDIRECT EFFECT:

protest     ->     approp     ->     eval

        sexism     Effect     BootSE   BootLLCI   BootULCI
X1     -0.8670     0.1396     0.1601    -0.1759     0.4604
X1      0.0030     0.4516     0.1301     0.2163     0.7176
X1      0.7790     0.7299     0.2049     0.3413     1.1464

     Index of moderated mediation:
              Index     BootSE   BootLLCI   BootULCI
sexism       0.3586     0.1603     0.0661     0.6877

        sexism     Effect     BootSE   BootLLCI   BootULCI
X2     -0.8670     0.3722     0.1476     0.1118     0.6960
X2      0.0030     0.6064     0.1492     0.3249     0.9191
X2      0.7790     0.8153     0.2220     0.4019     1.2735

     Index of moderated mediation:
              Index     BootSE   BootLLCI   BootULCI
sexism       0.2692     0.1444     0.0006     0.5748
```

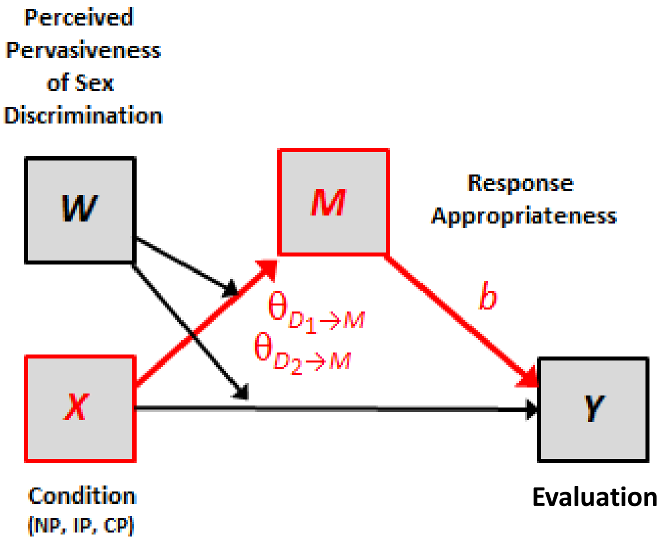
$a_4b = 0.359$

95% bootstrap confidence intervals for test of moderation of the relative indirect effect of individual protest by perceived pervasiveness of sex discrimination

$a_5b = 0.269$

95% bootstrap confidence intervals for test of moderation of the relative indirect effect of collective protest by perceived pervasiveness of sex discrimination

Now we can probe moderation of mediation



<u>**Individual**</u>

$$\theta_{D_1 \to M}\, b = (a_1 + a_4 W)b$$
$$= (-3.775 + 0.978W)0.367$$

<u>Collective</u>

$$\theta_{D_2 \to M}\, b = (a_2 + a_5 W)b$$
$$= (-2.104 + 0.734W)0.367$$

Using these functions, we can estimate the indirect effect for any value of the moderator we choose

	***SEXISM* (*W*)**	$\boldsymbol{\theta_{D_1 \to M}}$	$\boldsymbol{\theta_{D_2 \to M}}$	***b***	$\boldsymbol{\theta_{D_1 \to M}\, b}$	$\boldsymbol{\theta_{D_2 \to M}\, b}$
16^{th}	4.250	0.381	1.015	0.367	0.140	0.372
50^{th}	5.120	1.231	1.653	0.367	0.452	0.606
84^{th}	5.896	1.990	2.223	0.367	0.730	0.815

Relative conditional indirect effects

We can use bootstrap confidence intervals for inference about these relative conditional indirect effects. PROCESS will help.

PROCESS Output: Indices of moderated mediation

```
INDIRECT EFFECT:

protest     ->     approp     ->     eval

       sexism      Effect      BootSE   BootLLCI   BootULCI
X1    -0.8670      0.1396      0.1601    -0.1759     0.4604
X1     0.0030      0.4516      0.1301     0.2163     0.7176
X1     0.7790      0.7299      0.2049     0.3413     1.1464

      Index of moderated mediation:
             Index     BootSE   BootLLCI   BootULCI
sexism      0.3586     0.1603     0.0661     0.6877

       sexism      Effect      BootSE   BootLLCI   BootULCI
X2    -0.8670      0.3722      0.1476     0.1118     0.6960
X2     0.0030      0.6064      0.1492     0.3249     0.9191
X2     0.7790      0.8153      0.2220     0.4019     1.2735

      Index of moderated mediation:
             Index     BootSE   BootLLCI   BootULCI
sexism      0.2692     0.1444     0.0006     0.5748
```

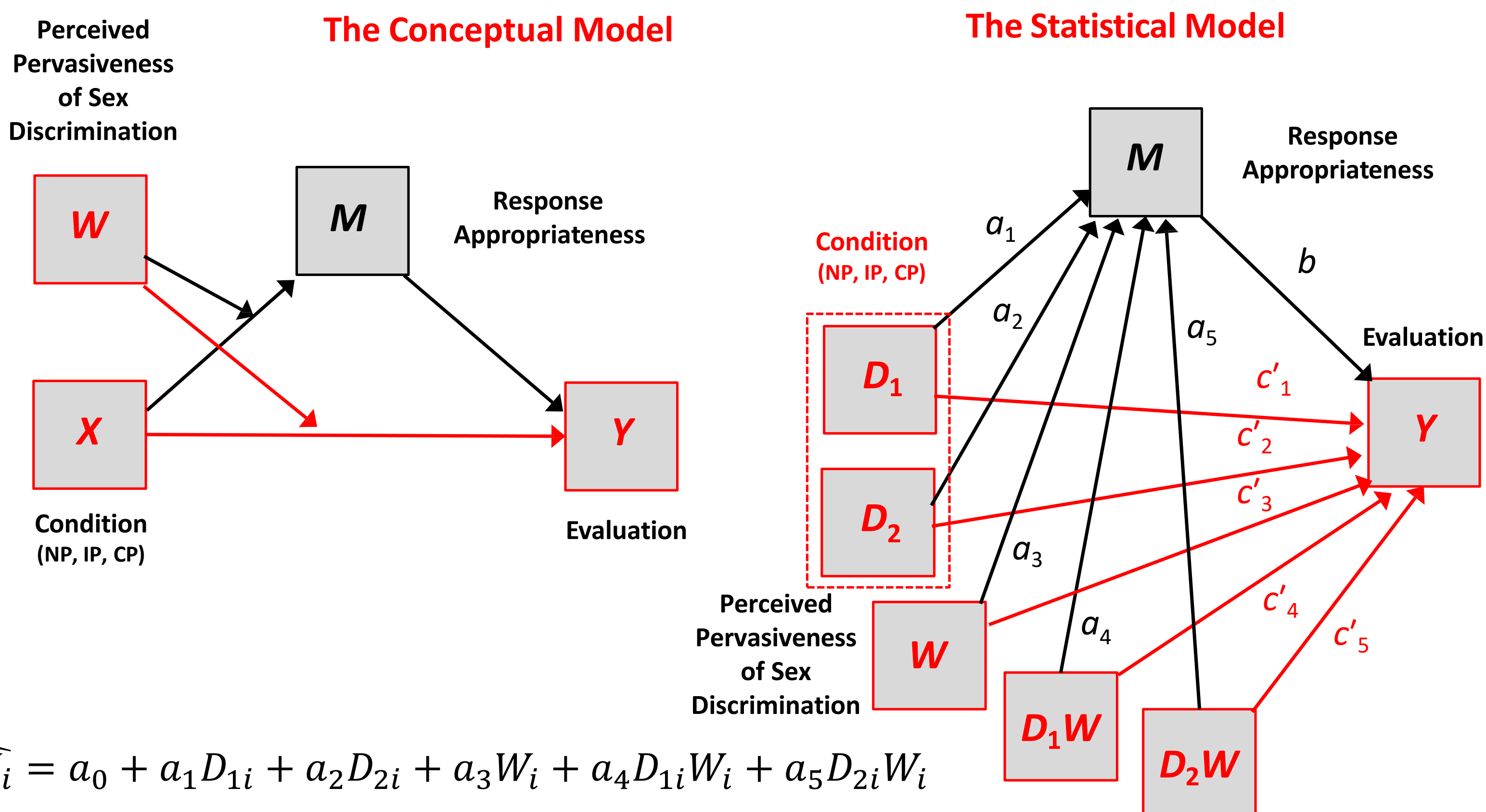
$\boldsymbol{\theta_{D1 \to M} b = (1.23 + 0.978W')0.367 = 0.45 + 0.359W'}$

95% bootstrap confidence intervals

$\boldsymbol{\theta_{D2 \to M} b = (1.65 + 0.734W')0.367 = 0.605 + 0.269W'}$

95% bootstrap confidence intervals

The (conditional) direct effect



$$\widehat{M_i} = a_0 + a_1 D_{1i} + a_2 D_{2i} + a_3 W_i + a_4 D_{1i} W_i + a_5 D_{2i} W_i$$

$$\widehat{Y_i} = c'_0 + \boxed{c'_1 D_{1i} + c'_2 D_{2i} + c'_3 W_i + c'_4 D_{1i} W_i + c'_5 D_{2i} W_i} + bM_i$$

The moderation of the direct effect of experimental condition on evaluation of the attorney by perceived pervasiveness of sex discrimination in society.

PROCESS Output

c'_1 = -2.744
c'_2 = -2.719
c'_4 = 0.543
c'_5 = 0.509

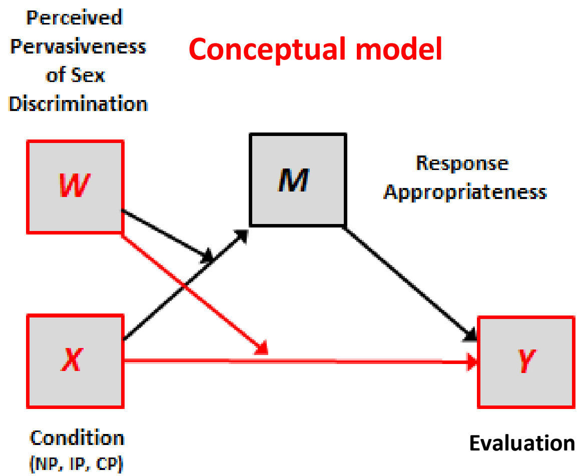
```
**************************************************************************
OUTCOME VARIABLE:
 eval

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .5355      .2868      .8245     8.1767     6.0000   122.0000      .0000

Model:
              coeff         se          t          p       LLCI       ULCI
constant     3.8728     0.3120    12.4137     0.0000     3.2552     4.4904
X1           0.0323     0.2184     0.1479     0.8826    -0.4001     0.4648
X2          -0.1163     0.2298    -0.5061     0.6137    -0.5712     0.3386
approp       0.3668     0.0720     5.0969     0.0000     0.2243     0.5092
sexism      -0.2785     0.1910    -1.4581     0.1474    -0.6565     0.0996
int_1        0.5426     0.2714     1.9992     0.0478     0.0053     1.0799
int_2        0.5086     0.2564     1.9839     0.0495     0.0011     1.0162

Product terms key:
 Int_1    :        X1        x        sexism
 Int_2    :        X2        x        sexism

Test(s) of highest order unconditional interaction(s):

        R2-chng          F        df1        df2          p
 X*W      .0298     2.5479     2.0000   122.0000      .0824
```

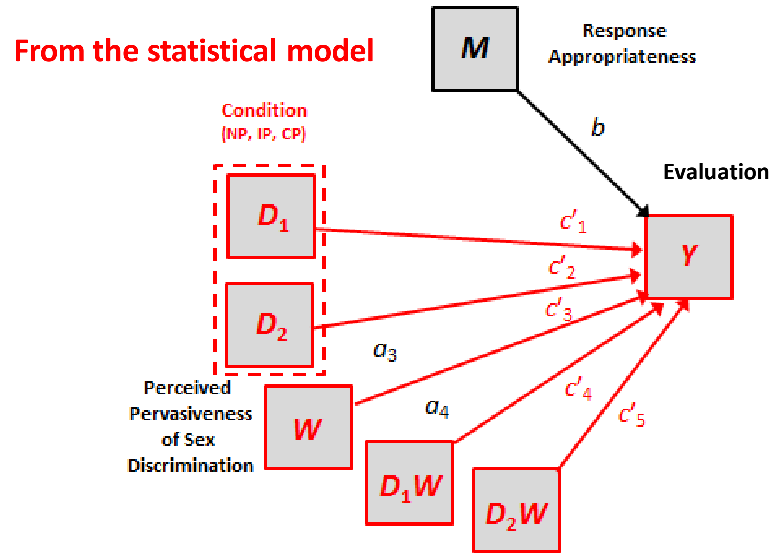
c'_1 → X1
c'_2 → X2
c'_3 → sexism
c'_4 → int_1
c'_5 → int_2

PROCESS creates the two products automatically

Test of interaction between X and W. This is a test of Moderation of the conditional direct effect of X.

The relative conditional direct effect





$$\hat{Y}_i = 3.87 + 0.03D_{1i} - 0.12D_{2i} - 0.278W'_i + 0.543D_{1i}W'_i + 0.509D_{2i}W'_i + 0.367M_i$$

which can be rewritten as

$$\hat{Y}_i = 3.87 + (0.03 + 0.543W'_i)D_{1i} + (-0.12 + 0.509W'_i)D_{2i} - 0.278W_i + 0.367M_i$$

or

$$\hat{Y}_i = 3.87 + \theta_{D_{1i}\to Y}D_{1i} + \theta_{D_{2i}\to Y}D_{2i} - 0.278W'_i + 0.367M_i$$

where

$$\theta_{D_{1i}\to Y} = 0.03 + 0.543W'_i$$

$$\theta_{D_{2i}\to Y} = -0.12 + 0.509W'_i$$

So the relative direct effects of individual and collective protest on evaluation of the attorney (relative to not protesting) are functions of perceived pervasiveness of sex discrimination.

PROCESS output

Use information provided by the PLOT option to generate a visual depiction of the model.

```
DATA LIST FREE/
   protest     sexism      eval         .
BEGIN DATA.
       .0000      4.2500      5.8991
      1.0000      4.2500      5.4610
      2.0000      4.2500      5.3418
       .0000      5.1200      5.6568
      1.0000      5.1200      5.6908
      2.0000      5.1200      5.5421
       .0000      5.8960      5.4407
      1.0000      5.8960      5.8957
      2.0000      5.8960      5.7207
END DATA.
GRAPH/SCATTERPLOT=
 sexism    WITH      eval      BY        protest   .
```

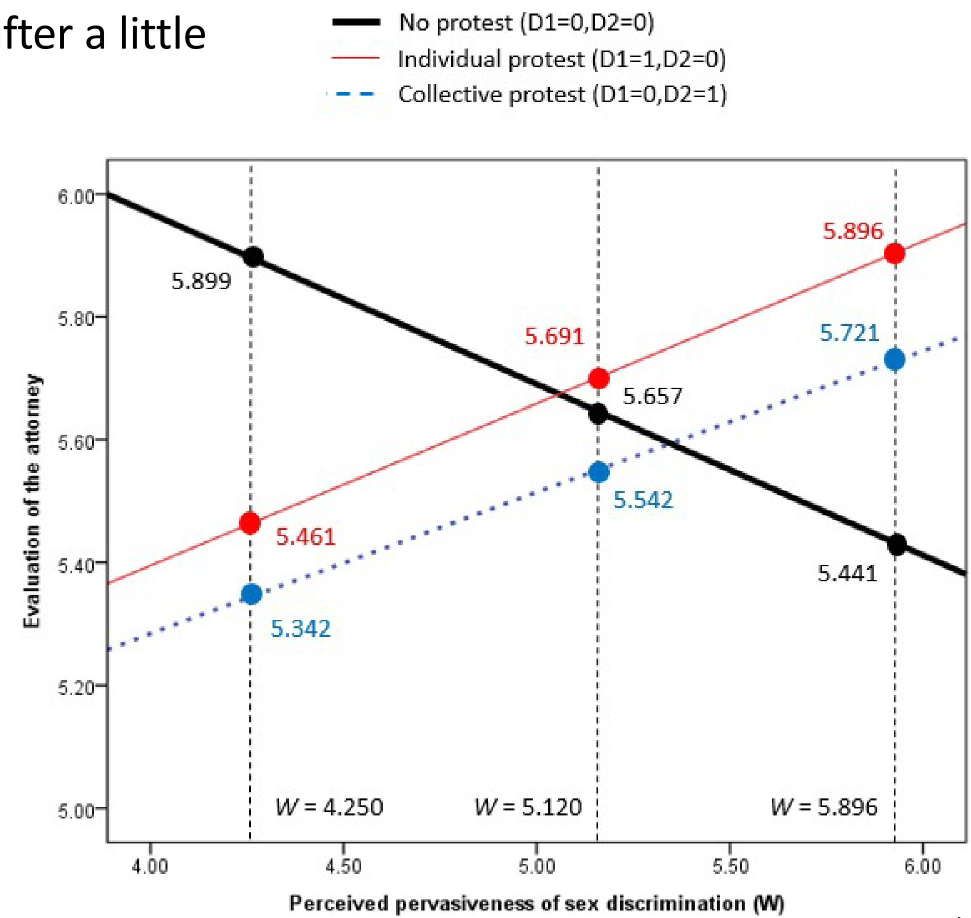
PROCESS for SPSS generates the data and the code to create the plot

```
data;
input protest sexism esteval;
datalines;
     .0000         4.2500     5.8991
    1.0000         4.2500     5.4610
    2.0000         4.2500     5.3418
     .0000         5.1200     5.6568
    1.0000         5.1200     5.6908
    2.0000         5.1200     5.5421
     .0000         5.8960     5.4407
    1.0000         5.8960     5.8957
    2.0000         5.8960     5.7207

run;
proc sgplot;reg x=sexism y=esteval/group=protest;run;
```

PROCESS for SAS only gives you this. The rest of the code you have to enter yourself.

Visualizing the moderation (from SPSS)

From SPSS, after a little editing:

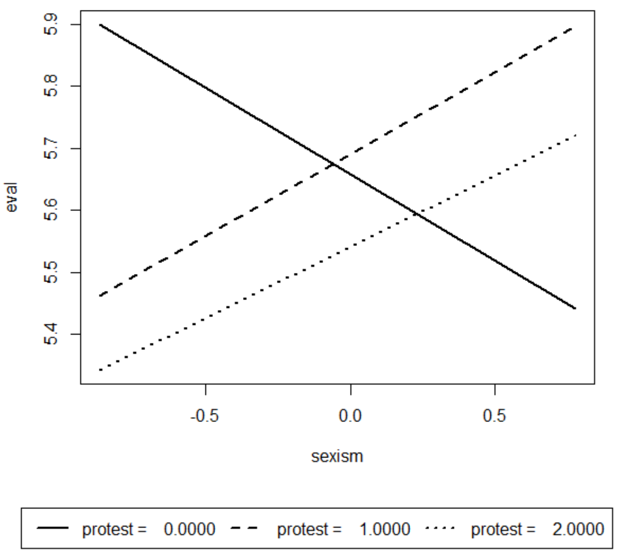


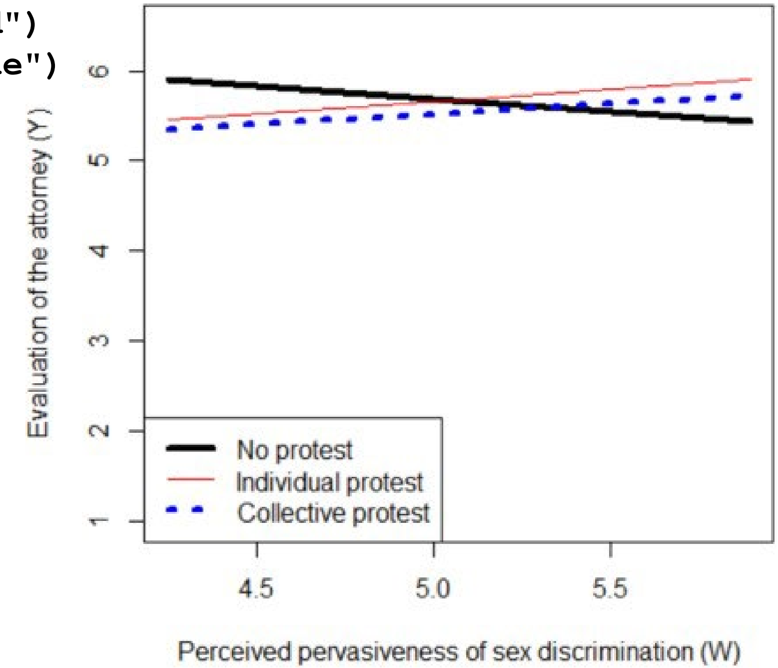
with *M* (response appropriateness) set to the sample mean

A visual representation of the model of *Y* (in R)

```
x<-c(0,1,2,0,1,2,0,1,2)
w<-c(4.25,4.25,4.25,5.12,5.12,5.12,5.896,5.896,5.896)
y<-c(5.899,5.461,5.342,5.657,5.691,5.542,5.441,5.896,5.721)
plot(y=y,x=w,pch=15,col="white",ylim=c(1,6.5),
xlab="Perceived pervasiveness of sex discrimination (W)",
ylab="Evaluation of the attorney (Y)")
legend.txt<-c("No protest","Individual protest","Collective protest")
legend("bottomleft",legend=legend.txt,lty=c(1,1,3),lwd=c(4,1,4),
col=c("black","red","blue"))
lines(w[x==0],y[x==0],lwd=4,lty=1,col="black")
lines(w[x==1],y[x==1],lwd=1,lty=1,col="red")
lines(w[x==2],y[x==2],lwd=4,lty=3,col="blue")
```

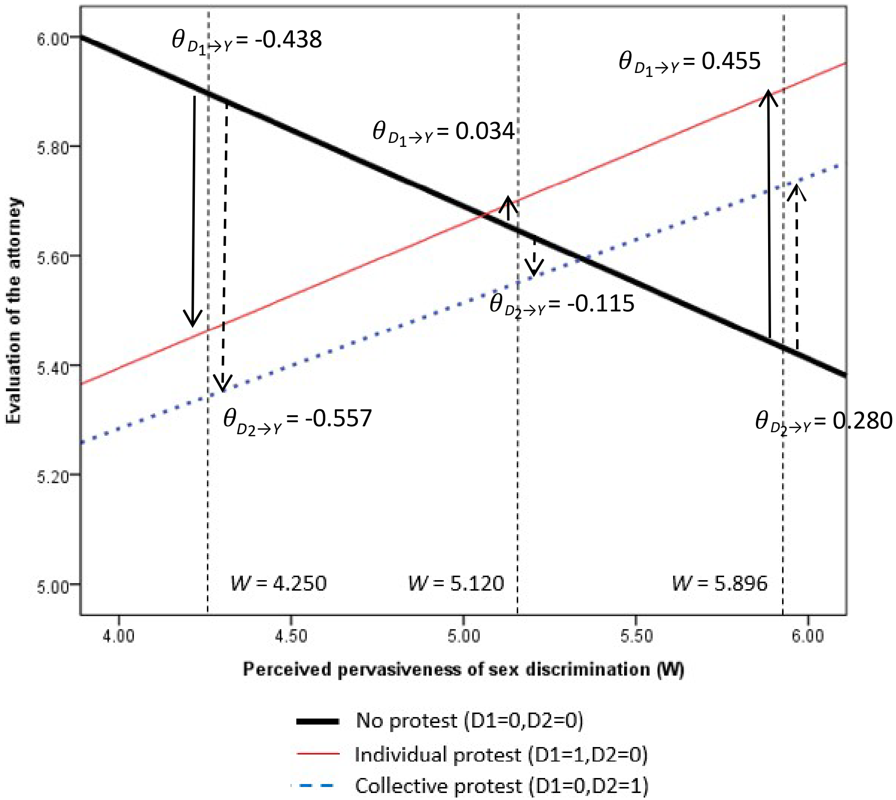
From PROCESS plot output





Conditional direct effects of protest form on evaluation of the attorney

$$\hat{Y}_i = 5.298 + (-2.744 + 0.543W_i)D_{1i} + (-2.719 + 0.509W_i)D_{2i} - 0.278W_i + 0.367M_i$$



$$\theta_{D_1\to Y} = -2.744 + 0.543W$$

$$\theta_{D_2\to Y} = -2.719 + 0.509W$$

SEXISM (W)	$\theta_{D_1\to Y}$	$\theta_{D_2\to Y}$
4.250	-0.438	-0.557
5.120	0.034	-0.115
5.896	0.455	0.280

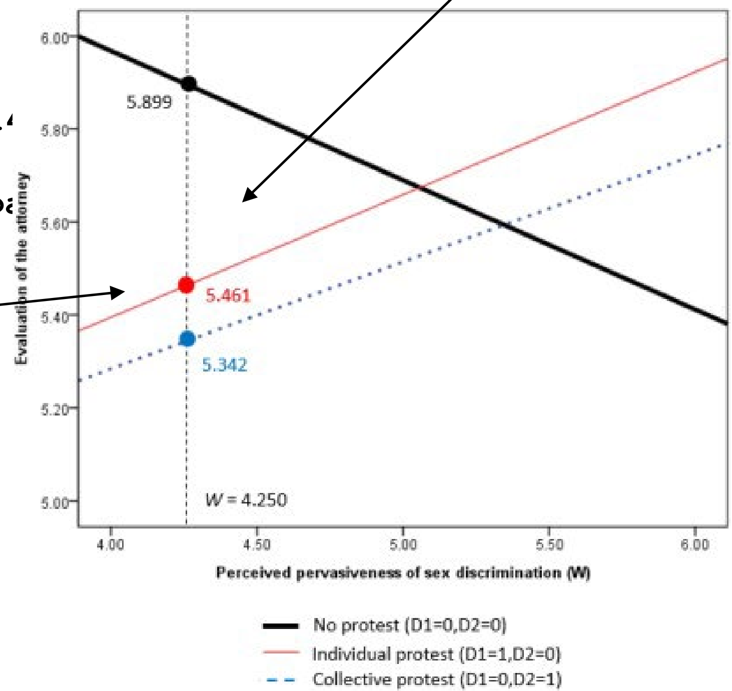
Holding perceived response appropriateness constant , it appears those who see sex discrimination as relatively less pervasive disliked the attorney more when she protested relative to when she did not, but the opposite is true for those who sex discrimination as relatively more pervasive. **But as will be seen, none of these relative direct effects are statistically different from zero.**

The relative conditional (direct) effect of *X* on *Y* when *W'* = -0.867, *W* = 4.250

```
Moderator value(s):

sexism    -0.8670

        effect         se          t          p       LLCI       ULCI
X1     -0.4381     0.3114    -1.4069     0.1620    -1.0545     0.1783
X2     -0.5573     0.2910    -1.9152     0.0578    -1.1332     0.0187

Test of equality of conditional means
          F        df1        df2
     1.9642     2.0000   122.0000     0.14

Estimated conditional means being compa
    protest       eval
     0.0000     5.8991
     1.0000     5.4610
     2.0000     5.3418
```

$\theta_{D_1 \rightarrow Y} = .03 + 0.534W = 0.03 + 0.534(-0.87) = -0.438$

$\theta_{D_2 \rightarrow Y} = -0.12 + 0.509W = -0.12 + 0.509(-0.87) = -0.557$

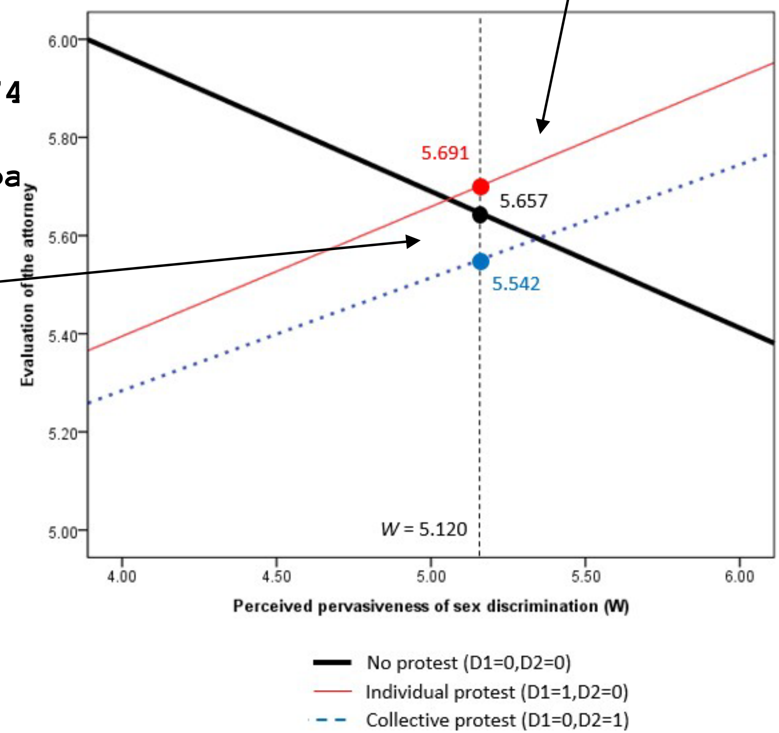


The relative conditional (direct) effect of *X* on *Y* when $W' = 0.003$, $W = 5.12$

```
Moderator value(s):
sexism     0.0030

        effect        se         t         p      LLCI      ULCI
X1      0.0340    0.2185    0.1554    0.8768   -0.3986    0.4665
X2     -0.1147    0.2299   -0.4991    0.6186   -0.5699    0.3404

Test of equality of conditional means
             F        df1        df2
        0.2942     2.0000   122.0000     0.74

Estimated conditional means being compa
       protest        eval
        0.0000      5.6568
        1.0000      5.6908
        2.0000      5.5421
```

$\theta_{D1\rightarrow Y} = .03 + 0.534W = 0.03 + 0.534(0.003) = 0.0340$

$\theta_{D2\rightarrow Y} = -0.12 + 0.509W = -0.12 + 0.509(0.003) = -0.1147$

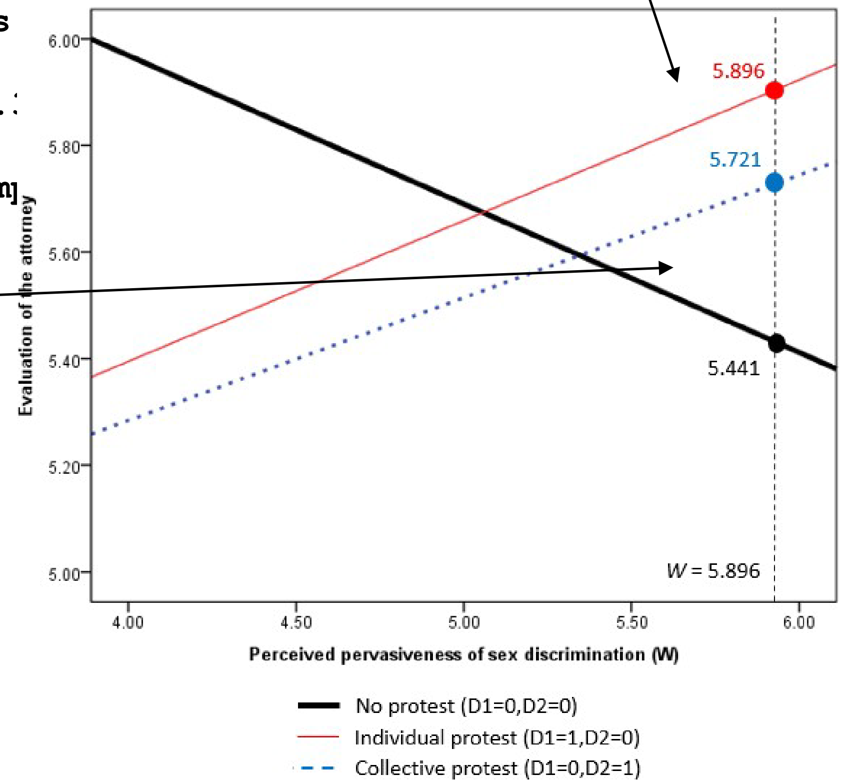


The relative conditional (direct) effect of X on Y when $W' = 0.7790$, $W = 5.896$

```
Moderator value(s):
sexism     0.7790

         effect        se         t         p      LLCI      ULCI
X1       0.4550    0.3129    1.4541    0.1485   -0.1645    1.0745
X2       0.2800    0.3293    0.8501    0.3970   -0.3720    0.9319

Test of equality of conditional means
          F        df1        df2
     1.0646     2.0000   122.0000     0.:

Estimated conditional means being com
    protest       eval
     0.0000     5.4407
     1.0000     5.8957
     2.0000     5.7207
```

$\theta_{D_1 \rightarrow Y} = .03 + 0.534W = 0.03 + 0.534(0.7790) = 0.4550$

$\theta_{D_2 \rightarrow Y} = -0.12 + 0.509W = -0.12 + 0.509(0.7790) = 0.2800$



PROCESS Output: Relative direct effects of *X* on *Y*

This information is also available in a summary section of the PROCESS output.

$\theta_{D_1 \rightarrow Y}$ = -2.744 + 0.534W

```
***************** DIRECT AND INDIRECT EFFECTS OF X ON Y ****************

Relative conditional direct effects of X on Y:
       sexism    effect        se         t         p      LLCI      ULCI
X1    -0.8670   -0.4381    0.3114   -1.4069    0.1620   -1.0545    0.1783
X1     0.0030    0.0340    0.2185    0.1554    0.8768   -0.3986    0.4665
X1     0.7790    0.4550    0.3129    1.4541    0.1485   -0.1645    1.0745
X2    -0.8670   -0.5573    0.2910   -1.9152    0.0578   -1.1332    0.0187
X2     0.0030   -0.1147    0.2299   -0.4991    0.6186   -0.5699    0.3404
X2     0.7790    0.2800    0.3293    0.8501    0.3970   -0.3720    0.9319
```

$\theta_{D_2 \rightarrow Y}$ = -2.719 + 0.509W

Putting it all together

SEXISM	Relative indirect effects (with 95% CI) Individual protest	 Collective protest
16^{th} %	0.140 (-0.188 to 0.450)	0.372 (0.102 to 0.685)*
50^{th} %	0.451 (0.214 to 0.728)*	0.606 (0.322 to 0.919)*
84^{th} %	0.730 (0.348 to 1.149)*	0.815 (0.407 to 1.283)*
	Relative direct effects (with 95% CI)	
16^{th} %	-0.438 (-1.055 to 0.178)	-0.557 (-1.113 to 0.019)
50^{th} %	0.034 (-0.399 to 0.467)	-0.115 (-0.567 to 0.340)
84^{th} %	0.455 (-0.165 to 1.075)	0.280 (-0.372 to 0.932)

Confidence intervals for indirect effects are percentile bootstrap intervals based on 5,000 bootstrap samples.
Confidence intervals for direct effects are OLS intervals.
* Statistically different from zero.

Protesting (regardless of form) indirectly influenced evaluations of the attorney through perceived response appropriateness, with both forms being perceived as a more appropriate response which in turn enhanced likeability. The exception was individual protest among people relatively low in perceived pervasiveness of sex discrimination, where the relative indirect effect was not definitively different from zero.

Models with Multiple Moderators

Stereotype Threat and the "Fragile" Self-Concept

Gerstenberg, F. X. R., Imhoff, R., & Schmitt, M. (2012). "Women are bad at math, but I am not, am I?": Fragile mathematical self-concept predicts vulnerability to a stereotype threat effect on mathematical performance*. European Journal of Personality, 26*, 588-599.

Stereotype Threat Theory: When others hold a belief that your group is weak in a particular area, performance in that area can be undermined as a result of fear and concern about confirming a negative stereotype about your group. Experimental research supports this. When a stereotype is made salient about a group with which people identify, their performance is hindered on stereotype-related tasks that are difficult.

Method

- 136 women were given 20 moderately difficult math problems to complete from various standardized tests (e.g., GMAT).

- Just before completing the math test, half of the participants had their sex made salient merely by having them report whether they were male or female. This was their manipulation of "stereotype threat" (i.e., those whose sex was made salient are in the the stereotype threat condition).

- Prior to the test, an explicit and an implicit measure of "mathematical self-concept" was administered.

Stereotype Threat and the "Fragile" Self-Concept

Gerstenberg, F. X. R., Imhoff, R., & Schmitt, M. (2012). "Women are bad at math, but I am not, am I?": Fragile mathematical self-concept predicts vulnerability to a stereotype threat effect on mathematical performance. *European Journal of Personality, 26*, 588-599.

Method

- **Explicit mathematical self-concept**. Average response to three questions to which the women expressed their level of identification with the labels "mathematical", "artistic" and "more mathematical than artistic". Higher scores represent a stronger concept of one's self as "mathematical" rather than "artistic".

- **Implicit mathematical self-concept**. Obtained from an implicit associations test (IAT). An IAT gauges one's self-concept or attitude by relative speed of "me" and "not me" or "like" and "don't like" when presented with words, photographs, and so forth that represent various concepts (in this case, mathematical or artistic). Higher scores represent a stronger concept of one's self as "mathematical" relative to "artistic".

Implicit measures are thought to be more valid measures of attitudes and self-concepts because people often are unwilling to admit to things that are socially undesirable (or they may not have accurate insights into their "real" thoughts and feelings). This is controversial to some extent. **Learn about the IAT at https://implicit.harvard.edu/implicit/**

The data: math

math.sav

*math.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs U

4 :

	threat	implms	explms	mathprob
1	0	-.56	15	8
2	1	-1.49	13	6
3	0	-.36	14	7
4	1	-.71	12	8
5	0	-.71	17	9
6	1	-.77	13	8
7	0	-.23	13	5
8	1	-.44	19	11
9	0	-.35	17	10
10	1	-.59	15	8

math.sas

math

```
data math;
  input threat implms explms mathprob;
  datalines;
  0   -.56    15  8
  1   -1.49   13  6
  0   -.36    14  7
  1   -.71    12  8
  0   -.71    17  9
  1   -.77    13  8
  0   -.23    13  5
  1   -.44    19  11
  0   -.35    17  10
  1   -.59    15  8
  0   -.63    15  8
  1   -.43    17  10
  0   -.77    15  7
  1   -.68    16  6
```

In R: Don't forget to change the path below to where your **math.csv** file is located.

```
> math<-read.table("c:/mmcpa/math.csv", sep=",",header=TRUE)
> head(math)
  threat implms explms mathprob
1      0  -0.56     15        8
2      1  -1.49     13        6
3      0  -0.36     14        7
4      1  -0.71     12        8
5      0  -0.71     17        9
6      1  -0.77     13        8
```

THREAT: 0 = no threat condition, 1 = threat condition

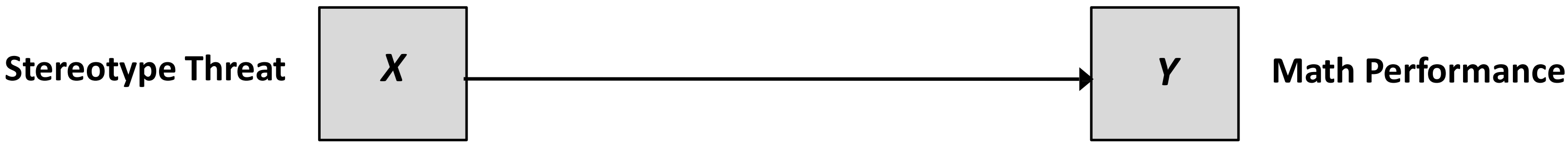
MATHPROB: Performance (math problems correctly solved)

IMPLMS: Implicit mathematical self concept

EXPLMS: Explicit mathematical self-concept

A stereotype threat effect?

Did women whose identity as female was made salient (stereotype threat condition) perform worse on average than those who's identity was not made salient?




```
regression/dep=mathprob/method=enter threat.
```

```
proc reg data=math;model mathprob=threat;run;
```

```
summary(lm(mathprob~threat,data=math))
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.2794     0.3758  22.034   <2e-16 ***
threat       -0.3235     0.5314  -0.609    0.544
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.099 on 134 degrees of freedom
Multiple R-squared:  0.002759,  Adjusted R-squared:  -0.004684
F-statistic: 0.3707 on 1 and 134 DF,  p-value: 0.5437
```

Women whose identity as female was primed answered slightly fewer questions correctly on average (7.955) than those whose identity was not primed (8.279), but this difference of 0.324 questions was not statistically significant, $t(134) = -0.609$, $p = 0.544$.

An "ANCOVA" (analysis of covariance) model

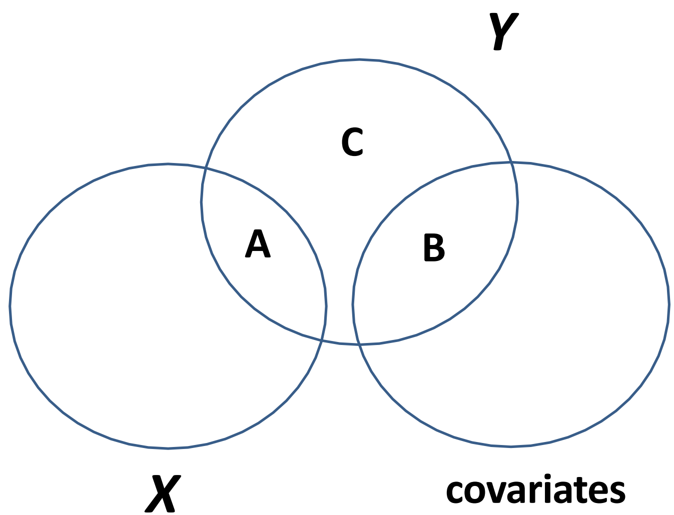
In experiments, power of tests of an experimental manipulation *X* can be increased by including "covariates" that are correlated with *Y* but not with *X*. Random assignment to values of *X* all but guarantees *X* is uncorrelated with everything measured prior to assignment.

```
correlations variables = threat explms implms mathprob.
```

```
proc corr data=math;var threat explms implms mathprob;run;
```

```
cor(math)
```

	threat	implms	explms	mathprob
threat	1.00000000	-0.09972272	0.09196297	-0.05252168
implms	-0.09972272	1.00000000	0.19780049	0.42972932
explms	0.09196297	0.19780049	1.00000000	0.45625390
mathprob	-0.05252168	0.42972932	0.45625390	1.00000000

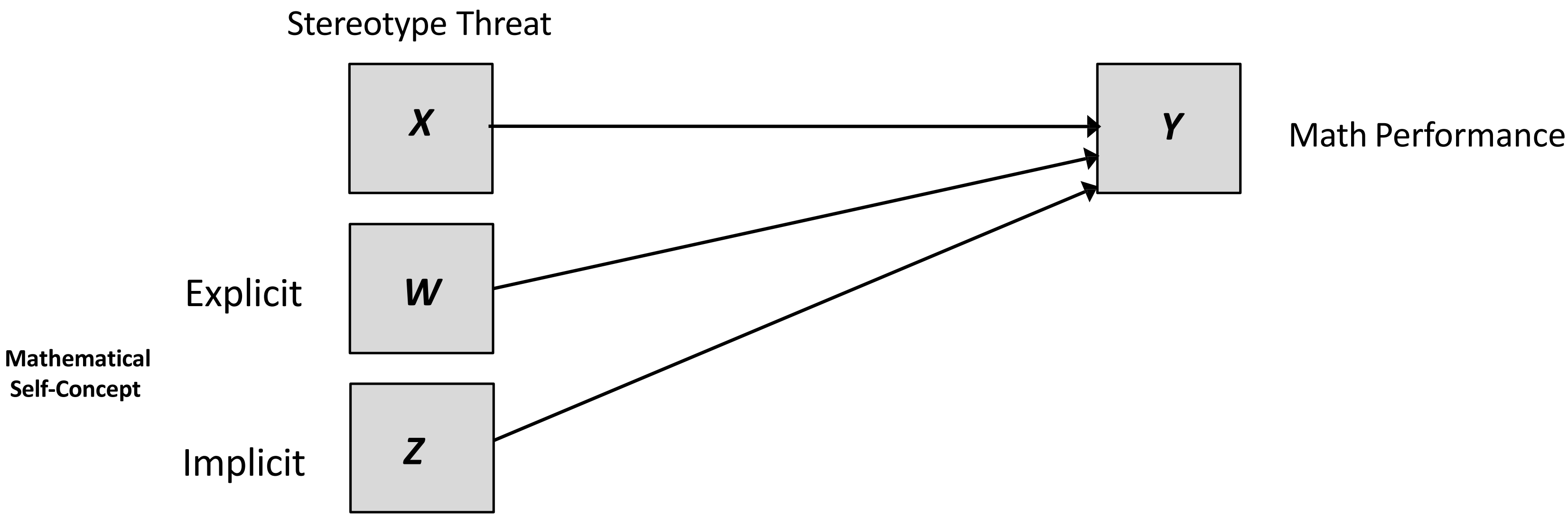


Observe that experimental condition is largely uncorrelated with math self-concept, but math self-concept is correlated with performance.

The experimental effect is "signal" relative to "noise." In the prior analysis, that ratio is A/(B+C). In ANCOVA, that ratio is A/C, which is typically larger. So power to detect *X*'s effect is increased by accounting for things correlated with *Y* but not *X*.

An "ANCOVA" (analysis of covariance) model

In regression terms, we can conduct analysis of covariance by including covariates in the model of *Y* along with the variable coding experimental condition.



In this model, the effect of the experimental manipulation of stereotype threat is assessed, "holding mathematical self-concept constant" (measured in two different ways). Variation in performance due to variance in self-concept is eliminated from the "noise" that is error variance in *Y*. In general, this increases the power of the test of the manipulation's effect when *X* is uncorrelated with the covariates.

An ANCOVA model

```
regression/dep = mathprob/method = enter threat explms implms.
```

```
proc reg data=math;model mathprob=threat explms implms;run;
```

```
summary(lm(mathprob~threat+explms+implms,data=math))
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.7676     1.7775   0.432    0.667
threat       -0.3330     0.4437  -0.751    0.454
explms        0.5817     0.1083   5.371 3.43e-07 ***
implms        2.5951     0.5477   4.738 5.49e-06 ***
```

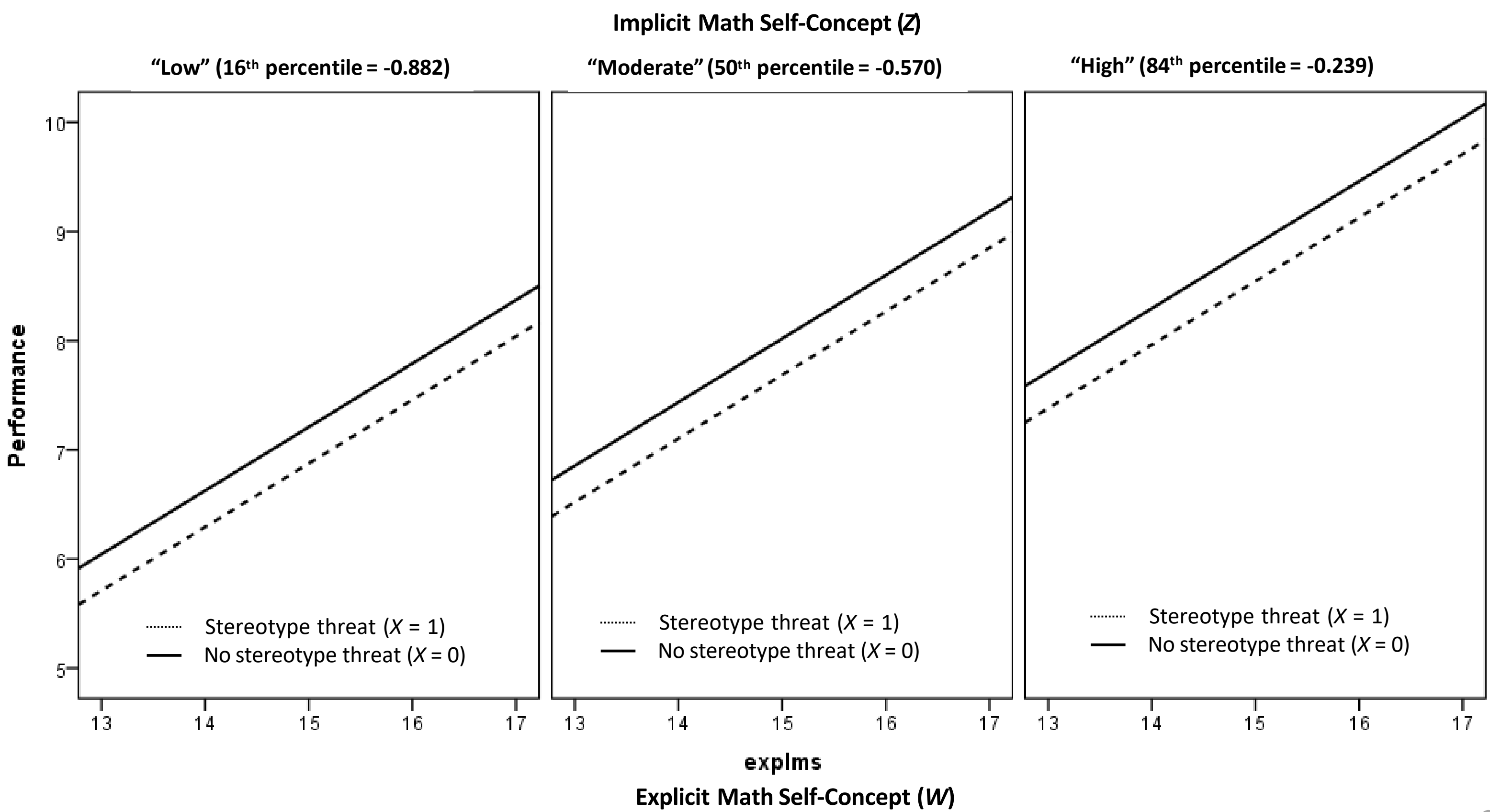
This doesn't seem to have helped in this case. Although the estimate is more precise (as expected; se = 0.444 vs. 0.531), the difference between conditions is still not statistically significant.

The stereotype threat manipulation appears to have had no effect on performance even after eliminating variation in performance due to math self-concept. Each measure of math self-concept explains some of the variation in performance and in the direction you would expect (stronger math self-concept = better performance). But stereotype threat still does not (although the effect is in the expected direction).

This is an "unconditional" model

This model forces stereotype threat's effect on performance to be invariant across individual differences in mathematical self-concept. Proper use and interpretation of ANCOVA not only forces this, but ASSUMES this.

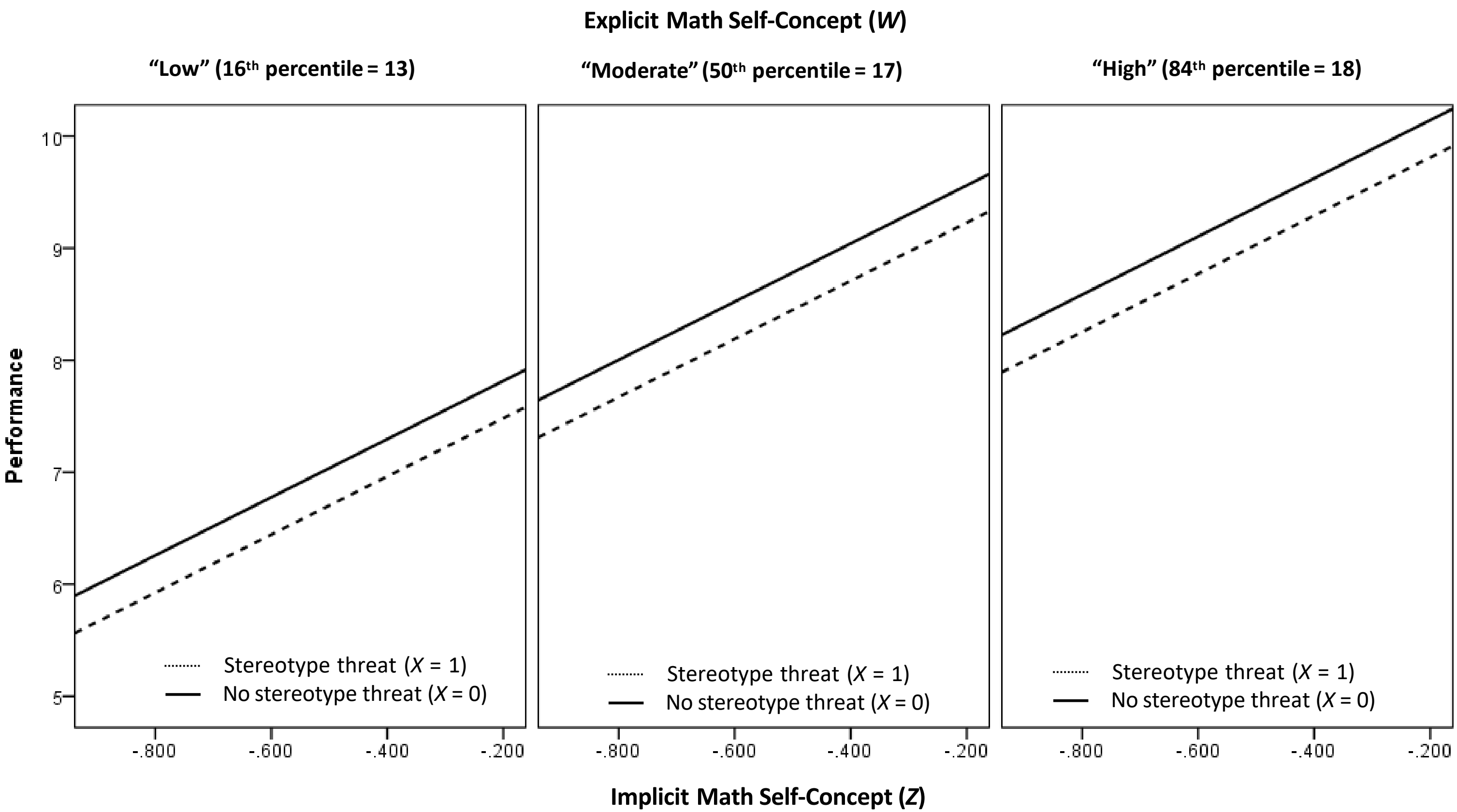
$$\hat{Y}_i = 0.768 - 0.333X_i + 0.582W_i + 2.595Z_i$$



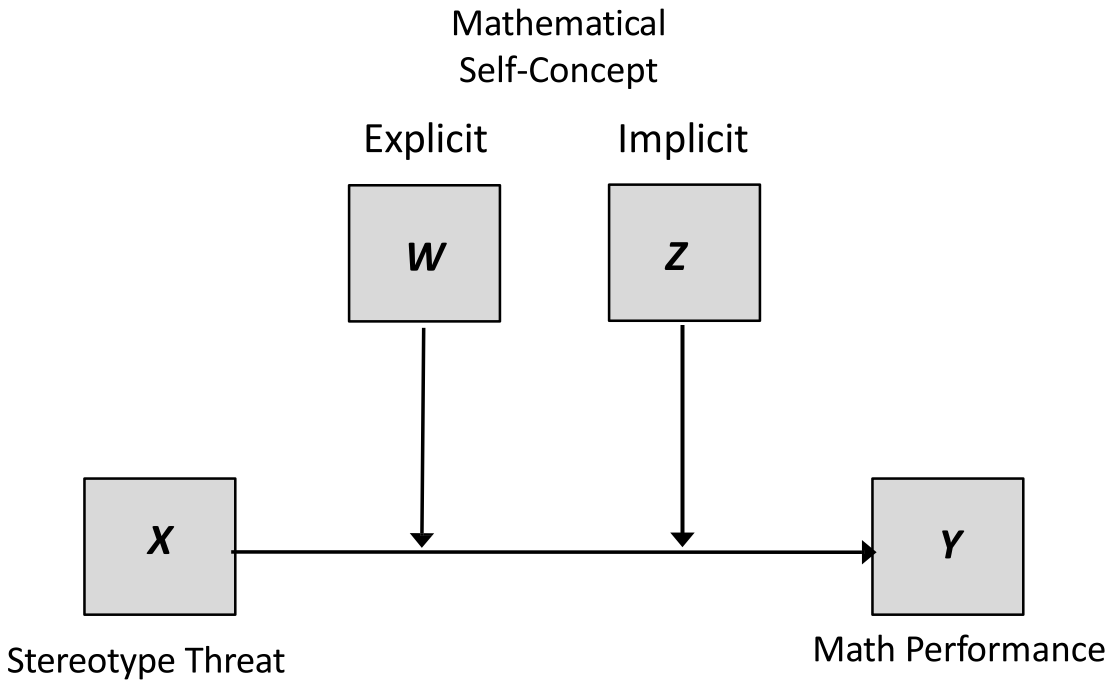
This is an "unconditional" model

This model forces stereotype threat's effect on performance to be invariant across individual differences in mathematical self-concept. Proper use and interpretation of ANCOVA not only forces this, but ASSUMES this.

$$\hat{Y}_i = 0.768 - 0.333X_i + 0.582W_i + 2.595Z_i$$



A conditional model



There is reason to believe (from existing theory and research) that threat of stereotype confirmation would be especially strong among those whose identities are consistent with the stereotype view (i.e., those who see themselves as mathematically weak). Those whose identities are inconsistent with the stereotype (those with a stronger mathematical self-concepts) have less reason to be concerned about their performance confirming the stereotype, so performance would be less hindered, if at all, by stereotype threat.

Relaxing this constraint

Suppose we let *X*'s effect be a function of both *W* and *Z*, *f*(*W*,*Z*), as in

$$\widehat{Y}_i = b_0 + f(W_i, Z_i)X_i + b_2W_i + b_3Z_i$$

For instance, let *f*(*W*,*Z*) be an additive linear function $b_1 + b_4W + b_5Z$. Thus,

$$\widehat{Y}_i = b_0 + (b_1 + b_4W_i + b_5Z_i)X_i + b_2W_i + b_3Z_i$$

This can be rewritten in an equivalent form as

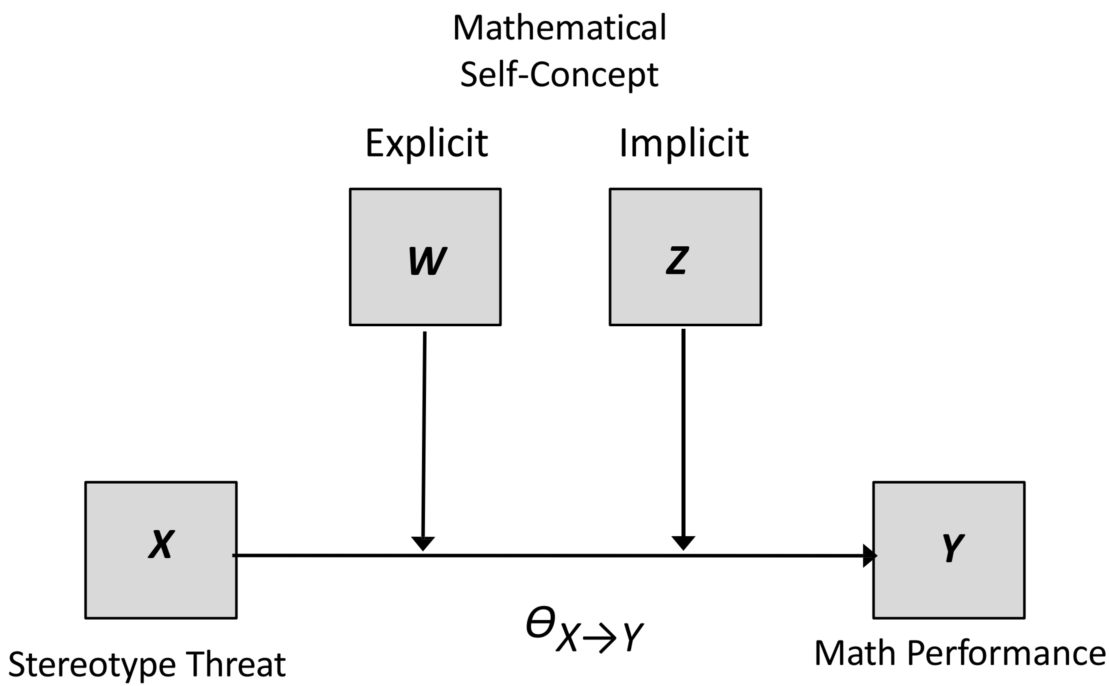
$$\widehat{Y}_i = b_0 + \theta_{X\rightarrow Y}X_i + b_2W_i + b_3Z_i \quad \text{where} \quad \theta_{X\rightarrow Y} = b_1 + b_4W_i + b_5Z_i$$

or, equivalently,

$$\widehat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3Z_i + b_4W_iX_i + b_5Z_iX_i$$

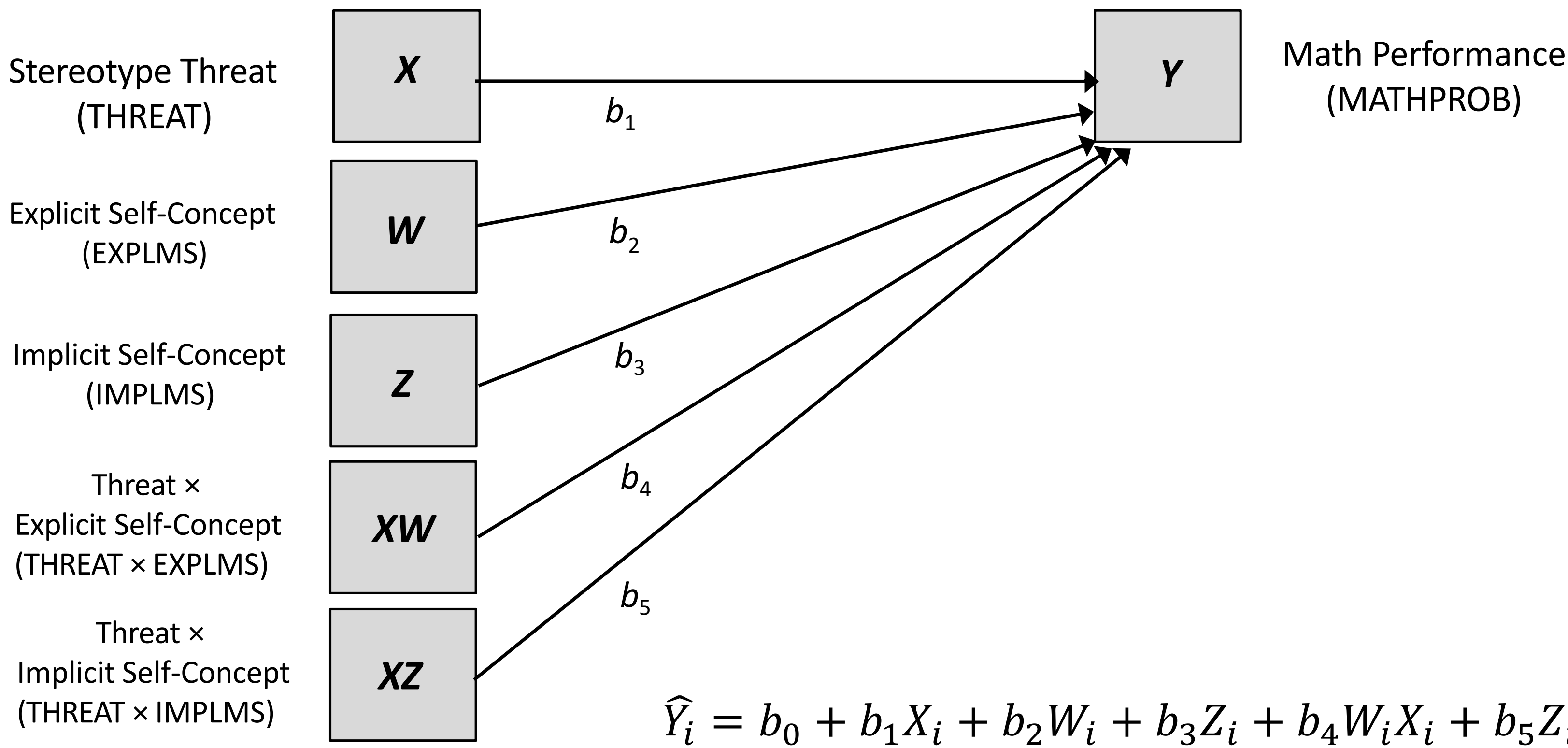
This model, the "additive multiple moderation model," allows *X*'s effect on *Y* to depend linearly and additively on *W* and *Z*. *X*'s effect is conditioned on both *W* and *Z*. *X*'s conditional effect on *Y*, $\Theta_{X\rightarrow Y}$, is $b_1 + b_4W + b_5Z$. *X*'s effect on *Y* depends on what you put in for *W* and *Z*.

A conditional model in conceptual form



Stereotype threat's (*X*) conditional effect on math performance (*Y*), $\theta_{X \rightarrow Y}$, depends on mathematical self-concept, measured explicitly (*W*) and implicitly (*Z*). Implicit and explicit math self-concepts are only weakly correlated (r = 0.20), frustrating a desire to simplify the analysis by aggregating them.

In path diagram form



$$\widehat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3Z_i + b_4W_iX_i + b_5Z_iX_i$$

or

$$\widehat{Y}_i = b_0 + \underline{(b_1 + b_4W_i + b_5Z_i)}X_i + b_2W_i + b_3Z_i$$

X's effect, $\theta_{X \to Y}$, is a linear additive function of *W* and *Z*

Estimating in SPSS, SAS, or R (the harder way)

```
compute thrtexpl = threat*explms.
compute thrtimpl = threat*implms.
regression/dep = mathprob/method = enter threat explms implms thrtexpl thrtimpl.
```

```
data math;set math;thrtexpl=threat*explms;thrtimpl=threat*implms;run;
proc reg data=math;model mathprob=threat explms implms thrtexpl thrtimpl;run;
```

```
summary(lm(mathprob~threat+explms+implms+threat*explms+threat*implms,data=math))
```

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     -2.1587     2.6965  -0.801   0.4249
threat           5.1751     3.6004   1.437   0.1530
explms           0.7344     0.1679   4.375 2.47e-05 ***
implms           1.5307     0.7908   1.936   0.0551 .
threat:explms   -0.2774     0.2186  -1.269   0.2068
threat:implms    2.0481     1.0881   1.882   0.0620 .
```

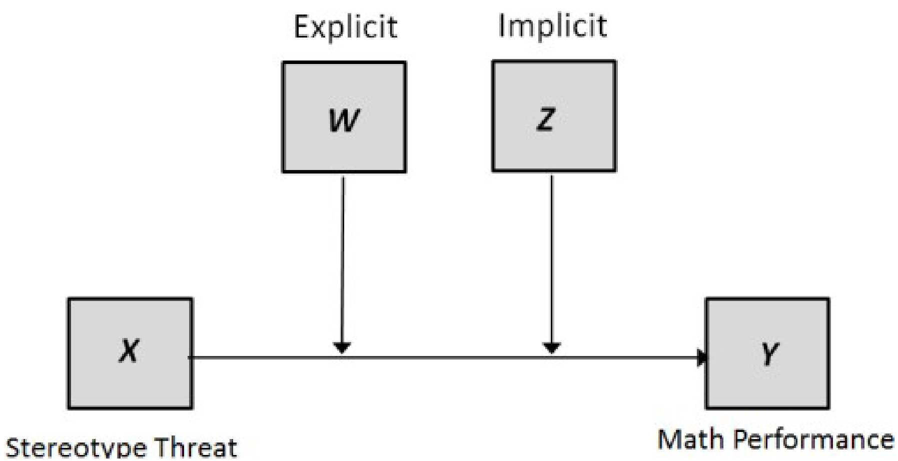
No significant evidence of moderation of the effect of threat by *implicit* math self-concept or *explicit* math self-concept. We'll return to this later.

$$\hat{Y}_i = b_0 + b_1 X_i + b_2 W_i + b_3 Z_i + b_4 W_i X_i + b_5 Z_i X_i$$

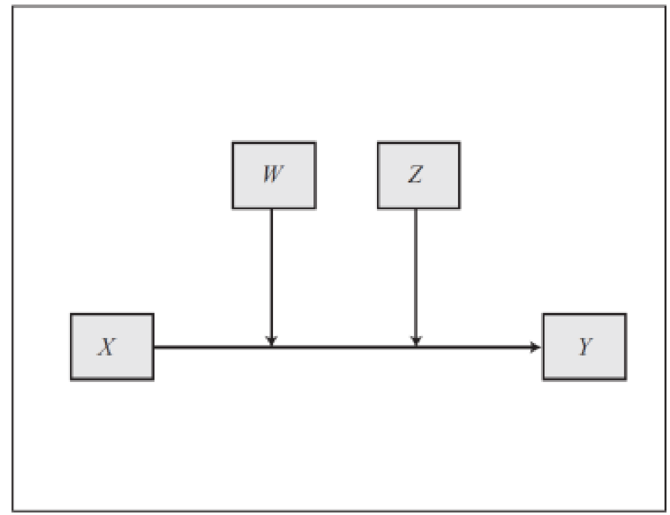
$$\hat{Y}_i = -2.159 + 5.175 X_i + 0.734 W_i + 1.531 Z_i - 0.277 W_i X_i + 2.048 Z_i X_i$$

Let PROCESS help you estimate, test, and visualize

We can use PROCESS model 2 to estimate this model, and do some visualization and a few other things.



Model 2




```
process y=mathprob/x=threat/w=explms/z=implms/model=2/plot=1.
```

```
%process (data=math,y=mathprob,x=threat,w=explms,z=implms,model=2,plot=1)
```

```
process(data=math,y="mathprob",x="threat",w="explms",z="implms",model=2,
   plot=1)
```

From PROCESS (Model 2)

```
Model  : 2
    Y  : mathprob
    X   :  threat
    W   :  explms
    Z  : implms

Sample
Size:  136
```

$$\widehat{Y}_i = -2.159 + 5.175X_i + 0.734W_i + 1.531Z_i - 0.277W_iX_i + 2.048Z_iX_i$$

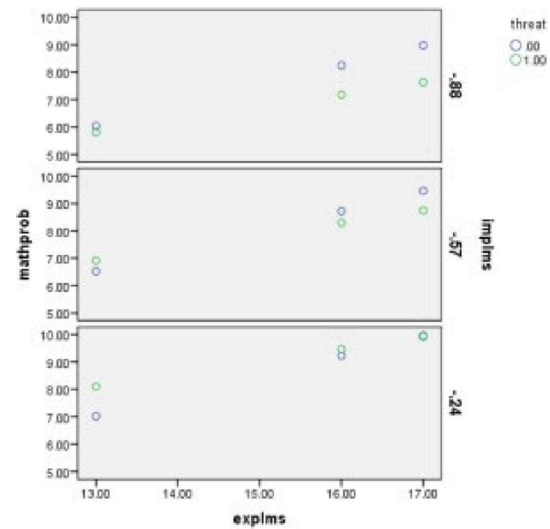
```
**************************************************************************
OUTCOME VARIABLE:
 mathprob

Model Summary
      R    R-sq   MSE                          F        df1        df2          p
   .5940  .3528 6.4229                   14.1722     5.0000   130.0000      .0000

Model
           coeff             se          t           p      LLCI        ULCI
Constant  -2.1587         2.6965    -.8005       .4249   -7.4934      3.1761
threat     5.1751         3.6004    1.4374       .1530   -1.9478     12.2981
explms      .7344          .1679    4.3747       .0000     .4023      1.0665
Int_1     -.2774          .2186   -1.2689       .2068    -.7100       .1551
implms     1.5307          .7908    1.9356       .0551    -.0338      3.0953
Int_2      2.0481         1.0881    1.8824       .0620    -.1045      4.2007

Product terms key:
 Int_1:    threat  x                explms
 Int_2:    threat  x                implms

Test(s) of highest order unconditional interaction(s):
    R2-chng              F     df1    df2            p
X*W   .0080          1.6100   1.0000  130.0000      .2068
X*Z   .0176          3.5433   1.0000  130.0000      .0620
BOTH  .0218          2.1916   2.0000  130.0000      .1159
```

A visual representation of the effect of *X* on *Y*

The **plot** option in SPSS generates code to visualize. Cut and paste this into syntax and run.

$$\widehat{Y_i} = -2.159 + 5.175X_i + 0.734W_i + 1.531Z_i - 0.277W_iX_i + 2.048Z_iX_i$$

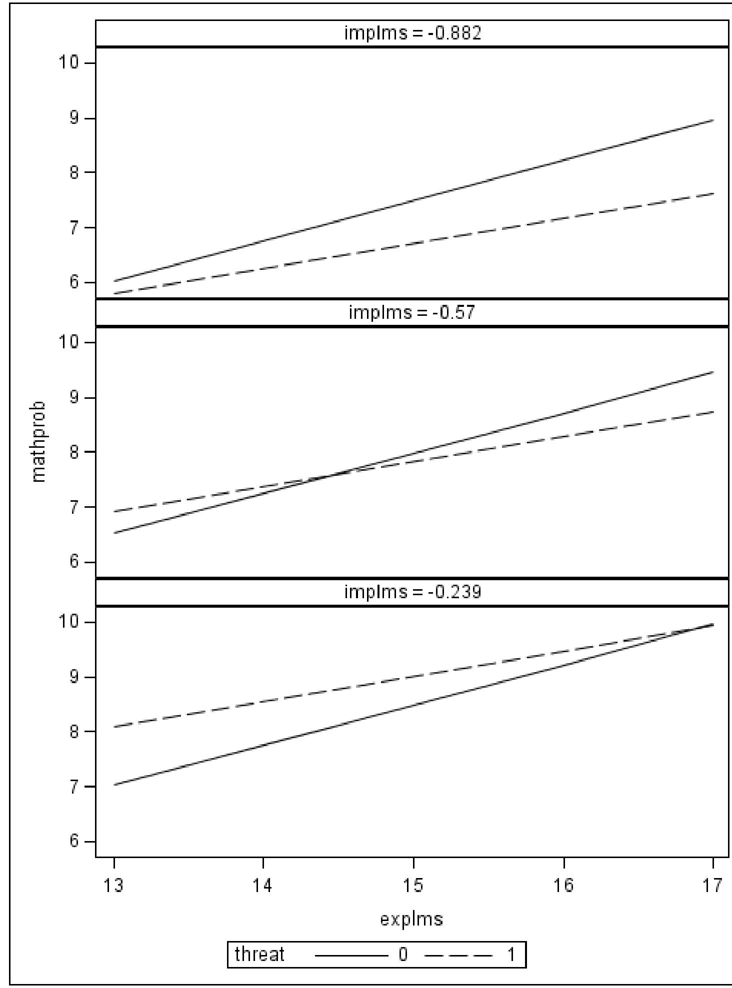
```
DATA LIST FREE/
   threat     explms      implms        mathprob  .
 BEGIN DATA.
        .0000      13.0000       -.8816       6.0393
       1.0000      13.0000       -.8816       5.8023
        .0000      13.0000       -.5700       6.5163
       1.0000      13.0000       -.5700       6.9174
        .0000      13.0000       -.2392       7.0226
       1.0000      13.0000       -.2392       8.1013
        .0000      16.0000       -.8816       8.2425
       1.0000      16.0000       -.8816       7.1733
        .0000      16.0000       -.5700       8.7195
       1.0000      16.0000       -.5700       8.2884
        .0000      16.0000       -.2392       9.2259
       1.0000      16.0000       -.2392       9.4723
        .0000      17.0000       -.8816       8.9769
       1.0000      17.0000       -.8816       7.6302
        .0000      17.0000       -.5700       9.4539
       1.0000      17.0000       -.5700       8.7454
        .0000      17.0000       -.2392       9.9603
       1.0000      17.0000       -.2392       9.9293

END DATA.
GRAPH/SCATTERPLOT=
 explms   WITH      mathprob BY        threat    /PANEL    ROWVAR=   implms
```



This can be edited to make it look prettier.

In SAS

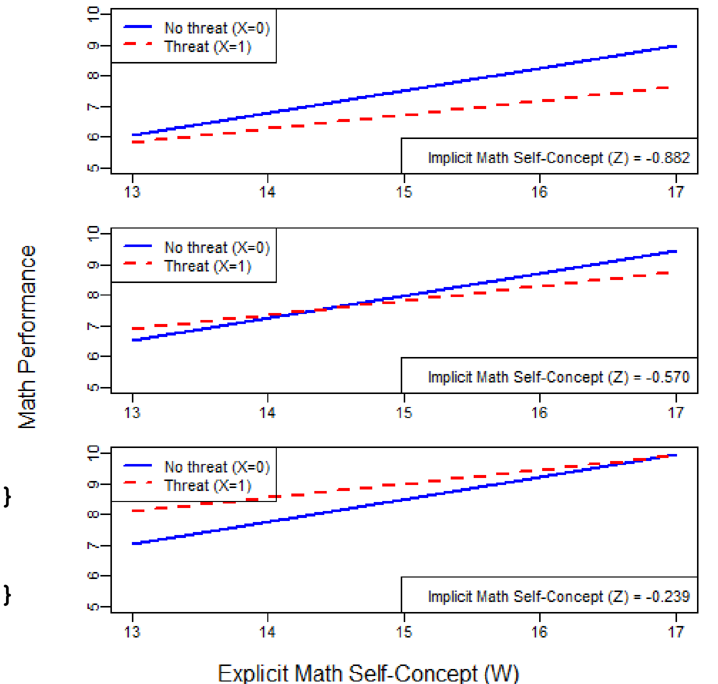
```
data;input threat explms implms mathprob;
datalines;
     .0000    13.0000     -.8816       6.0393
    1.0000    13.0000     -.8816       5.8023
     .0000    13.0000     -.5700       6.5163
    1.0000    13.0000     -.5700       6.9174
     .0000    13.0000     -.2392       7.0226
    1.0000    13.0000     -.2392       8.1013
     .0000    16.0000     -.8816       8.2425
    1.0000    16.0000     -.8816       7.1733
     .0000    16.0000     -.5700       8.7195
    1.0000    16.0000     -.5700       8.2884
     .0000    16.0000     -.2392       9.2259
    1.0000    16.0000     -.2392       9.4723
     .0000    17.0000     -.8816       8.9769
    1.0000    17.0000     -.8816       7.6302
     .0000    17.0000     -.5700       9.4539
    1.0000    17.0000     -.5700       8.7454
     .0000    17.0000     -.2392       9.9603
    1.0000    17.0000     -.2392       9.9293

run;
proc sgpanel;
panelby implms / columns=1;
series x=explms y=mathprob/group=threat
lineattrs =(color=black);run;
```



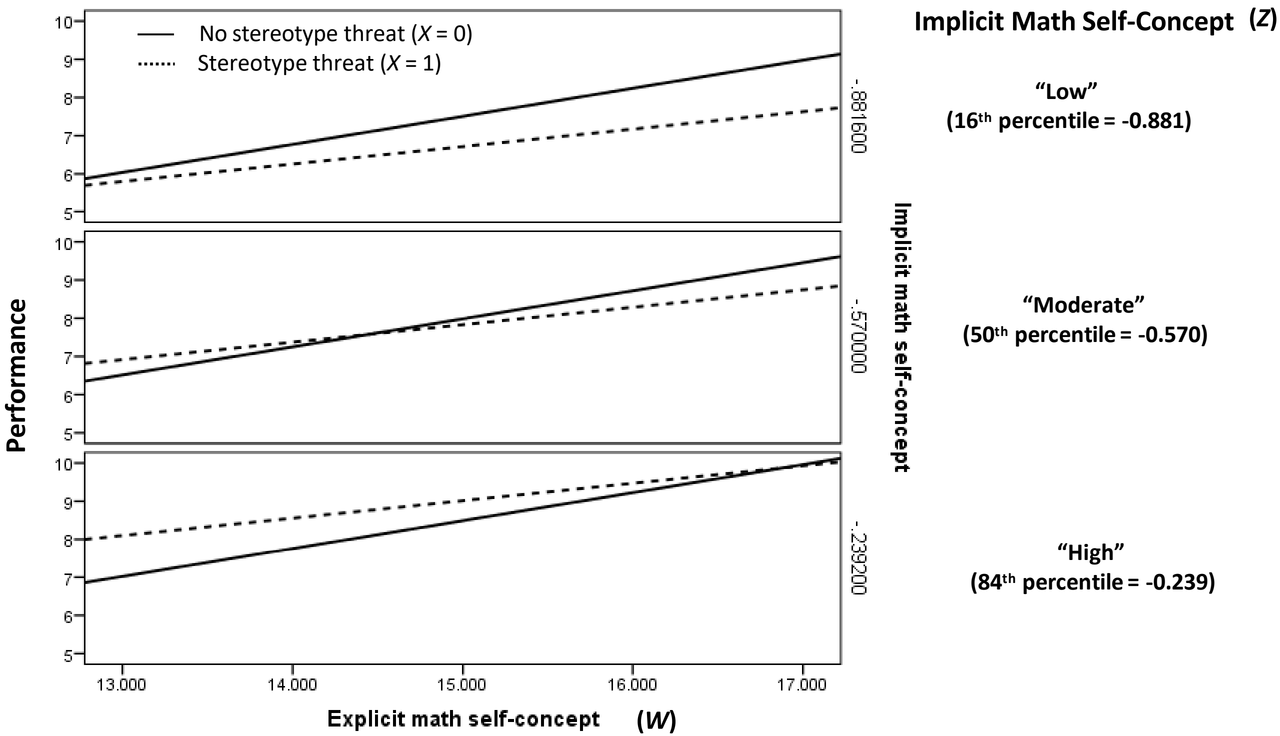
In R

```
oldp<-par(mfrow=c(3,1),mar=c(3,4,0,0),
oma=c(2,2,2,2),mgp=c(5,0.5,0))
w<-c(13,16,17,13,16,17)
x<-c(0,0,0,1,1,1)
yzlow<-c(6.039,8.242,8.977,5.802,7.173,7.630)
yzmod<-c(6.516,8.719,9.454,6.917,8.288,8.745)
yzhigh<-c(7.023,9.226,9.960,8.101,9.472,9.929)
wt<-x
x<-w
w<-wt
legend.txt<-c("No threat (X=0)","Threat (X=1)")
for (i in 1:3){
if (i==1)
  {y<-yzlow
  legend2.txt<-c("Implicit Math Self-Concept (Z) = -0.882")  }
if (i==2)
  {y<-yzmod
  legend2.txt<-c("Implicit Math Self-Concept (Z) = -0.570")  }
if (i==3)
  {y<-yzhigh
  legend2.txt<-c("Implicit Math Self-Concept (Z) = -0.239")  }
plot(y=y,x=x,col="white",ylim=c(5,10),cex=1.5,xlim=c(13,17),tcl=-0.5)
lines(x[w==0],y[w==0],lwd=2,lty=1,col="blue")
lines(x[w==1],y[w==1],lwd=2,lty=2,col="red")
legend("topleft", legend=legend.txt,lwd=2,lty=c(1,2),
col=c("blue","red"))
legend("bottomright",legend=legend2.txt)
box}
mtext("Explicit Math Self-Concept (W)",side=1,outer=TRUE)
mtext("Math Performance",side=2,outer=TRUE)
par(oldp)
```

From PROCESS **plot** output



A visual representation of the effect of X on Y



$$\hat{Y}_i = -2.159 + 5.175X_i + 0.734W_i + 1.531Z_i - 0.277W_iX_i + 2.048Z_iX_i$$

$$\hat{Y}_i = -2.159 + (5.175 - 0.277W_i + 2.048Z_i)X_i + 0.734W_i + 1.531Z_i$$

$\hat{Y}_i = -2.159 + \theta_{X\rightarrow Y}X_i + 0.734W_i + 1.531Z_i$ where $\theta_{X\rightarrow Y} = 5.175 - 0.277W_i + 2.048Z_i$

$\theta_{X\rightarrow Y}$ is the distance between the solid and the dotted line at values of W and Z.

Interpretation of regression coefficients determining *X*'s effect on *Y*

$$\hat{Y}_i = b_0 + b_1 X_i + b_2 W_i + b_3 Z_i + b_4 W_i X_i + b_5 Z_i X_i$$

$$\hat{Y}_i = b_0 + (b_1 + b_4 W_i + b_5 Z_i) X_i + b_2 W_i + b_3 Z_i$$

$$\hat{Y}_i = b_0 + \theta_{X \to Y} X_i + b_2 W_i + b_3 Z_i \qquad \theta_{X \to Y} = b_1 + b_4 W_i + b_5 Z_i$$

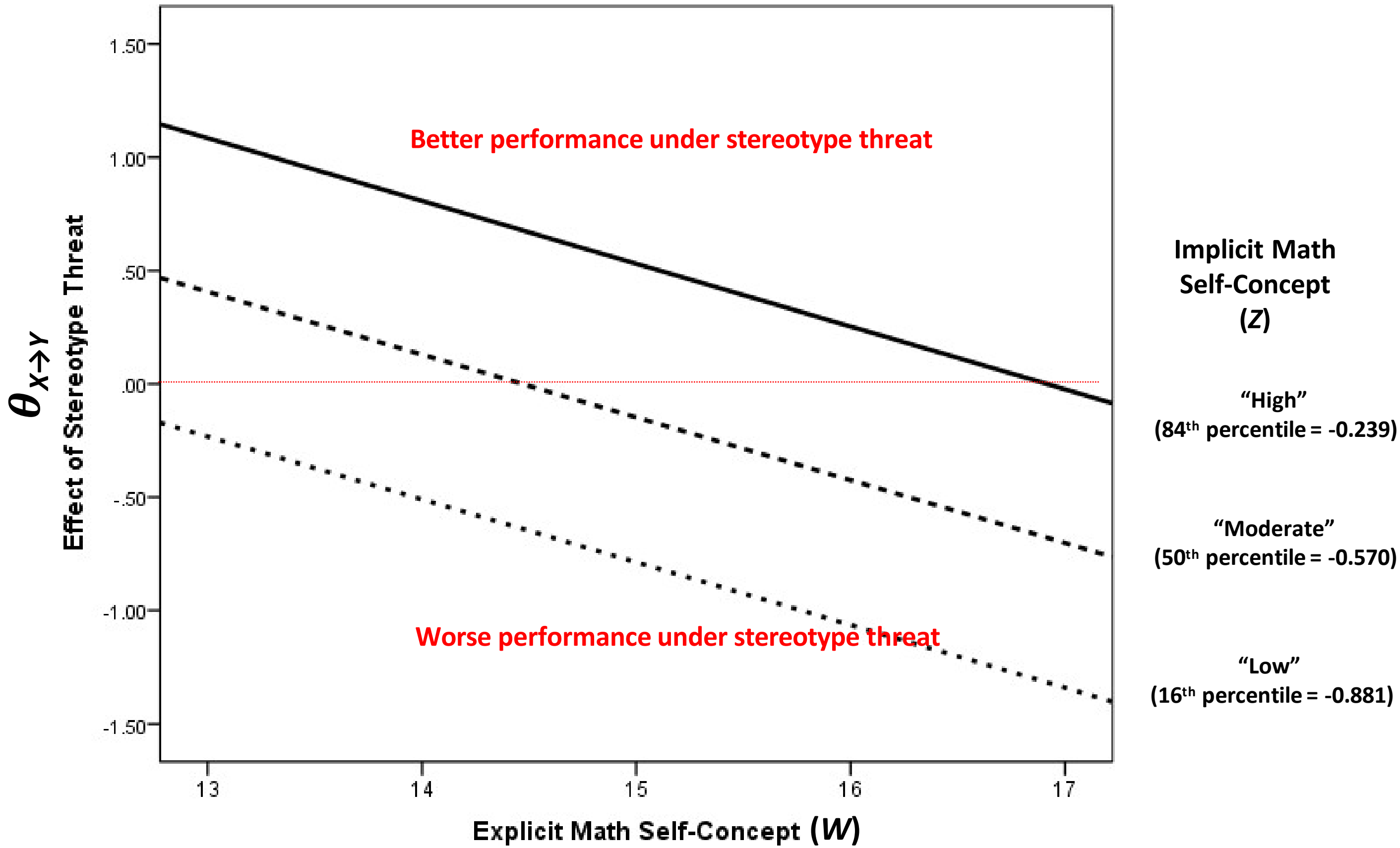
b_1: The amount by which two cases that differ by one unit on *X* but that are equal to zero on *W* and *Z* are estimated to differ on *Y*.

b_4: How much the amount by which two cases that differ by one unit on *X* are estimated to differ on *Y* changes as *W* changes by one unit but *Z* is held fixed.

b_5: How much the amount by which two cases that differ by one unit on *X* are estimated to differ on *Y* changes as *Z* changes by one unit but *W* is held fixed.

Another visual representation

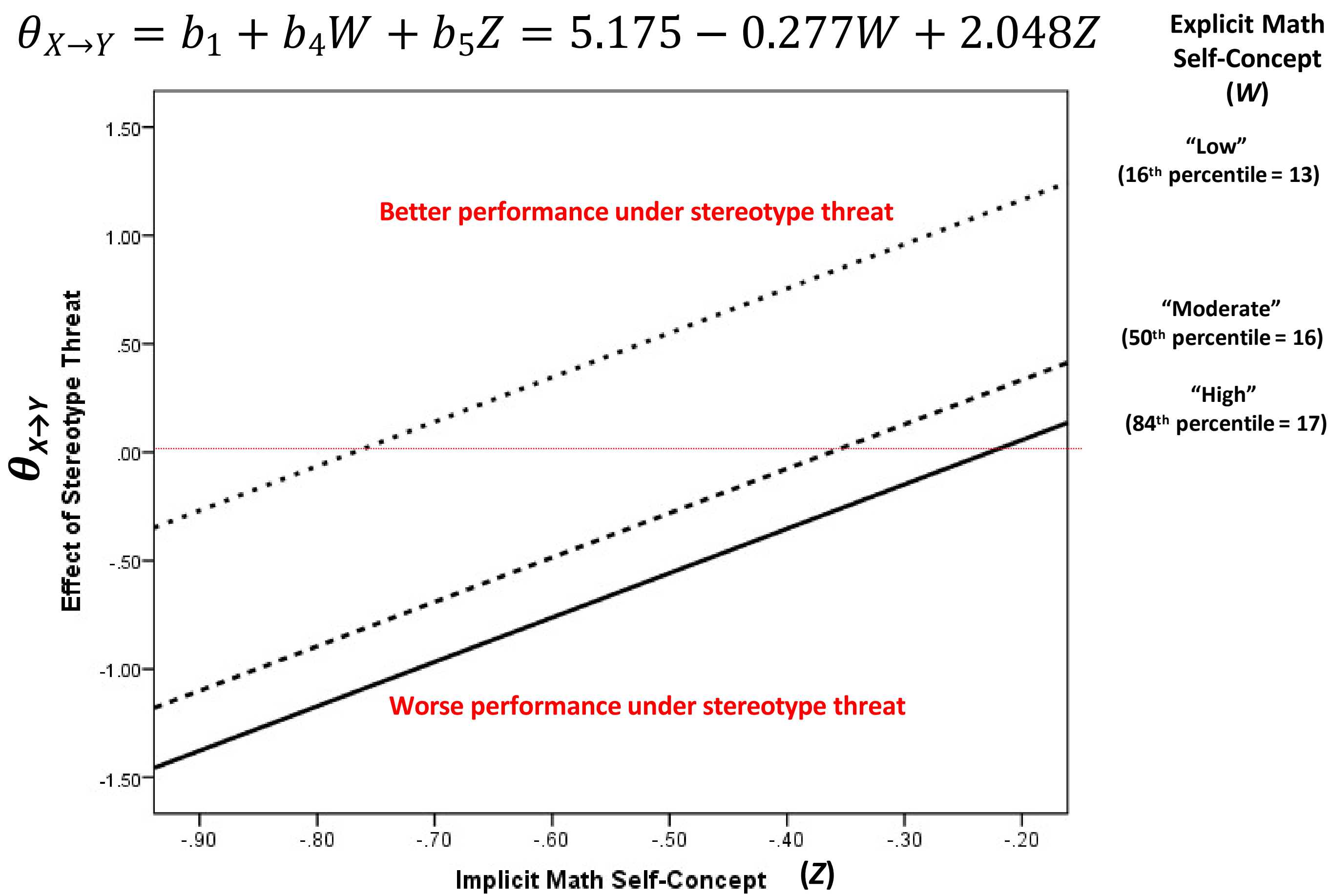
$$\theta_{X\to Y} = b_1 + b_4 W + b_5 Z = 5.175 - 0.277W + 2.048Z$$



b_4 is the slope of these lines. It quantifies how much the effect of X on Y changes as W changes by one unit when Z is held fixed. Note that W's effect on $\theta_{X\to Y}$ doesn't depend on Z.

Another visual representation

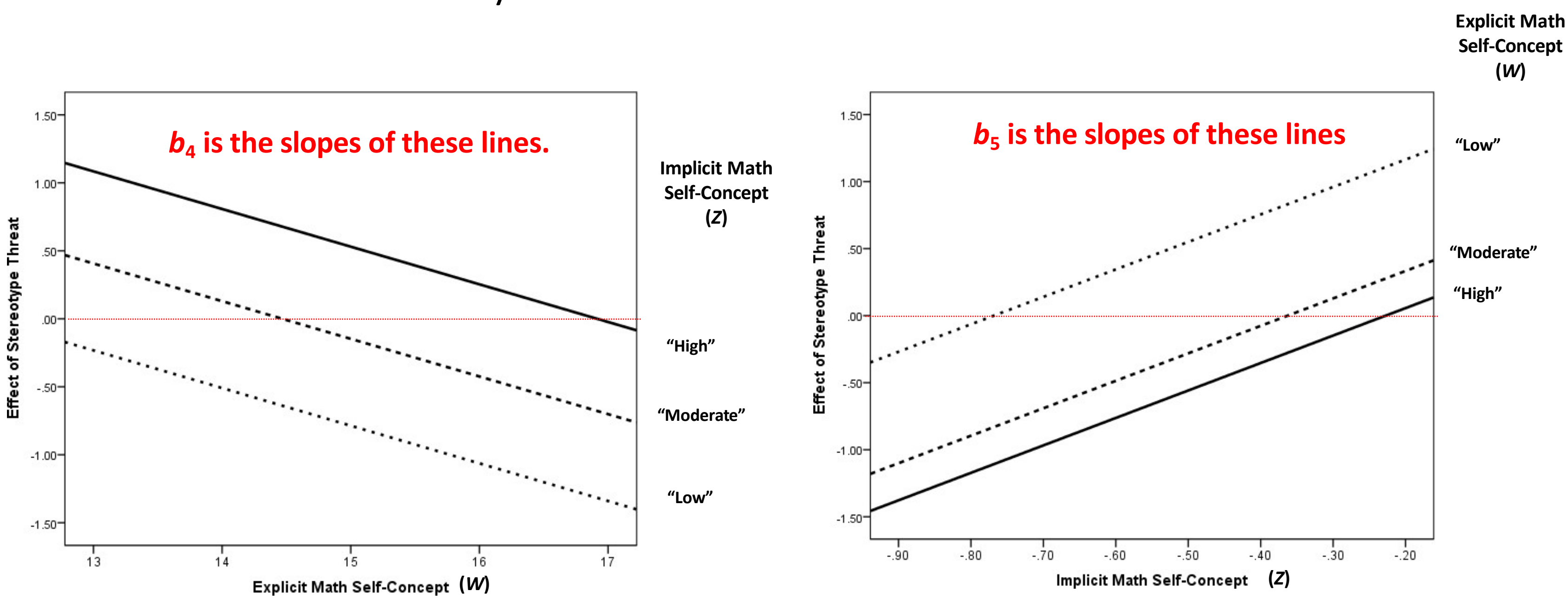
$$\theta_{X\to Y} = b_1 + b_4W + b_5Z = 5.175 - 0.277W + 2.048Z$$



b_5 is the slope of these lines. It quantifies how much the effect of X on Y changes as Z changes by one unit when W is held fixed. Note that Z's effect on $\theta_{X\to Y}$ doesn't depend on W.

Statistical Inference about moderation

These visually depict the association between each putative moderator and the effect of stereotype threat on math performance. Moderation implies that these slopes are different from zero. We need to formally test this.



$$\theta_{X \to Y} = b_1 + b_4 W + b_5 Z = 5.175 - 0.277W + 2.048Z$$

We can test whether these slopes are different from zero individually or as a set.

From PROCESS (Model 2)

```
Model  : 2
    Y  : mathprob
    X  :  threat
    W  :  explms
    Z  : implms

Sample
Size:  136
```

Does explicit math self concept moderate the effect of stereotype threat on math performance, after controlling for the moderating effect of implicit math self concept?

No. b_4 is not statistically different from zero.

```
**************************************************************************
OUTCOME VARIABLE:
 mathprob

Model Summary

          R
   .5940
              R-sq        MSE          F       df1        df2          p
Model        .3528     6.4229    14.1722    5.0000   130.0000      .0000
              coeff         se          t         p       LLCI       ULCI
constant    -2.1587     2.6965     -.8005     .4249    -7.4934     3.1761
threat       5.1751     3.6004     1.4374     .1530    -1.9478    12.2981
explms        .7344      .1679     4.3747     .0000      .4023     1.0665
Int_1        -.2774      .2186    -1.2689     .2068     -.7100      .1551
implms       1.5307      .7908     1.9356     .0551     -.0338     3.0953
Int_2        2.0481     1.0881     1.8824     .0620     -.1045     4.2007

Product terms key:
 Int_1    :        threat    x        explms
 Int_2    :        threat    x        implms

Test(s) of highest order unconditional interaction(s):
        R2-chng          F        df1        df2          p
X*W       .0080     1.6100     1.0000   130.0000      .2068
X*Z       .0176     3.5433     1.0000   130.0000      .0620
BOTH      .0218     2.1916     2.0000   130.0000      .1159
```

Any OLS regression routine would give you this information.

From PROCESS (Model 2)

Does implicit math self concept moderate the effect of stereotype threat on math performance, after controlling for the moderating effect of explicit math self concept?

No (or perhaps). b_5 is not statistically different from zero.

```
Model  : 2
    Y  : mathprob
    X   :  threat
    W   :  explms
    Z  : implms

Sample
Size:  136

**************************************************************************
OUTCOME VARIABLE:
 mathprob

Model Summary

           R
   .5940
              R-sq        MSE          F        df1        df2          p
Model        .3528     6.4229    14.1722     5.0000   130.0000      .0000
              coeff        se          t          p       LLCI       ULCI
constant    -2.1587    2.6965     -.8005      .4249    -7.4934     3.1761
threat       5.1751    3.6004     1.4374      .1530    -1.9478    12.2981
explms        .7344     .1679     4.3747      .0000      .4023     1.0665
Int_1        -.2774     .2186    -1.2689      .2068     -.7100      .1551
implms       1.5307     .7908     1.9356      .0551     -.0338     3.0953
Int_2        2.0481    1.0881     1.8824      .0620     -.1045     4.2007

Product terms key:
 Int_1     :        threat    x        explms
 Int_2     :        threat    x        implms

Test(s) of highest order unconditional interaction(s):

        R2-chng          F        df1        df2          p
X*W       .0080     1.6100     1.0000   130.0000      .2068
X*Z       .0176     3.5433     1.0000   130.0000      .0620
BOTH      .0218     2.1916     2.0000   130.0000      .1159
```

Any OLS regression routine would give you this information.

Testing moderation by the set

We can simultaneously test moderation of X's effect by the set of putative moderators W and Z by comparing the fit of two models using hierarchical regression.

$$\text{Model A: } \widehat{Y}_i = b_0 + b_1 X_i + b_2 W_i + b_3 Z_i$$

$$\text{Model B: } \widehat{Y}_i = b_0 + b_1 X_i + b_2 W_i + b_3 Z_i + b_4 X_i W_i + b_5 X_i Z_i$$

Model A is nested under Model B, so the increase in R^2 that results when the 2 products XW and XZ are added to model A can be converted to an F-ratio and a p-value derived from the $F(2, df_{residual\text{B}})$ distribution, where $df_{residual\text{B}}$ is the residual degrees of freedom for model B. Rejection of the null hypothesis that the change in R^2 equals zero implies that the effect of X depends on math self-concept (explicit and/or implicit)

$$\theta_{X \rightarrow Y} = b_1 + b_4 W + b_5 Z$$

We are testing whether the regression weights for the products are both equal to zero. If so, this implies that the effect of *X* on *Y* is a constant, not a function of *W* or *Z*. The null hypothesis is that they are both zero; the alternative is that at least one is not.

Implementation in SPSS and SAS (including PROCESS)

```
compute thrtexpl = threat*explms.
compute thrtimpl = threat*implms.
regression/statistics defaults change/dep = mathprob/method = enter threat explms implms
                                               /method = enter thrtexpl thrtimpl.
```

```
data math;set math;thrtexpl=threat*explms;thrtimpl=threat*implms;run;
proc reg data=math;model mathprob = threat explms implms thrtexpl thrtimpl;
test thrtexpl=0,thrtimpl=0;run;
```

```
model1<-(lm(mathprob~threat+explms+implms,data=math))
model2<-(lm(mathprob~threat+explms+implms+threat*explms+threat*implms,data=math))
anova(model1,model2)
```

```
Model 1: mathprob ~ threat + explms + implms
Model 2: mathprob ~ threat + explms + implms + threat * explms + threat *
    implms
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    132 863.14
2    130 834.98  2    28.153 2.1916 0.1159
```

Math self concept does not significantly moderate the effect of stereotype threat on math performance, $F(2,130) = 2.192$, $p = 0.12$

From PROCESS (Model 2)

```
Model  : 2
    Y  : mathprob
    X   :  threat
    W   :  explms
    Z  : implms

Sample
Size:  136

**************************************************************************
OUTCOME VARIABLE:
 mathprob

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .5940      .3528     6.4229    14.1722     5.0000   130.0000      .0000

Model
                coeff         se          t          p       LLCI       ULCI
constant      -2.1587     2.6965     -.8005      .4249    -7.4934     3.1761
threat         5.1751     3.6004     1.4374      .1530    -1.9478    12.2981
explms          .7344      .1679     4.3747      .0000      .4023     1.0665
Int_1          -.2774      .2186    -1.2689      .2068     -.7100      .1551
implms         1.5307      .7908     1.9356      .0551     -.0338     3.0953
Int_2          2.0481     1.0881     1.8824      .0620     -.1045     4.2007

Product terms key:
 Int_1    :        threat    x        explms
 Int_2    :        threat    x        implms

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*W      .0080     1.6100     1.0000   130.0000      .2068
X*Z      .0176     3.5433     1.0000   130.0000      .0620
BOTH     .0218     2.1916     2.0000   130.0000      .1159
```

Any OLS regression routine would give you this information.

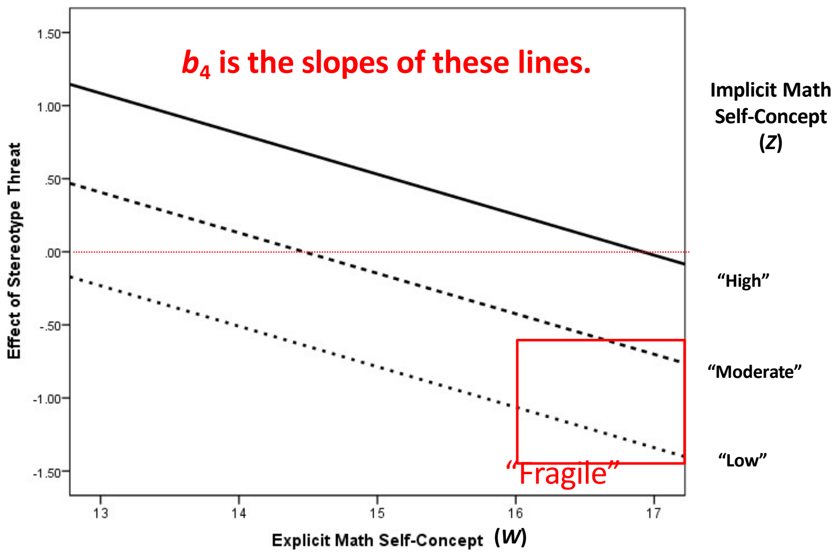
Disappointing? Maybe not.

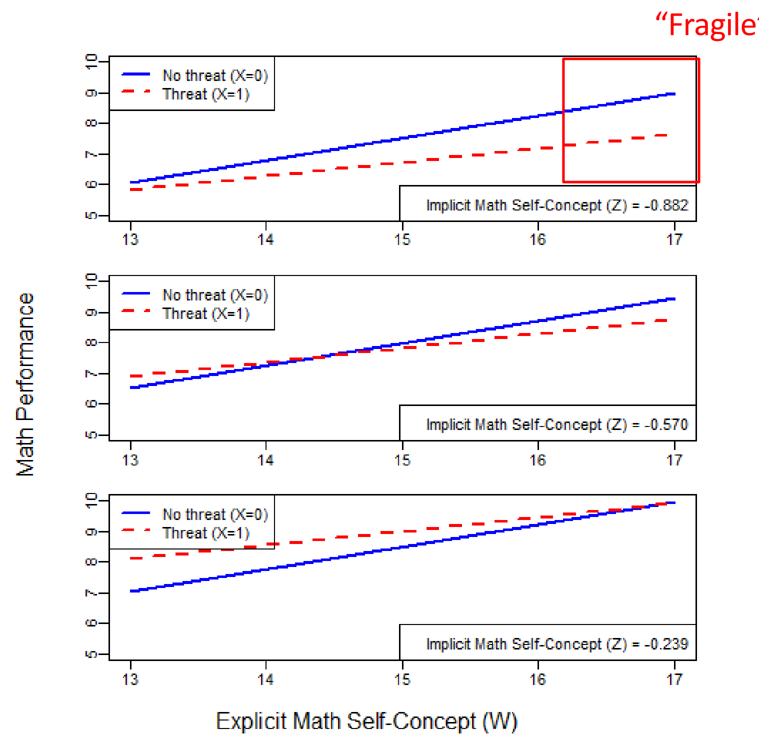
This analysis leads to the (disappointing?) conclusion that threat of stereotype confirmation is not differentially affecting women's performance depending on their math self-concept. But this analysis, thus far, doesn't quite get at what they were predicting.

		Implicit mathematical self-concept	
		High	Low
Explicit mathematical self-concept	High	Consistently high	Fragile
	Low	Damaged	Consistently low

Gerstenberg et al. (2012) singled out a select type of woman who would be affected by threat of stereotype confirmation: one with a "fragile" mathematical self concept, who claims to be mathematical (High Explicit) but who doesn't really identify as such (Low Implicit). People with fragile self- concepts in the stereotype-related domain will feel anxiety and worry about stereotype confirmation, which inhibits performance. They predicted no stereotype threat effect among other types of women.

Where is any stereotype threat effect biggest?





Gerstenberg et al. (2012) predicted a stereotype threat effect (worse performance under stereotype threat than when not under threat) among women with a fragile math self-concept---the combination of high math self concept explicitly measured and low math self concept implicitly measured. It is in that region of the model where we see just such a stereotype threat effect…not so much elsewhere.

Estimating and testing any conditional effect of *X*

We can estimate the effect of *X* on *Y* at any combination of values of *W* and *Z* we choose using the pick-a-point approach. The regression centering strategy makes this fairly easy.

$$\hat{Y}_i = b_0 + (b_1 + b_4 W_i + b_5 Z_i)X_i + b_2 W_i + b_3 Z_i \qquad \theta_{X \to Y} = b_1 + b_4 W + b_5 Z$$

In this model, b_1 estimates the effect of *X* on *Y* when *W* and *Z* = 0. Its standard error, *t* value, and *p*-value tests the null that *X*'s conditional effect equals zero when *W* and *Z* = 0. But if we center *W* and *Z* around values we care about, *w* and *z*

$$W_i' = W_i - w \qquad Z_i' = Z_i - z$$

and estimate:

$$\hat{Y}_i = b_0 + (b_1 + b_4 W'_i + b_5 Z'_i)X_i + b_2 W'_i + b_3 Z'_i \quad \theta_{X \to Y} = b_1 + b_4 W' + b_5 Z'$$

then b_1 estimates the effect of *X* on *Y* when *W* = *w* and *Z* = *z*, and its standard error, *t* value, and *p*-value tests the null that *X*'s conditional effect equals zero at those two values of *W* and *Z*.

No categorization or subgroups analysis needed. We use all the data, operationalizing the "category" as position on the continuum, to quantify the effect of *X* among members of the category.

Estimating and testing any conditional effect of *X*

Let's estimate the effect of stereotype threat among those with a "fragile" math self-concept—"relatively high" in explicit (*W*) but "relatively low" in implicit (*Z*). We'll define these as the 84th percentile of *W* and the 16th percentile of *Z*. Thus, $w = 17$ and $z = -0.882$.

Center *W* around "relatively high" and *Z* around "relatively low" prior to constructing products

```
compute explms_c=explms-17.
compute implms_c=implms-(-0.882).
compute thrtexpl = threat*explms_c.
compute thrtimpl = threat*implms_c.
regression/dep = mathprob/method = enter threat explms_c implms_c thrtexpl thrtimpl.
```

```
data math;set math;explms_c=explms-17;implms_c=implms-(-0.882);
thrtexpl=threat*explms_c;thrtimpl=threat*implms_c;run;
proc reg data=math;model mathprob=threat explms_c implms_c thrtexpl thrtimpl;run;
```

```
explms_c<-math$explms-17;implms_c<-math$implms-(-0.882)
summary(lm(mathprob~threat+explms_c+implms_c+threat*explms_c+threat*implms_c,data=math))
```

$$\theta_{X\rightarrow Y}|(W = 17, Z = -0.882) = -1.348$$

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       8.9763     0.5136  17.476  < 2e-16 ***
threat           -1.3475     0.6649  -2.027   0.0448 *
explms_c          0.7344     0.1679   4.375 2.47e-05 ***
implms_c          1.5307     0.7908   1.936   0.0551 .
threat:explms_c  -0.2774     0.2186  -1.269   0.2068
threat:implms_c   2.0481     1.0881   1.882   0.0620 .
```

Significantly worse performance in the stereotype threat condition among those with a "fragile" math self-concept.

From PROCESS (Model 2)

PROCESS does this automatically. It sees that *W* and *Z* are both continuous, so by default it estimates the conditional effect of *X* at all nine combinations of “relatively low,” “moderate,” and “relatively high” on *W* and *Z*, defined as the 16th, 50th, and 84th percentiles of the distribution.

Conditional effects of the focal predictor at values of the moderator(s):

explms	implms	Effect	se	t	p	LLCI	ULCI
13.0000	-.8816	-.2370	.7271	-.3260	.7450	-1.6755	1.2015
13.0000	-.5700	.4012	.7073	.5672	.5715	-.9981	1.8005
13.0000	-.2392	1.0787	.8495	1.2698	.2064	-.6019	2.7593
16.0000	-.8816	-1.0693	.5658	-1.8899	.0610	-2.1886	.0500
16.0000	-.5700	-.4311	.4518	-.9541	.3418	-1.3250	.4628
16.0000	-.2392	.2464	.5768	.4273	.6699	-.8946	1.3875
17.0000	-.8816	-1.3467	.6647	-2.0261	.0448	-2.6617	-.0317
17.0000	-.5700	-.7085	.5447	-1.3008	.1956	-1.7861	.3691
17.0000	-.2392	-.0310	.6278	-.0493	.9607	-1.2731	1.2111

Use the **moments** option to condition on the mean, a standard deviation below the mean, and a standard deviation above the mean. Or use the **wmodval** and **zmodval** options to specify specific values of *W* and *Z* to condition at (e.g., **in SPSS**: “/wmodval=16/zmodval= -0.882”; **in SAS or R**: “ wmodval=16,zmodval=-0.882”). When a moderator is dichotomous, PROCESS just uses the two values in the data for conditioning.

From PROCESS (Model 2)

		Implicit mathematical self-concept	
		High	Low
Explicit mathematical self-concept	High	Consistently high	Fragile
	Low	Damaged	Consistently low

Conditional effects of the focal predictor at values of the moderator(s):

	explms	implms	Effect	se	t	p	LLCI	ULCI
"consistently-low"	13.0000	-.8816	-.2370	.7271	-.3260	.7450	-1.6755	1.2015
	13.0000	-.5700	.4012	.7073	.5672	.5715	-.9981	1.8005
"damaged"	13.0000	-.2392	1.0787	.8495	1.2698	.2064	-.6019	2.7593
	16.0000	-.8816	-1.0693	.5658	-1.8899	.0610	-2.1886	.0500
	16.0000	-.5700	-.4311	.4518	-.9541	.3418	-1.3250	.4628
	16.0000	-.2392	.2464	.5768	.4273	.6699	-.8946	1.3875
"fragile"	17.0000	-.8816	-1.3467	.6647	-2.0261	.0448	-2.6617	-.0317
	17.0000	-.5700	-.7085	.5447	-1.3008	.1956	-1.7861	.3691
"consistently high"	17.0000	-.2392	-.0310	.6278	-.0493	.9607	-1.2731	1.2111

These results are completely consistent with what they expected to find: A statistically significant effect of stereotype threat only among those with a fragile mathematical self-concept. Among such women, threat of stereotype confirmation lowered performance.

Some problems (to varying degrees)

PROBLEM 1: We have arbitrarily operationally defined “fragile,” “damaged” and so forth as combinations of values of W and Z arbitrarily selected.

True, and if we changed our operationalization, our results might change. But there is no way around this. On most things, people aren’t precategorized nor can their category membership be objectively determined. Operationalization compromises or “conventions” for categorization are often required to some degree.

PROBLEM 2: Difference in significance ≠ significantly different.

Yes. Just because there is an effect of X for one combination of values of W and Z but not for others by a hypothesis testing standard doesn’t mean that those effects of X differ between “groups” defined by different values of W and Z. Indeed, we know from this analysis that the effect of X **doesn’t** depend on W or Z (or both as a set)!

PROBLEM 3: This “additive multiple moderation” model could not actually accurately model results that are exactly as predicted.

Imagine if there was no error in estimation. A model of this form could not actually produce estimates of Y consistent with the predicted results. This is simply not the ideal model to test the question they were asking.

PROBLEM 2: Difference in significance ≠ significantly different.

```
Conditional effects of the focal predictor at values of the moderator(s):
```

	explms	implms	Effect	se	t	p	LLCI	ULCI
"consistently-low"	13.0000	-.8816	-.2370	.7271	-.3260	.7450	-1.6755	1.2015
"damaged"	13.0000	-.2392	1.0787	.8495	1.2698	.2064	-.6019	2.7593
"fragile"	17.0000	-.8816	-1.3467	.6647	-2.0261	.0448	-2.6617	-.0317
"consistently high"	17.0000	-.2392	-.0310	.6278	-.0493	.9607	-1.2731	1.2111

$$\widehat{Y}_i = b_0 + b_1 X_i + b_2 W_i + b_3 Z_i + b_4 X_i W_i + b_5 X_i Z_i$$

- It can be shown that the inferential test for b_4 is also an inferential test for

$$[\theta_{X \to Y} | Z, W = w_1] - [\theta_{X \to Y} | Z, W = w_2]$$

regardless of Z and the two values of W chosen (w_2 and w_1). So a claim that W does not moderate X's effect means that no two conditional effects of X on Y defined by the same value of Z but different values of W are statistically different from each other.

- It can be shown that the inferential test for b_5 is also an inferential test for

$$[\theta_{X \to Y} | W, Z = z_1] - [\theta_{X \to Y} | W, Z = z_2]$$

regardless of W and the two values of Z chosen (z_2 and z_1). So a claim that Z does not moderate X's effect means that no two conditional effects of X on Y defined by the the same value of W but different values of Z are statistically different from each other.

Comparing conditional effects conditioned on two moderators

Nothing in the analysis we've done allows for an inference about the difference between conditional effects of *X* when both *W* and *Z* differ:

$$[\theta_{X\to Y}|Z = z_1, W = w_1] - [\theta_{X\to Y}|Z = z_2, W = w_2]$$

For example, is the effect of stereotype threat among the "fragile" (w_1 = 17, z_1 = -0.8816) different than the effect among the "damaged"? (w_2 = 13, z_2 =-0.2392).

Conditional effects of the focal predictor at values of the moderator(s):

	explms	implms	Effect	se	t	p	LLCI	ULCI
"consistently-low"	13.0000	-.8816	-.2370	.7271	-.3260	.7450	-1.6755	1.2015
"damaged"	13.0000	-.2392	1.0787	.8495	1.2698	.2064	-.6019	2.7593
"fragile"	17.0000	-.8816	-1.3467	.6647	-2.0261	.0448	-2.6617	-.0317
"consistently high"	17.0000	-.2392	-.0310	.6278	-.0493	.9607	-1.2731	1.2111

$$[\theta_{X\to Y}|Z = 17, W = -0.8816] - [\theta_{X\to Y}|Z = 13, W = -0.2392]$$

-1.347 – 1.079 = -2.426

Is this difference statistically different than zero?

Dawson & Richter (2006, *Journal of Applied Psychology,* p. 917-926) provide a discussion of the comparison of conditional effect in models with more than one moderator. The mathematics is fairly tedious and mistakes are easily made. PROCESS makes it simple.

Comparing conditional effects with PROCESS (models 2 and 3 only)

The **contrast** option in PROCESS can be used for model 2 to compare to conditional effects of X when $W = w_1$ and $Z = z_1$ compared to when $W = w_2$ and $Z = z_2$.

For example, is the effect of stereotype threat among the "fragile" (w_1 = 17, z_1 = -0.8816) different than the effect among the "damaged"? (w_2 = 13, z_2 =-0.2392).

w_1 z_1 w_2 z_2

```
process  y=mathprob/x=threat/w=explms/z=implms/model=2/contrast=17,-0.8816;13,-0.2392.
```

```
%process  (data=math,y=mathprob,x=threat,w=explms,z=implms,model=2,
    contrast=17 -0.8816 13 -0.2392);
```

```
process(data=math,y="mathprob",x="threat",w="explms",z="implms",model=2,
   contrast=c(17,-0.8816,13,-0.2392))
```

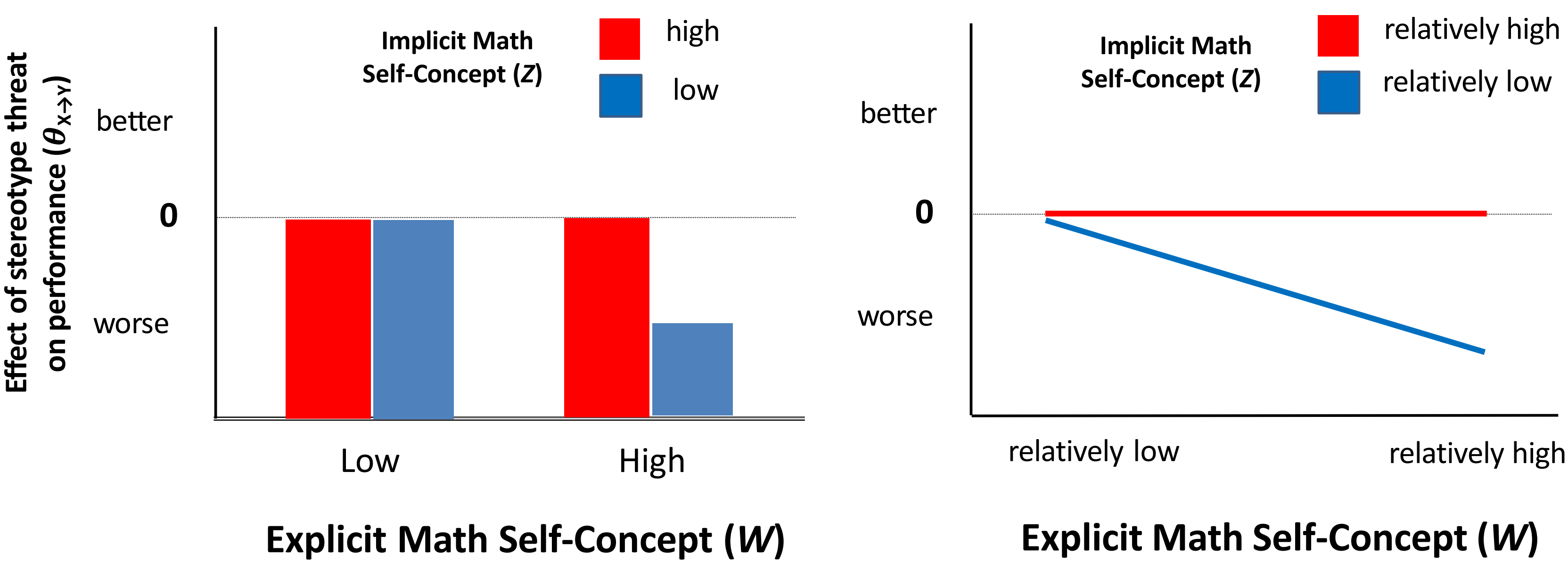
```
        Contrast between conditional effects of X:

                    explms       implms       Effect
        Effect1:    17.0000      -.8816      -1.3467
        Effect2:    13.0000      -.2392       1.0787

        Test of Effect1 minus Effect2
            Contrast          se          t            p        LLCI       ULCI
             -2.4254      1.2223    -1.9843        .0493     -4.8435     -.0073
```

The effect of stereotype threat is different among the fragile than among the damaged, $t(130) = -1.985$, $p = .049$.

PROBLEM 3: Wrong model.

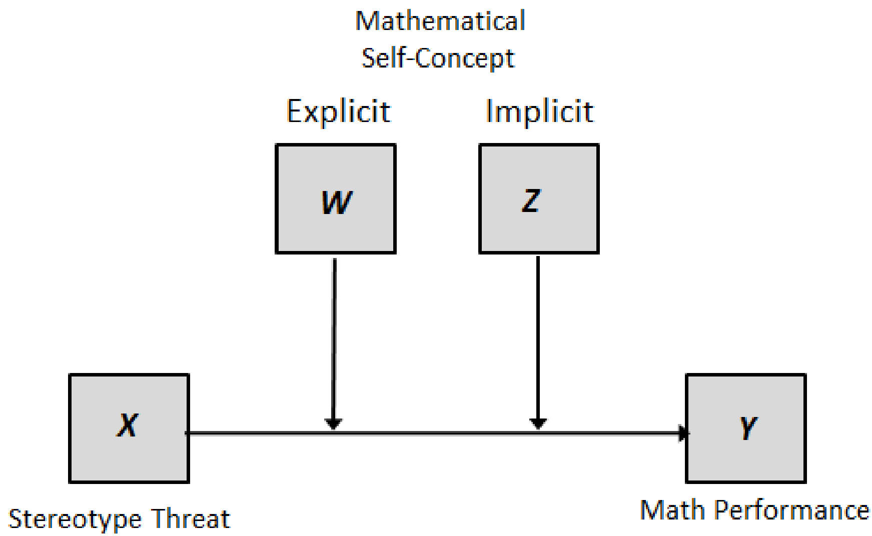
Gerstenberg et al. (2012) argue that among those higher in explicit math self-concept but lower in implicit math concept, stereotype threat should lower performance. Among others, stereotype threat should have little to no effect on performance. They predict something like:

In dichotomous space: **In continuous space:**



So they predict that the effect of explicit math self-concept (W) on the effect of stereotype threat on performance ($\theta_{X\to Y}$) would depend on implicit math self-concept (Z). That is, the moderation of X's effect on Y by W should depend on Z.

The additive multiple moderator model has a constraint



$$Y_i = b_0 + b_1 X_i + b_2 W_i + b_3 Z_i + b_4 X_i W_i + b_5 X_i Z_i$$

or

$$Y_i = b_0 + (b_1 + b_4 W_i + b_5 Z_i) X_i + b_2 W_i + b_3 Z_i$$

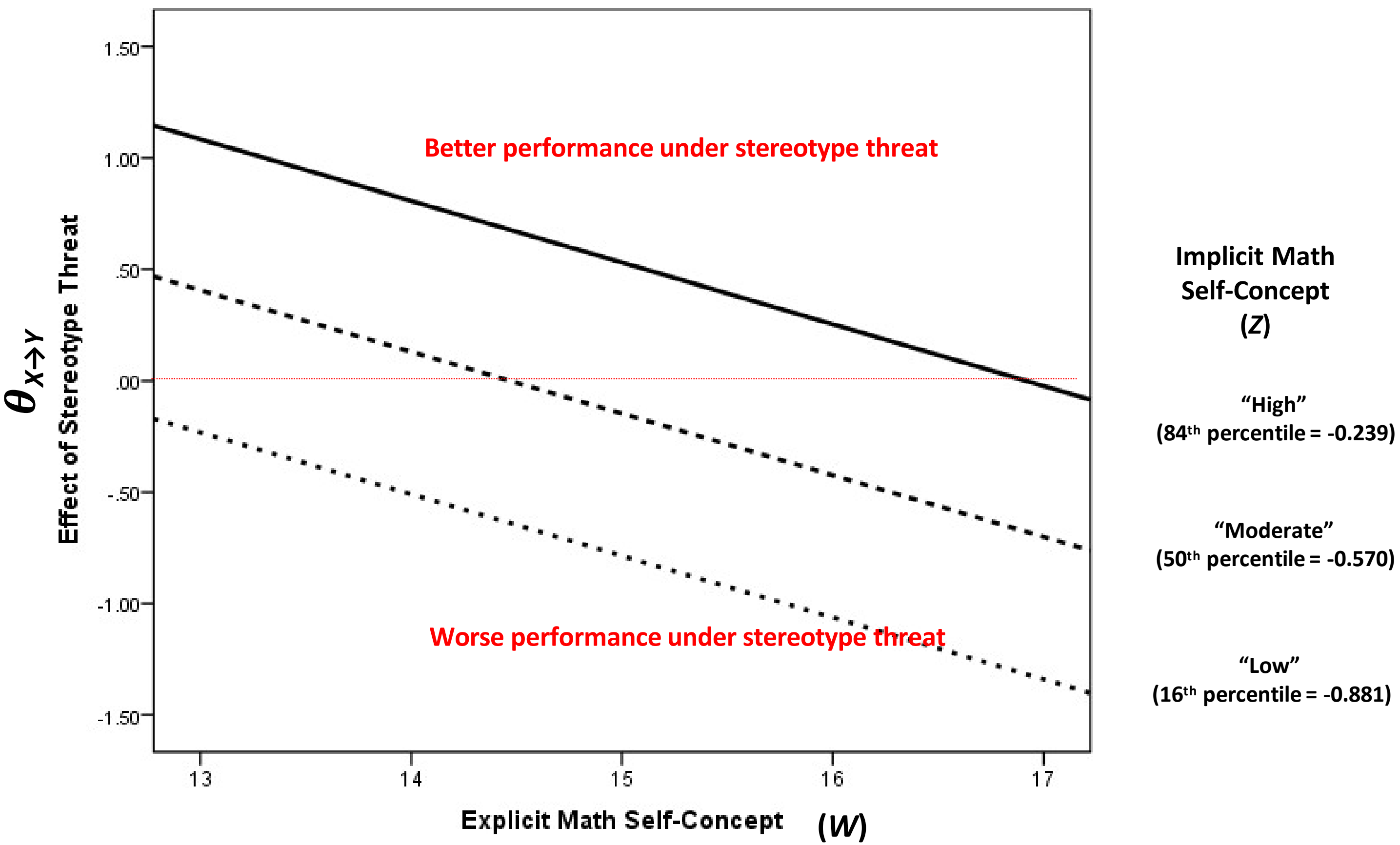
or

$$Y_i = b_0 + \theta_{X \to Y} X_i + b_2 W_i + b_3 Z_i \quad \text{where} \quad \theta_{X \to Y} = b_1 + b_4 W + b_5 Z$$

In this model, *W*'s effect on $\theta_{X \to Y}$ is b_4. It is a constant. *W*'s effect on $\theta_{X \to Y}$ is not related to or determined by *Z*. *W*'s effect on $\theta_{X \to Y}$ is fixed to be invariant across values of *Z*.

Recall our visual representation of the model

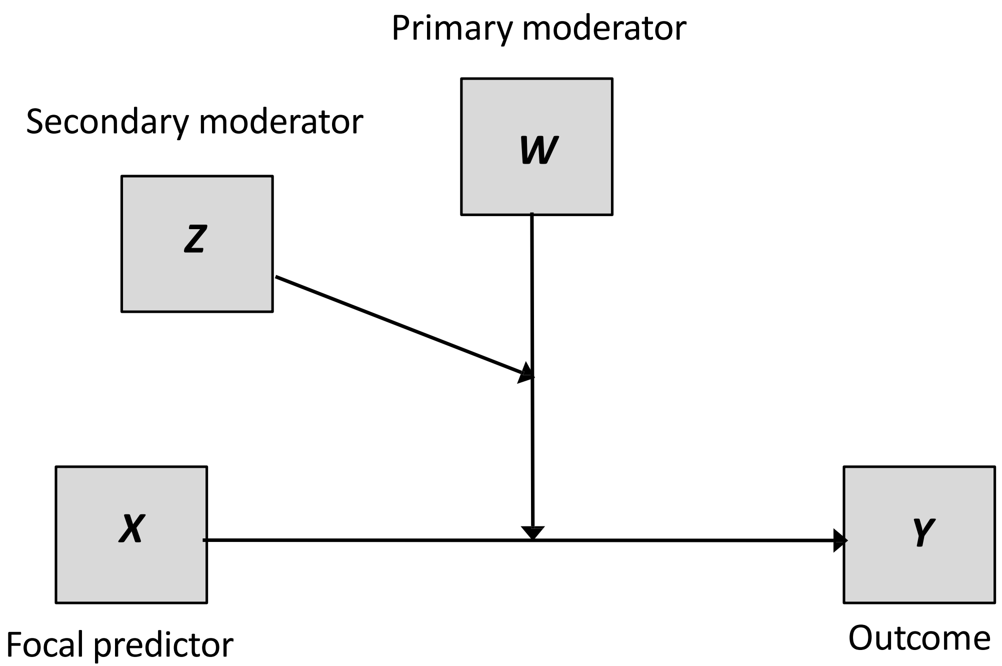
$$\theta_{X\to Y} = b_1 + b_4 W + b_5 Z = 5.175 - 0.277W + 2.048Z$$



b_4 is the slope of these lines. It quantifies how much the effect of *X* on *Y* changes as *W* changes by one unit when *Z* is held fixed. *W*'s effect on $\theta_{X\to Y}$ doesn't depend on *Z*. **It can't by this model.**

Relaxing the constraint: The "moderated moderation" model

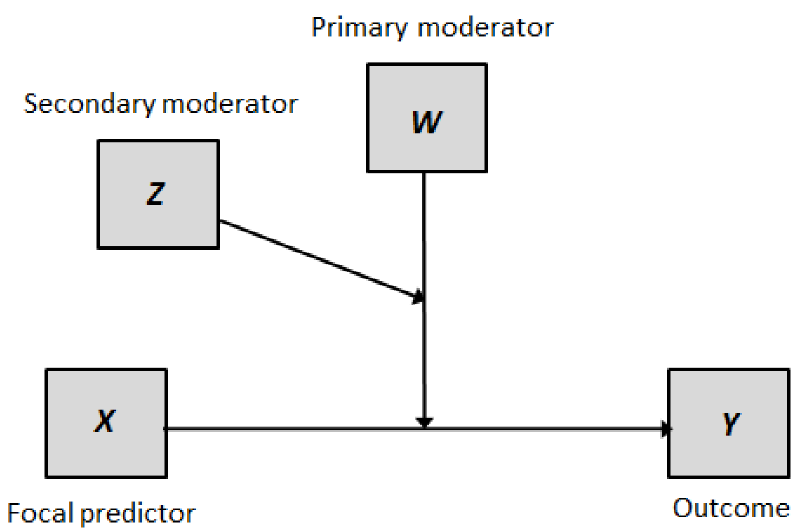
Moderation can be moderated, as depicted conceptually below. Here, *W*'s influence on the effect of *X* on *Y* is moderated by *Z*. So *W*'s role as a moderator of *X*'s effect on *Y* depends on *Z*. Such a model allows the effect of *X* on *Y* to vary as a **nonadditive combination** of *W* and *Z*.



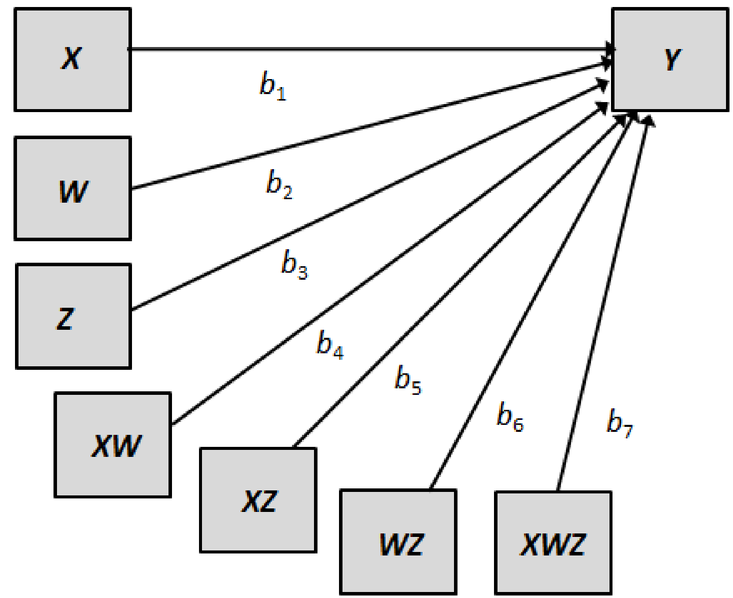
W's effect on the effect of *X* on *Y* is estimated with an interaction between *W* and *X*, sometimes called a "two way" interaction (i.e., two variables). *Z*'s effect on the effect of *W*'s effect on the effect *X* on *Y* is estimated with a *three-way* interaction between *X*, *W*, and *Z*.

Moderated Moderation ("multiplicative multiple moderation")

Conceptual diagram



Statistical diagram



$$\widehat{Y}_i = b_0 + b_1 X_i + b_2 W_i + b_3 Z_i + b_4 X_i W_i + b_5 X_i Z_i + b_6 W_i Z_i + b_7 X_i W_i Z_i$$

W's effect on *X*'s effect on *Y* now depends linearly on *Z*.

$$\hat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3Z_i + b_4X_iW_i + b_5X_iZ_i + b_6W_iZ_i + b_7X_iW_iZ_i$$

can be written equivalently as

$$\hat{Y}_i = b_0 + (b_1 + b_4W_i + b_5Z_i + b_7W_iZ_i)X_i + b_2W_i + b_3Z_i + b_6W_iZ_i$$

or

$$\hat{Y}_i = b_0 + \theta_{X\to Y}X_i + b_7X_iW_iZ_i + b_2W_i + b_3Z_i + b_6W_iZ_i$$

where $\theta_{X\to Y} = b_1 + b_4W + b_5Z + b_7WZ$

is the conditional effect of *X* on *Y*. The conditional effect of *X* on *Y* can be equivalently written as

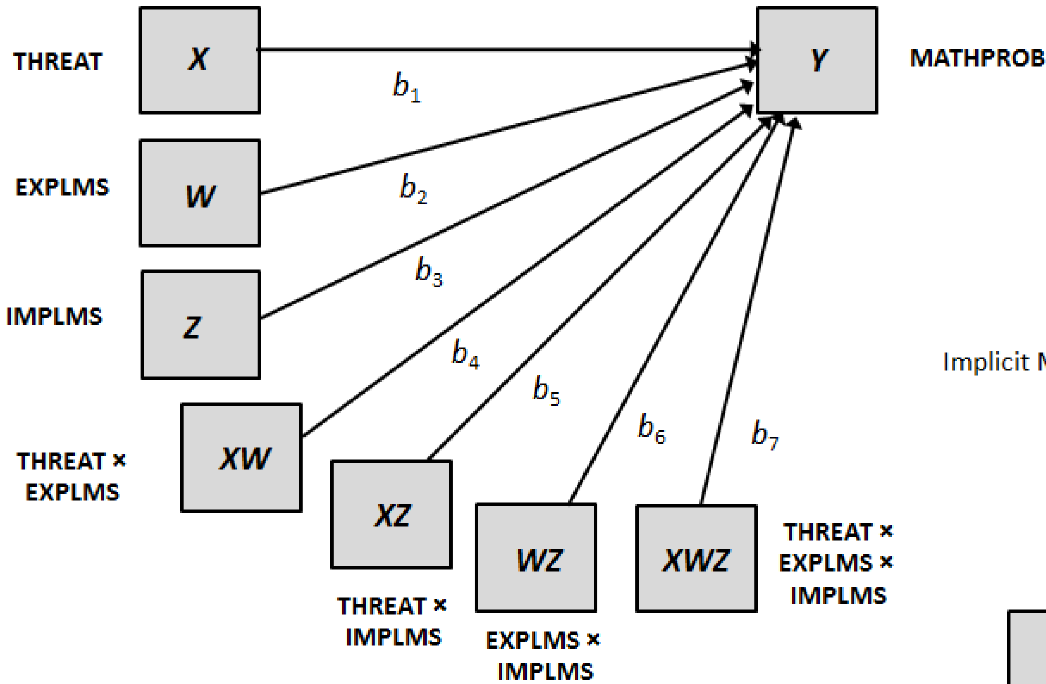
$$\theta_{X\to Y} = b_1 + (b_4 + b_7Z_i)W_i + b_5Z_i)$$

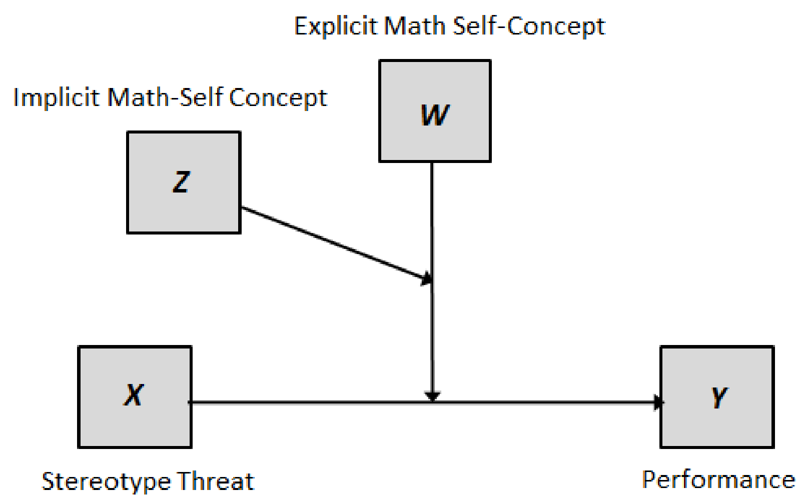
or

$\theta_{X\to Y} = b_1 + \theta_{XW\to Y}W + b_5Z$, where $\theta_{XW\to Y} = b_4 + b_7Z$ **is the conditional moderation of *X*'s effect on *Y* by *W*** (or a "conditional *X* by *W* interaction"). **It depends linearly on *Z*.**

$\theta_{XW\to Y} = b_4 + b_7Z$ is a function of *Z* **unless b_7 = 0**. We test a hypothesis about moderated moderation by testing whether the regression weight for *XWZ* equals zero.

Estimating the model





Does implicit math self-concept moderate the moderation of the effect of explicit math self-concept on the effect of stereotype threat on performance? b_7 gives us the answer.

Estimation in SPSS, SAS, and R

```
compute thrtexpl = threat*explms.  compute thrtimpl = threat*implms.  compute
explimpl = explms*implms.  compute condexim = threat*explms*implms.
regression/dep = mathprob/method = enter threat explms implms thrtexpl thrtimpl explimpl condexim.
```

```
data math;set math;thrtexpl=threat*explms;thrtimpl=threat*implms;
explimpl=explms*implms;condexim=threat*explms*implms;run;
proc reg data=math;model mathprob=threat explms implms thrtexpl thrtimpl explimpl
condexim;run;
```

```
summary(lm(mathprob~threat+explms+implms+threat*explms+threat*implms+explms*impl
ms+threat*explms*implms, data=math))
```

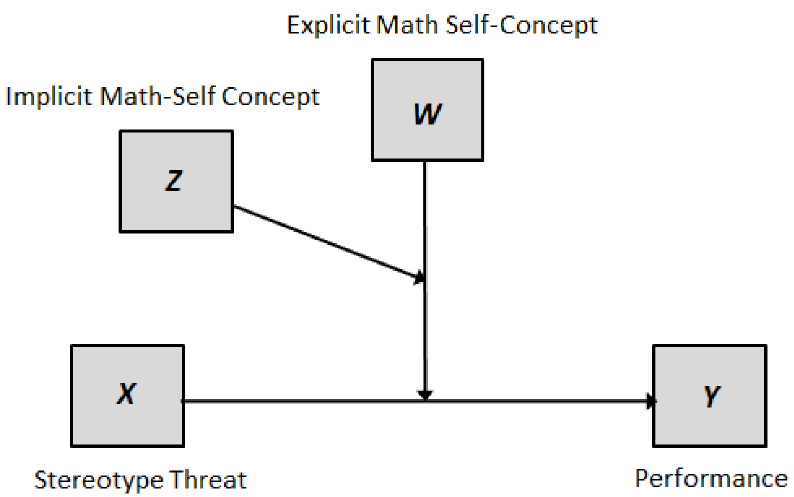
$$\widehat{Y}_i = -0.4611 - 4.2911X_i + 0.6235W_i + 4.5829Z_i + 0.2767X_iW_i - 13.9269X_iZ_i - 0.2021W_iZ_i + 0.9616X_iW_iZ_i$$

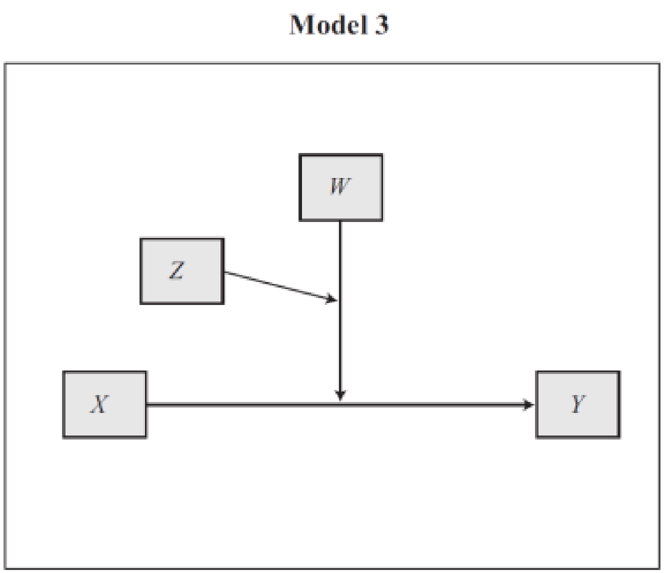
```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           -0.4611     4.2791  -0.108   0.9144
threat                -4.2911     5.4802  -0.783   0.4351
explms                 0.6235     0.2747   2.270   0.0249 *
implms                 4.5829     6.1320   0.747   0.4562
threat:explms          0.2767     0.3386   0.817   0.4154
threat:implms        -13.9269     7.4633  -1.866   0.0643 .
explms:implms         -0.2021     0.4029  -0.502   0.6167
threat:explms:implms   0.9616     0.4722   2.036   0.0438 *
```

Moderated moderation! The effect of explicit math self-concept on the effect of stereotype threat depends on implicit math self-concept. To understand this, we draw a picture and probe.

Let PROCESS help you estimate, test, and visualize

We can use PROCESS model 3 to estimate this model, do some visualization, and lots of other things.






```
process y=mathprob/x=threat/w=explms/z=implms/model=3/plot=1/jn=1.
```

```
%process (data=math,y=mathprob,x=threat,w=explms,z=implms,model=3,plot=1,jn=1);
```

```
process(data=math,y="mathprob",x="threat",w="explms",z="implms",model=3,
 plot=1,jn=1)
```

PROCESS output (Model 3)

```
Model  : 3
    Y  : mathprob
    X  : threat
    W  : explms
    Z  : implms

Sample
Size:  136

**************************************************************************
OUTCOME VARIABLE:
 mathprob

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .6314      .3986     6.0613    12.1205     7.0000   128.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant     -.4611     4.2791     -.1078      .9144    -8.9281     8.0059
threat      -4.2911     5.4802     -.7830      .4351   -15.1347     6.5525
explms        .6235      .2747     2.2699      .0249      .0800     1.1671
Int_1         .2767      .3386      .8171      .4154     -.3933      .9466
implms       4.5829     6.1320      .7474      .4562    -7.5503    16.7161
Int_2      -13.9269     7.4633    -1.8661      .0643   -28.6944      .8405
Int_3        -.2021      .4029     -.5017      .6167     -.9993      .5950
Int_4         .9616      .4722     2.0363      .0438      .0272     1.8960

Product terms key:
 Int_1    :        threat    x        explms
 Int_2    :        threat    x        implms
 Int_3    :        explms    x        implms
 Int_4    :        threat    x        explms    x        implms

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*W*Z     .0195     4.1464     1.0000   128.0000      .0438
```

$$Y_i = -0.461 - 4.291X_i + 0.624W_i + 4.583Z_i$$
$$+0.277X_iW_i - 13.927X_iZ_i - 0.202W_iZ_i + 0.962X_iW_iZ_i$$

PROCESS knows what products to construct. It does it for you.

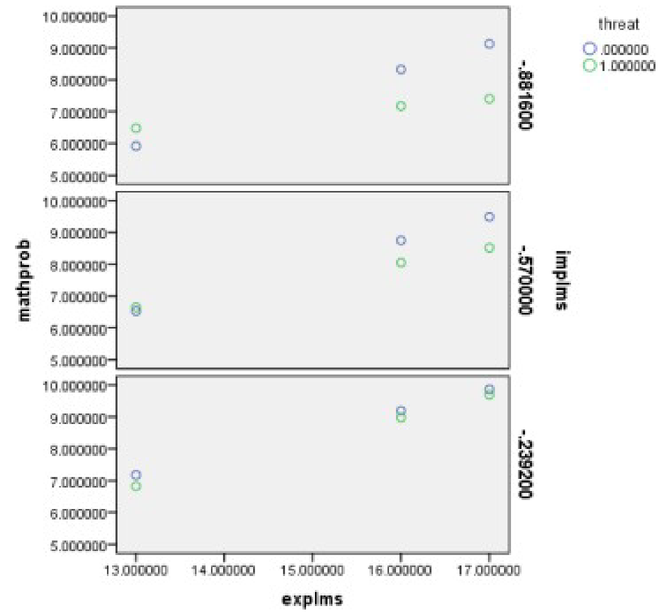
Mathematically equivalent tests of moderated moderation.

A visual representation of the effect of *X* on *Y*

The **plot** option in SPSS generates code to visualize. Cut and paste this into syntax and run.

$$\widehat{Y}_i = -0.4611 - 4.2911X_i + 0.6235W_i + 4.5829Z_i + 0.2767X_iW_i - 13.9269X_iZ_i - 0.2021W_iZ_i + 0.9616X_iW_iZ_i$$

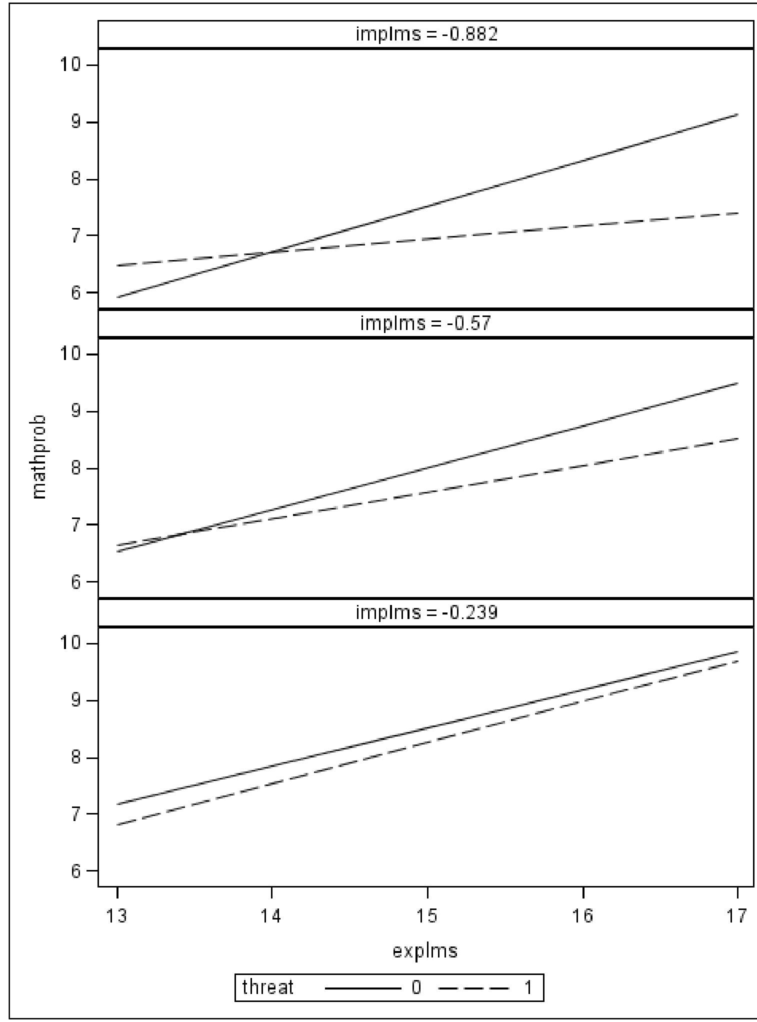
```
DATA LIST FREE/
   threat      explms       implms        mathprob  .
 BEGIN DATA.
       .0000     13.0000      -.8816      5.9209
      1.0000     13.0000      -.8816      6.4840
       .0000     13.0000      -.5700      6.5302
      1.0000     13.0000      -.5700      6.6488
       .0000     13.0000      -.2392      7.1770
      1.0000     13.0000      -.2392      6.8238
       .0000     16.0000      -.8816      8.3260
      1.0000     16.0000      -.8816      7.1760
       .0000     16.0000      -.5700      8.7464
      1.0000     16.0000      -.5700      8.0508
       .0000     16.0000      -.2392      9.1926
      1.0000     16.0000      -.2392      8.9794
       .0000     17.0000      -.8816      9.1277
      1.0000     17.0000      -.8816      7.4067
       .0000     17.0000      -.5700      9.4851
      1.0000     17.0000      -.5700      8.5181
       .0000     17.0000      -.2392      9.8645
      1.0000     17.0000      -.2392      9.6979
END DATA.
GRAPH/SCATTERPLOT=
 explms    WITH    mathprob BY   threat    /PANEL    ROWVAR=   implms.
```



This can be edited to make it look prettier.

In SAS

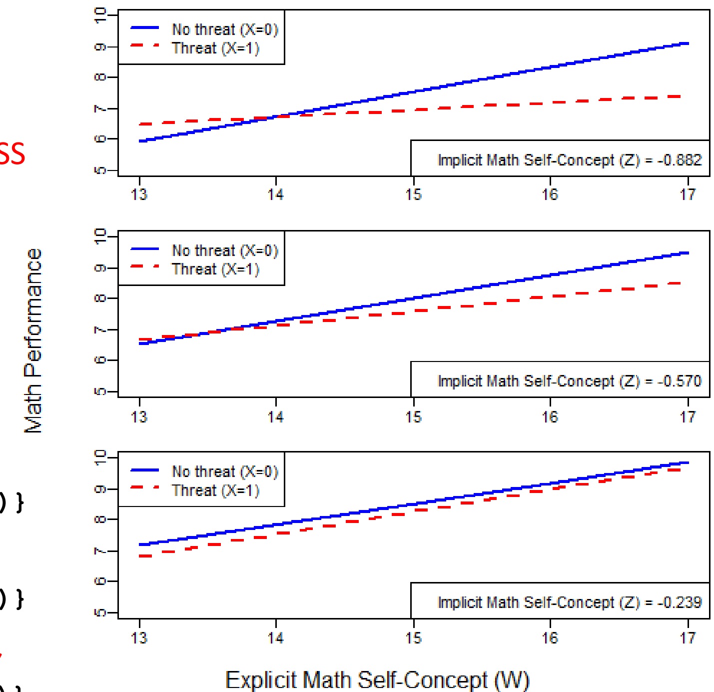
```
data;input threat explms implms mathprob;
datalines;
     .0000    13.0000     -.8816       5.9209
    1.0000    13.0000     -.8816       6.4840
     .0000    13.0000     -.5700       6.5302
    1.0000    13.0000     -.5700       6.6488
     .0000    13.0000     -.2392       7.1770
    1.0000    13.0000     -.2392       6.8238
     .0000    16.0000     -.8816       8.3260
    1.0000    16.0000     -.8816       7.1760
     .0000    16.0000     -.5700       8.7464
    1.0000    16.0000     -.5700       8.0508
     .0000    16.0000     -.2392       9.1926
    1.0000    16.0000     -.2392       8.9794
     .0000    17.0000     -.8816       9.1277
    1.0000    17.0000     -.8816       7.4067
     .0000    17.0000     -.5700       9.4851
    1.0000    17.0000     -.5700       8.5181
     .0000    17.0000     -.2392       9.8645
    1.0000    17.0000     -.2392       9.6979

run;
proc sgpanel;
panelby implms / columns=1;
series x=explms y=mathprob/group=threat
lineattrs =(color=black);run;
```



In R

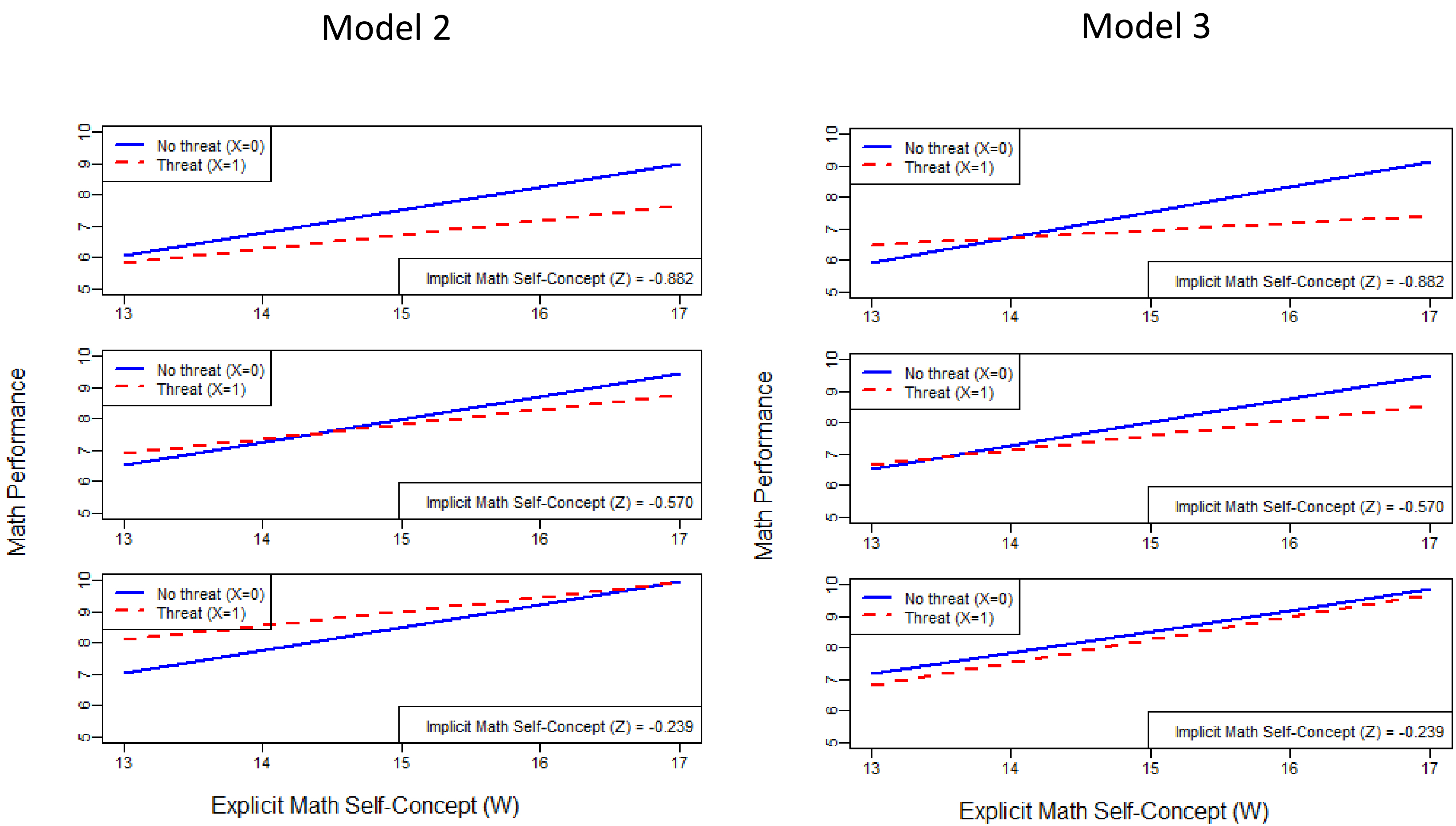
```
oldp<-par(mfrow=c(3,1),mar=c(3,4,0,0),oma=c(2,2,2,2),
mgp=c(5,0.5,0))
w<-c(13,16,17,13,16,17)
x<-c(0,0,0,1,1,1)
yzlow<-c(5.921,8.326,9.128,6.484,7.176,7.407)
yzmod<-c(6.530,8.746,9.485,6.649,8.051,8.518)
yzhigh<-c(7.177,9.193,9.865,6.824,8.979,9.698)
wt<-x
x<-w
w<-wt
legend.txt<-c("No threat (X=0)","Threat (X=1)")
for (i in 1:3){
if (i==1)
  {y<-yzlow
  legend2.txt<-c("Implicit Math Self-Concept (Z) = -0.882")}
if (i==2)
  {y<-yzmod
  legend2.txt<-c("Implicit Math Self-Concept (Z) = -0.570")}
if (i==3)
  {y<-yzhigh
  legend2.txt<-c("Implicit Math Self-Concept (Z) = -0.239")}
plot(y=y,x=x,col="white",ylim=c(5,10),cex=1.5,xlim=c(13,17),tcl=-0.5)
lines(x[w==0],y[w==0],lwd=2,lty=1,col="blue")
lines(x[w==1],y[w==1],lwd=2,lty=2,col="red")
legend("topleft", legend=legend.txt,lwd=2,lty=c(1,2),
col=c("blue","red"))
legend("bottomright",legend=legend2.txt)
box}
mtext("Explicit Math Self-Concept (W)",side=1,outer=TRUE)
mtext("Math Performance",side=2,outer=TRUE)
par<-oldp
```

From PROCESS **plot** output



OR USE THE SAVE OPTION

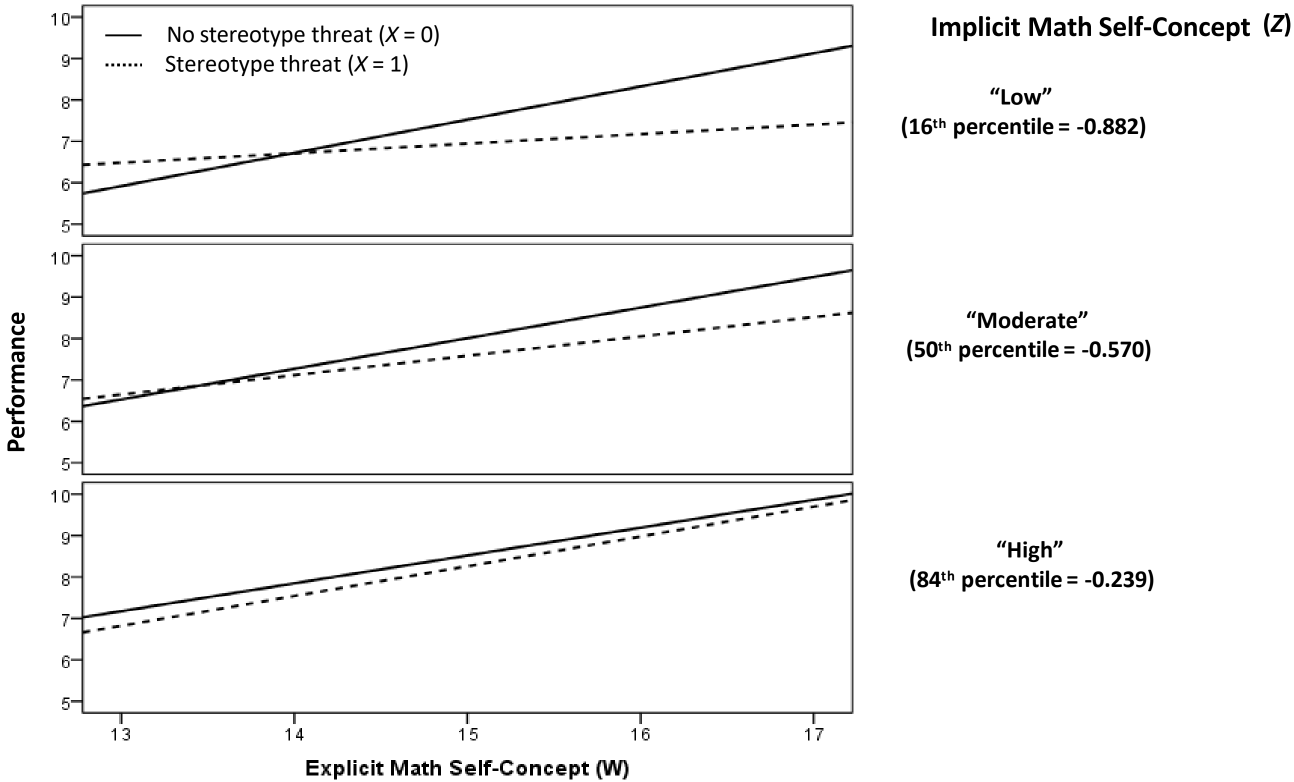
The rate of change of the effect of *X* as *W* changes



Rate of change independent of Z

Rate of change depends on Z

A visual representation of the effect of *X* on *Y*



$$\hat{Y}_i = -0.4611 - 4.2911X_i + 0.6235W_i + 4.5829Z_i + 0.2767X_iW_i - 13.9269X_iZ_i - 0.2021W_iZ_i + 0.9616X_iW_iZ_i$$

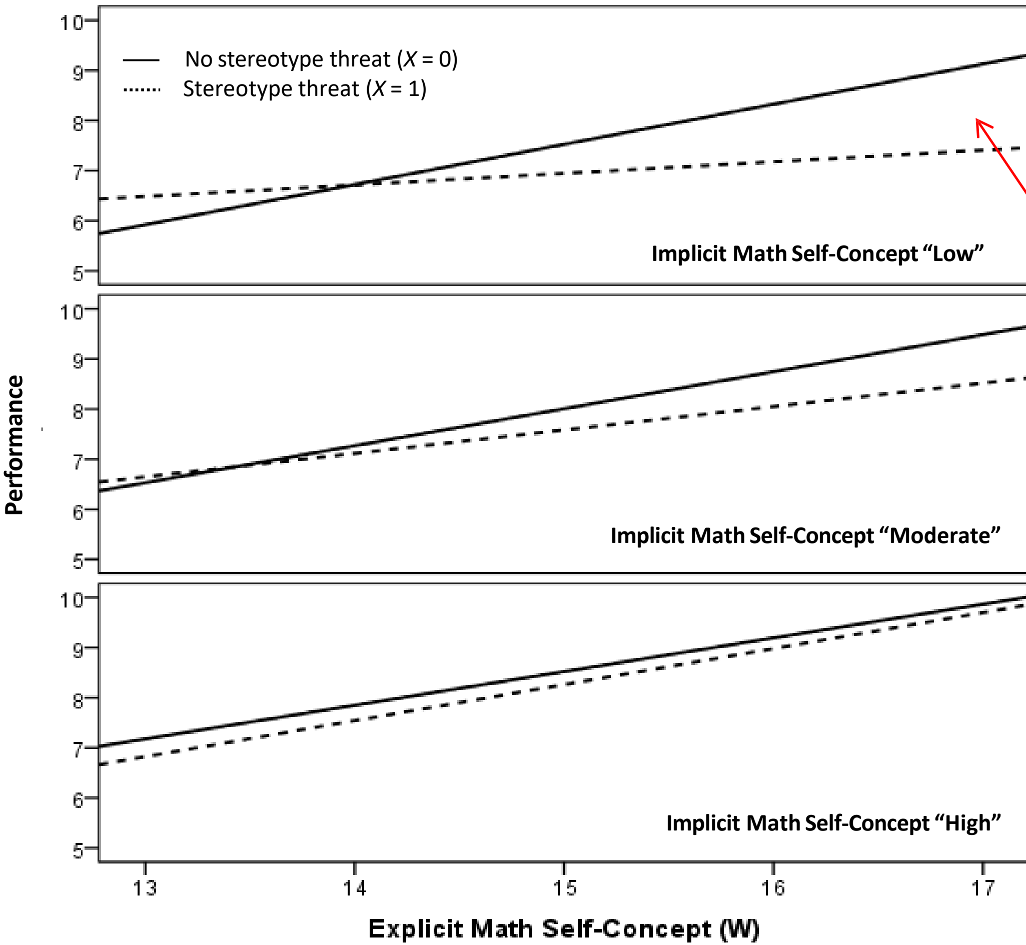
$$\hat{Y}_i = -0.4611 + (-4.2911 + 0.2767W_i - 13.9269Z_i + 0.9616W_iZ_i)X_i + 0.6235W_i + 4.5829Z_i - 0.2021W_iZ_i$$

$$\hat{Y}_i = -0.4611 + \theta_{X\rightarrow Y}X_i + 0.6235W_i + 4.5829Z_i - 0.2021W_iZ_i$$

where $\theta_{X\rightarrow Y} = -4.921 + 0.277W - 13.927Z + 0.962WZ$

$\boldsymbol{\theta_{X\rightarrow Y}}$ is the distance between the solid and the dotted line at values of *W* and *Z*.

Visualizing the model

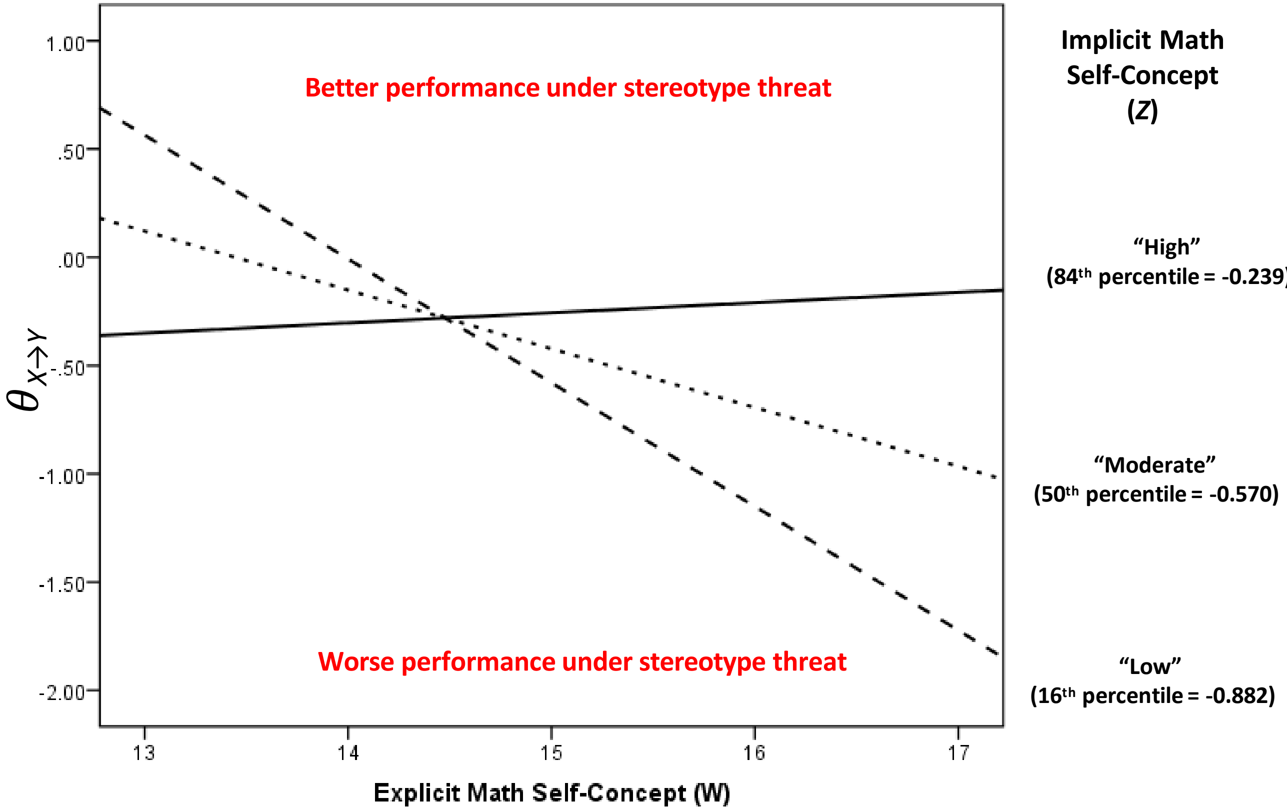


Observe that the moderation of the effect of stereotype threat by explicit self-concept (the $X \times W$ interaction) varies as a function of implicit self-concept (Z).

Primarily, the effect seems to be driven by a reduction in the number of problems solved as a result of stereotype threat that occurs primarily when relatively high explicit math concept is paired with relatively low implicit self-concept

Another visual representation

$$\theta_{X\to Y} = b_1 + b_4W + b_5Z + b_7WZ = -4.291 + 0.277W - 13.927Z + 0.962WZ$$



b_7 quantifies the change in the slope of the line linking W to $\theta_{X\to Y}$ as Z changes. W's effect on $\theta_{X\to Y}$ depends on Z.

Probing moderated moderation

A sensible strategy for probing the moderation of a moderated effect is to estimate the moderation of the effect of the focal predictor *X* by the primary moderator *W* at values of the secondary moderator *Z*. This is a pick-a-point approach to probing three-way interaction.

$$\hat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3Z_i + b_4X_iW_i + b_5X_iZ_i + b_6W_iZ_i + b_7X_iW_iZ_i$$

can be rewritten as

$$\hat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3Z_i + (b_4 + b_7Z_i)X_i\,W_i + b_5X_iZ_i + b_6W_iZ_i$$

which illustrates that the moderation of *X*'s effect on *Y* by *W* depends on *Z* and is defined by the function

$$\theta_{XW\rightarrow Y} = b_4 + b_7Z$$

So b_4 estimates *conditional* moderation of *X's* effect by *W* when *Z* = 0 and b_7 estimates how much the moderation of *X*'s effect by *W* changes as *Z* changes by one unit.

$$\hat{Y}_i = -0.4611 - 4.2911X_i + 0.6235W_i + 4.5829Z_i + 0.2767X_iW_i - 13.9269X_iZ_i - 0.2021W_iZ_i + 0.9616X_iW_iZ_i$$

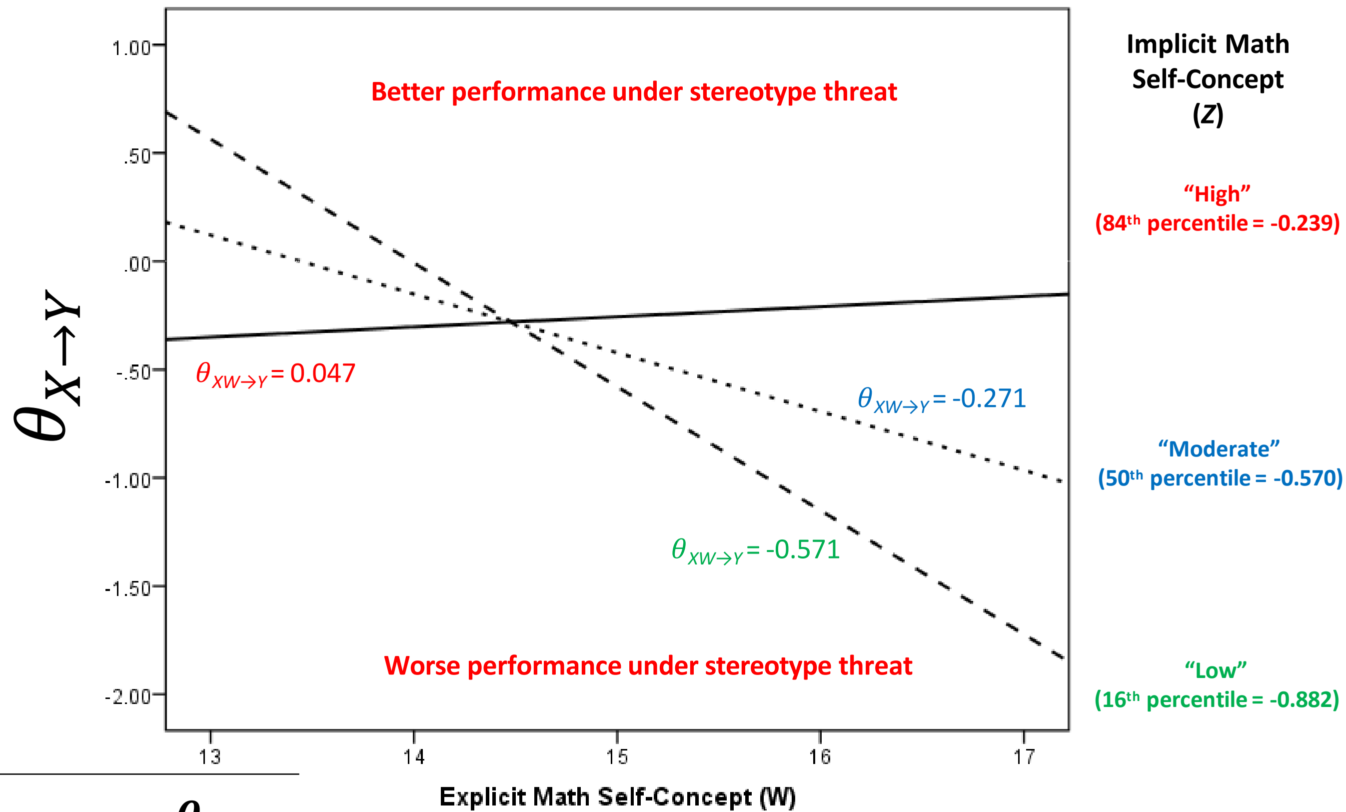
$$\theta_{XW\rightarrow Y} = b_4 + b_7Z$$

$$\theta_{XW\rightarrow Y} = 0.277 + 0.962Z$$

Z	$\boldsymbol{\theta_{XW\rightarrow Y}}$
-0.882	**-0.571**
-0.570	**-0.271**
-0.239	**0.047**

A visual representation

$$\theta_{X\to Y} = b_1 + b_4W + b_5Z + b_7WZ = -4.291 + 0.277W - 13.927Z + 0.962WZ$$



Z	$\theta_{XW\to Y}$
-0.882	-0.571
-0.570	-0.271
-0.239	0.047

These are the slopes of these lines linking *W* to the effect of *X* on *Y*

Implementation using regression centering

$$\hat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3Z_i + b_4X_iW_i + b_5X_iZ_i + b_6W_iZ_i + b_7X_iW_iZ_i$$

$$\hat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3Z_i + (b_4 + b_7Z_i)X_i\,W_i + b_5X_iZ_i + b_6W_iZ_i$$

b_4 estimates the conditional interaction between *X* and *W* when *Z* = 0. We can reparameterize the model such that b_4 estimates the *XW* interaction at any desired value *Z* = *z* by centering *Z* around z, recomputing products involving *Z*, and reestimating the model. That is, we will estimate

$$\hat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3(Z_i - z) + (b_4 + b_7(Z_i - z))X_i\,W_i + b_5X_i(Z_i - z) + b_6W_i(Z_i - z)$$

by creating a new variable *Z*′ defined as

$$Z_i' = Z_i - z$$

and then estimating

$$\hat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3Z'_i + (b_4 + b_7Z'_i)X_i\,W_i + b_5X_iZ'_i + b_6W_iZ'_i$$

In this model, b_4 estimates the conditional interaction between *X* and *W* when *Z*′ = 0, but *Z*′ = 0 when *Z* = *z*, which is what we want. We also get a test of significance for the *XW* interaction when *Z* = *z*.

Threat × explicit self-concept w/implicit math self-concept “relatively high’

center implicit self concept (*Z*) around -0.239 (relatively high).

```
compute implms_p = implms-(-0.239).
compute thrtim_p = threat*implms_p.
compute expimp_p = explms*implms_p.
compute thrtexpl = threat*explms.
compute thexim_p=threat*explms*implms_p.
regression/dep = mathprob/method = enter threat explms implms_p thrtexpl thrtim_p expimp_p thexim_p.
```

```
data math;set math;implms_p = implms-(-0.239);thrtim_p = threat*implms_p;expimp_p = explms*implms_p;
thrtexpl = threat*explms;thexim_p=threat*explms*implms_p;run;
proc reg data=math;model mathprob = threat explms implms_p thrtexpl thrtim_p expimp_p thexim_p;run;
```

```
implms_p<-math$implms-(-0.239)
summary(lm(mathprob~threat+explms+implms_p+threat*explms+threat*implms_p+explms*implms_p+
  threat*explms*implms_p,data=math))
```

```
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            -1.55642    3.21280  -0.484  0.62890
threat                 -0.96258    4.20522  -0.229  0.81931
explms                  0.67183    0.20533   3.272  0.00137 **
implms_p                4.58286    6.13199   0.747  0.45621
threat:explms           0.04686    0.26054   0.180  0.85756
threat:implms_p       -13.92695    7.46331  -1.866  0.06432 .
explms:implms_p        -0.20212    0.40287  -0.502  0.61675
threat:explms:implms_p  0.96158    0.47223   2.036  0.04379 *
```

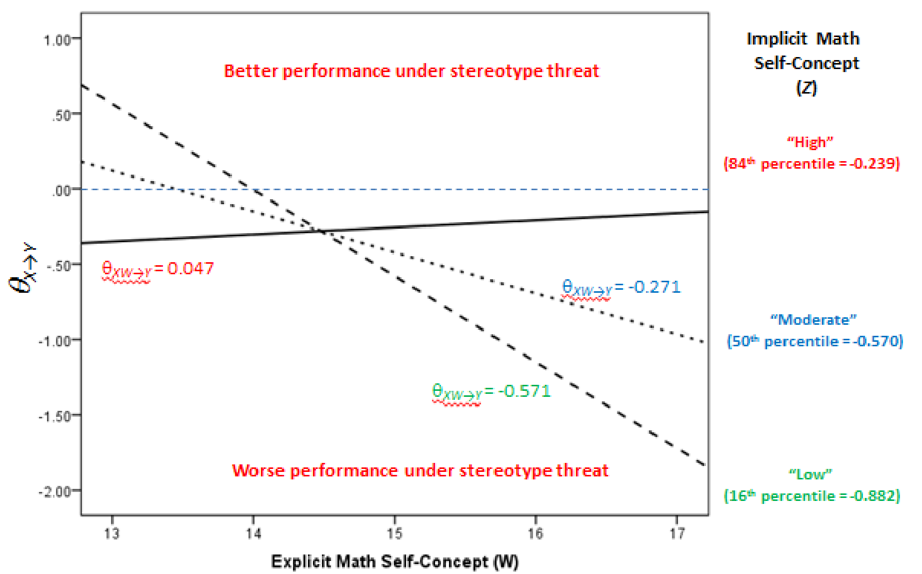
The two-way interaction between stereotype threat and explicit math self-concept is not statistically significant among those “relatively high” in implicit math self-concept (IMPLMS = -0.239). **Repeat for other values of implicit math self-concept. Or just look at the PROCESS output.**

$$\theta_{XW\rightarrow Y} = b_4 + b_7 Z = 0.277 + 0.962(\,-0.239) \;= 0.047$$

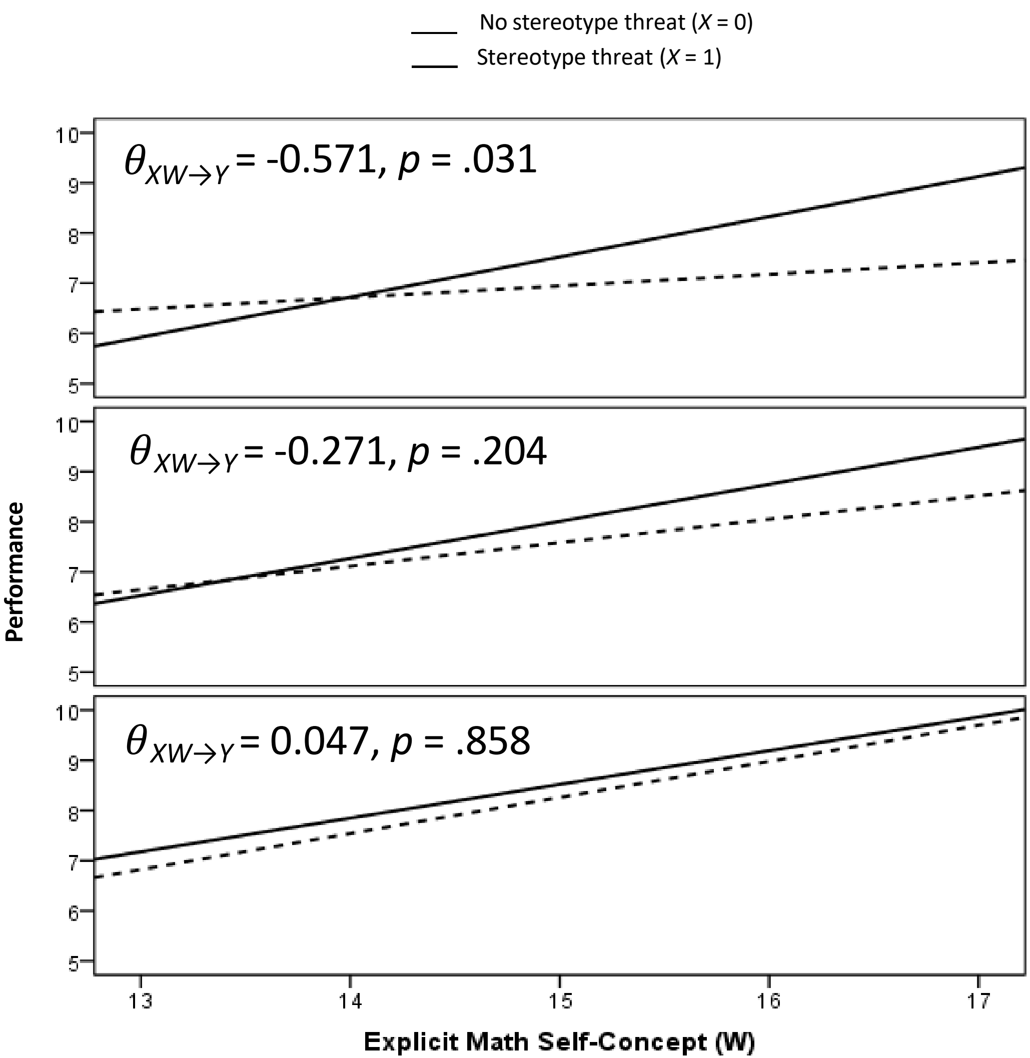
PROCESS does this automatically

The regression centering approach is tedious, though not difficult. PROCESS does this for you automatically for various values of the secondary moderator. Seeing that the secondary moderator (*Z*) is continuous, it selects values of *Z* corresponding to the 16th, 50th, and 84th percentiles of the distribution of *Z* and estimates the interaction between *X* and *W* at these values of *Z*. Use the **moments** or **zmodval** option if you prefer.

```
Test of conditional X*W interaction at value(s) of Z:
      implms       Effect           F        df1        df2          p
      -.8816       -.5711      4.7611     1.0000   128.0000      .0309
      -.5700       -.2714      1.6300     1.0000   128.0000      .2040
      -.2392        .0467       .0321     1.0000   128.0000      .8581
```



The moderation is moderated



(Z)

Implicit Math Self-Concept

"Low"

Explicit math self-concept × stereotype threat is statistically significant

"Moderate"

Explicit math self-Concept × stereotype threat is NOT statistically significant

"High"

Explicit math self-concept × stereotype threat is NOT statistically significant

The Johnson-Neyman Technique

The Johnson-Neyman technique can be applied to find the value or values of the secondary moderator (*Z*), if any, that define the points of transition for the statistical significance of the moderation of *X*'s effect by the primary moderator (*W*). PROCESS does this if you ask.

$$\theta_{XW \to Y} = b_4 + b_7 Z$$

$$\theta_{XW \to Y} = 0.277 + 0.962Z$$

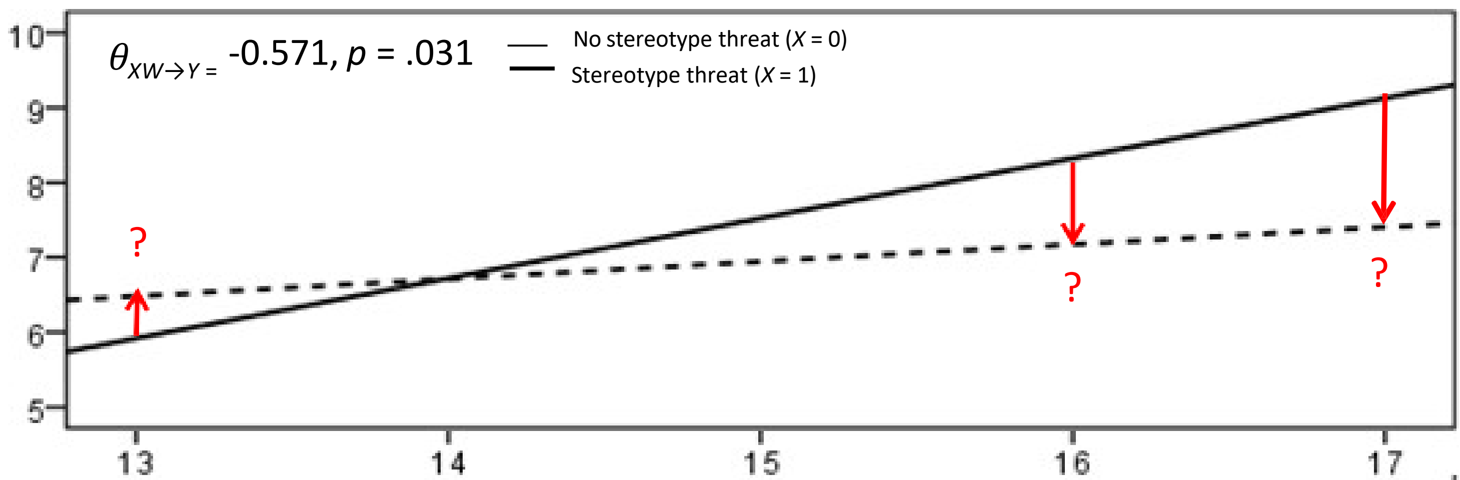
Statistically significant moderation of the effect of stereotype threat on performance by explicit math self-concept among those below -0.772 (the 24th percentile) in implicit math self-concept.

```
Moderator value(s) defining Johnson-Neyman significance region(s):
     Value    % below    % above
    -.7718    24.2647    75.7353

Conditional X*W interaction at values of the moderator Z:
    implms     Effect         se          t          p       LLCI       ULCI
   -1.6500    -1.3099      .5577    -2.3489      .0204    -2.4134     -.2065
   -1.5140    -1.1792      .4989    -2.3635      .0196    -2.1663     -.1920
   -1.3780    -1.0484      .4416    -2.3738      .0191    -1.9223     -.1745
   -1.2420     -.9176      .3866    -2.3736      .0191    -1.6825     -.1527
   -1.1060     -.7868      .3348    -2.3501      .0203    -1.4493     -.1244
    -.9700     -.6561      .2881    -2.2774      .0244    -1.2261     -.0860
    -.8340     -.5253      .2493    -2.1073      .0370    -1.0185     -.0321
    -.7718     -.4655      .2353    -1.9787      .0500     -.9310      .0000
    -.6980     -.3945      .2225    -1.7727      .0787     -.8349      .0458
    -.5620     -.2637      .2125    -1.2409      .2169     -.6843      .1568
    -.4260     -.1330      .2215     -.6003      .5494     -.5712      .3053
    -.2900     -.0022      .2474     -.0088      .9930     -.4917      .4873
    -.1540      .1286      .2856      .4502      .6533     -.4366      .6938
    -.0180      .2594      .3320      .7812      .4361     -.3976      .9163
     .1180      .3901      .3836     1.0171      .3110     -.3688     1.1491
     .2540      .5209      .4385     1.1880      .2370     -.3467     1.3885
     .3900      .6517      .4956     1.3149      .1909     -.3290     1.6324
     .5260      .7825      .5543     1.4116      .1605     -.3144     1.8793
     .6620      .9132      .6141     1.4870      .1395     -.3019     2.1284
     .7980     1.0440      .6748     1.5472      .1243     -.2911     2.3791
     .9340     1.1748      .7360     1.5962      .1129     -.2815     2.6311
    1.0700     1.3056      .7977     1.6367      .1042     -.2728     2.8839
```

Probing a conditional two way interaction

With evidence of moderation of the effect of stereotype threat on performance by explicit math self-concept among those low in implicit math self-concept, we might then probe this conditional moderation of *X*'s effect on *Y*. We'll use the pick-a-point approach.

Implicit Math Self Concept (*Z*) "Relatively Low" (-0.822)



13.0 (16th percentile) "Relatively low"

16.0 (50th percentile) "Moderate"

17 (84th percentile) "Relatively high"

Explicit math self-concept (*W*)

Probing a conditional two way interaction

A sensible strategy for probing a conditional two way interaction between *X* and *W* is to estimate the effect of the focal predictor (*X*) at various values of the primary moderator (*W*), setting the secondary moderator (*Z*) to the value at which the interaction is conditioned.

$$\widehat{Y_i} = b_0 + b_1 X_i + b_2 W_i + b_3 Z_i + b_4 X_i W_i + b_5 X_i Z_i + b_6 W_i Z_i + b_7 X_i W_i Z_i$$

can be rewritten as

$$\widehat{Y_i} = b_0 + (b_1 + b_4 W_i + b_5 Z_i + b_7 W_i Z_i) X_i + b_2 W_i + b_3 Z_i + b_6 W_i Z_i$$

which illustrates that the effect of *X* is conditional on *W* and *Z* and defined by the function

$$\theta_{X \rightarrow Y} = b_1 + b_4 W + b_5 Z + b_7 WZ$$

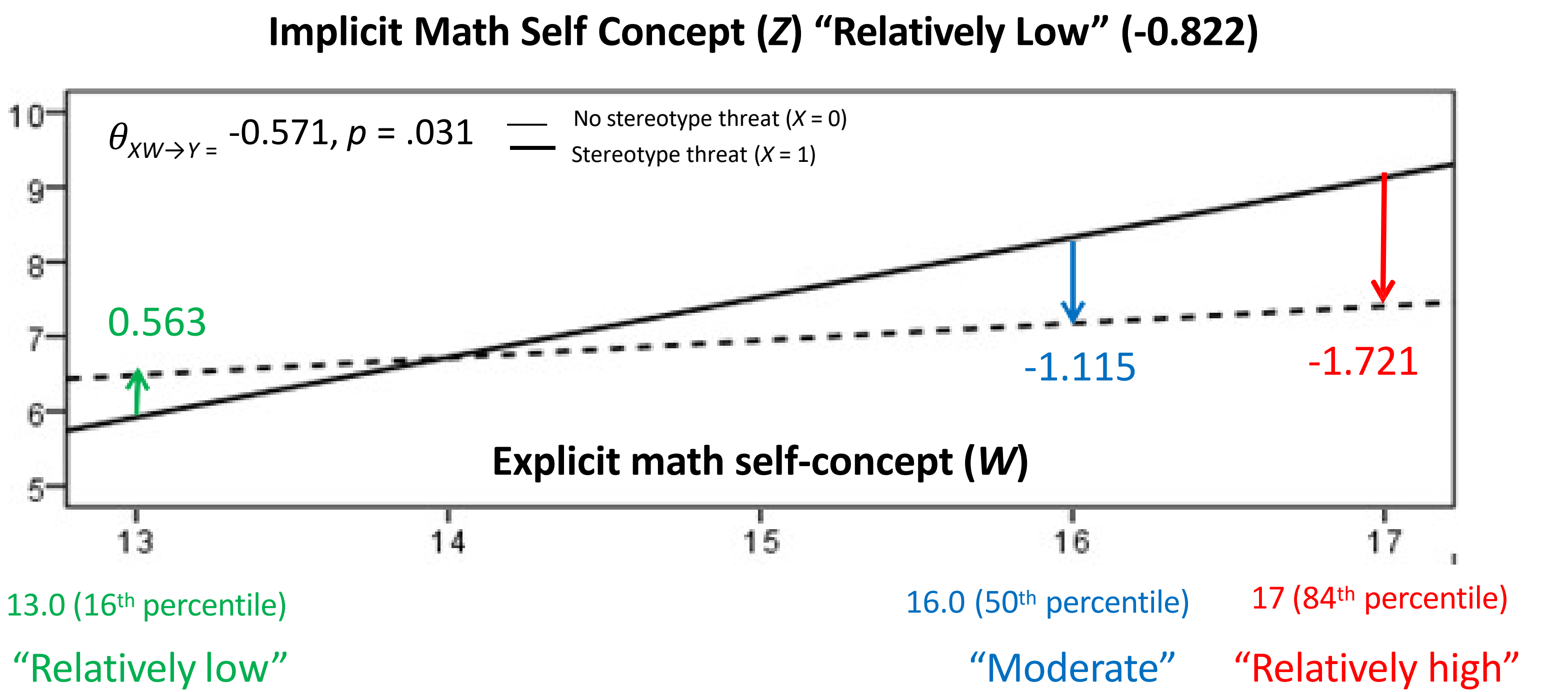
So b_1 estimates the conditional effect of *X* on *Y* when *W* and *Z* are both zero.

$$\widehat{Y_i} = -0.4611 - 4.2911 X_i + 0.6235 W_i + 4.5829 Z_i + 0.2767 X_i W_i - 13.9269 X_i Z_i - 0.2021 W_i Z_i + 0.9616 X_i W_i Z_i$$

$$\theta_{X \rightarrow Y} = b_1 + b_4 W + b_5 Z + b_7 WZ$$
$$\theta_{X \rightarrow Y} = -4.291 + 0.277W - 13.927Z + 0.962WZ$$

W	***Z***	$\boldsymbol{\theta_{X \rightarrow Y}}$
13	**-0.882**	**0.563**
16	**-0.882**	**-1.150**
17	**-0.882**	**-1.721**

Probing a conditional two way interaction



W	Z	$\theta_{X \to Y}$
13	-0.882	0.563
16	-0.882	-1.115
17	-0.882	-1.721

Probing the conditional two way interaction using the regression centering approach

$$\hat{Y}_i = b_0 + b_1X_i + b_2W_i + b_3Z_i + b_4X_iW_i + b_5X_iZ_i + b_6W_iZ_i + b_7X_iW_iZ_i$$

$$= b_0 + (b_1 + b_4W_i + b_5Z_i + b_7W_iZ_i)X_i + b_2W_i + b_3Z_i + b_6W_iZ_i$$

b_1 estimates the *conditional effect* of *X* on *Y* when *W* and *Z* are both zero. We can reparameterize the model such that b_1 estimates the *conditional effect of X* at any desired values *W* = *w and Z* = *z* by centering *W* around *w* and *Z* around *z*, recomputing products involving *W* and *Z*, and reestimating the model. That is, we will estimate

$$\hat{Y}_i = b_0 + b_1X_i + b_2(W_i - w) + b_3(Z_i - z) + b_4X_i(W_i - w) + b_5X_i(Z_i - z) +$$
$$b_6(W_i - w)(Z_i - z) + b_7X_i(W_i - w)(Z_i - z)$$

by creating new variables *W′* and *Z′* defined as

$$W_i' = W_i - w \qquad\qquad Z_i' = Z_i - z$$

and then estimating

$$\hat{Y}_i = b_0 + b_1X_i + b_2W'_i + b_3Z'_i + b_4X_iW'_i + b_5X_iZ'_i + b_6W'_iZ'_i + b_7X_iW'_iZ'_i$$

In this model, b_1 estimates the conditional effect of *X* on *Y* when *W′* = 0 and *Z′* = 0, but *W′* = 0 and *Z′* = 0 when *W* = *w* and *Z* = *z*, which is what we want. We also get a test of significance for the conditional effect of *X* on *Y* when *W* = *w* and *Z* = *z*.

The conditional effect of stereotype threat for high explicit and low implicit math self-concept

Let's test the effect of stereotype threat among those with a **fragile** math self-concept: Those high in explicit but low in implicit math self-concept.

```
compute explms_p  =  explms-17.
compute implms_p = implms-(-0.882).
compute thrtim_p = threat*implms_p.
compute expimp_p = explms_p*implms_p.
compute thrtexpl = threat*explms_p.
compute thexim_p=threat*explms_p*implms_p.
regression/dep = mathprob/method = enter threat explms_p implms_p thrtexpl thrtim_p expimp_p thexim_p.
```

center implicit math self-concept around -0.882 (relatively low) and explicit math self-concept around 17 (relatively high).

```
data math;set math;explms_p = explms-17;implms_p = implms-(-.882);
thrtim_p=threat*implms_p;expimp_p=explms_p*implms_p;thrtexpl=threat*explms_p;thexim_p=threat*explms_p*implms_p;run;
proc reg data=math;model mathprob=threat explms_p implms_p thrtexpl thrtim_p expimp_p thexim_p;run;
```

```
explms_p<-math$explms-17;implms_p<-math$implms-(-0.882);
summary(lm(mathprob~threat+explms_p+implms_p+threat*explms_p+threat*implms_p+explms_p*implms_p+
   threat*explms_p*implms_p,data=math))
```

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                9.1273     0.5826  15.665  < 2e-16 ***
threat                    -1.7220     0.7162  -2.404 0.017637 *
explms_p                   0.8018     0.2113   3.795 0.000227 ***
implms_p                   1.1469     1.0843   1.058 0.292179
threat:explms_p           -0.5714     0.2618  -2.183 0.030895 *
threat:implms_p            2.4199     1.3049   1.854 0.065970 .
explms_p:implms_p         -0.2021     0.4029  -0.502 0.616747
threat:explms_p:implms_p   0.9616     0.4722   2.036 0.043787 *
```

Stereotype threat significantly reduces performance among those with a fragile mathematical self-concept, $t(130)=-2.404$, $p = .018$. **Repeat for other values of explicit math self-concept. Or just look at PROCESS output.**

$$\theta_{X\to Y} = b_1 + b_4W + b_5Z + b_7WZ = -4.291 + 0.277W - 13.927Z + 0.962WZ$$
$$= -4.291 + 0.277(17) - 13.927(-0.882) + 0.962(17)(-0.882) = -1.722$$

PROCESS does this automatically

The regression centering approach is tedious, though not difficult. PROCESS does this for you automatically for various combinations of the primary and secondary moderator. Seeing that both the primary (*W*) and secondary (*Z*) moderators are continuous, it selects values of *W* and *Z* corresponding to the 16th, 50th, and 84th percentiles of the distributions of *W* and *Z* estimates the conditional effect of *X* on *Y* at the 9 combinations of the moderator values.

Conditional effects of the focal predictor at values of the moderator(s):

explms	implms	Effect	se	t	p	LLCI	ULCI
13.0000	-.8816	.5632	.7769	.7249	.4698	-.9740	2.1003
13.0000	-.5700	.1187	.6931	.1712	.8643	-1.2528	1.4902
13.0000	-.2392	-.3532	.9734	-.3628	.7173	-2.2791	1.5728
16.0000	-.8816	-1.1500	.5743	-2.0026	.0473	-2.2863	-.0137
16.0000	-.5700	-.6956	.4489	-1.5497	.1237	-1.5837	.1925
16.0000	-.2392	-.2132	.5864	-.3635	.7168	-1.3735	.9471
17.0000	-.8816	-1.7211	.7159	-2.4041	.0176	-3.1376	-.3045
17.0000	-.5700	-.9670	.5378	-1.7980	.0745	-2.0312	.0972
17.0000	-.2392	-.1665	.6435	-.2588	.7962	-1.4398	1.1068

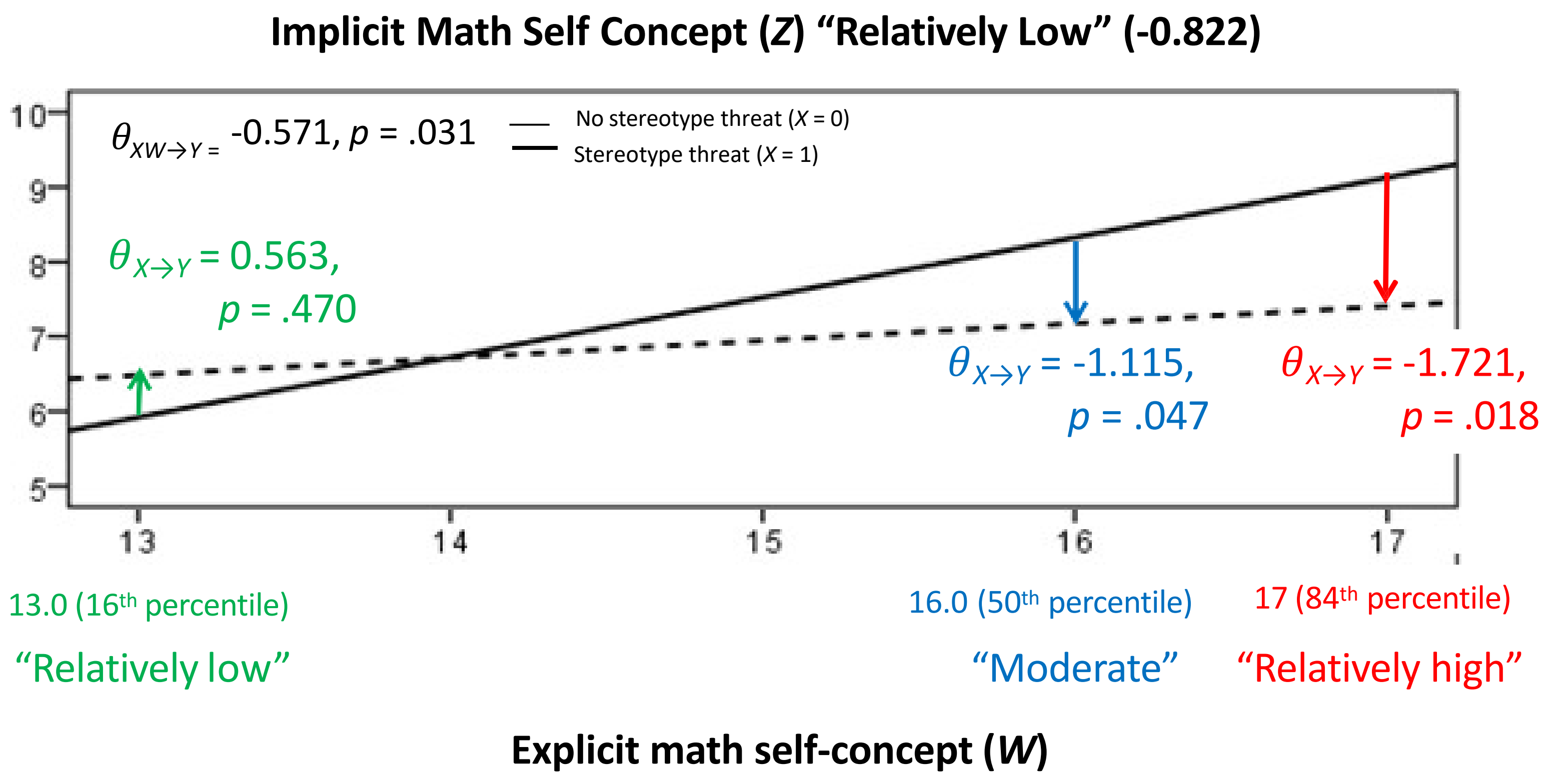
↑

$$\theta_{X\to Y} = b_1 + b_4 W + b_5 Z + b_7 WZ$$

$$\theta_{X\to Y} = -4.291 + 0.277W - 13.927Z + 0.962WZ$$

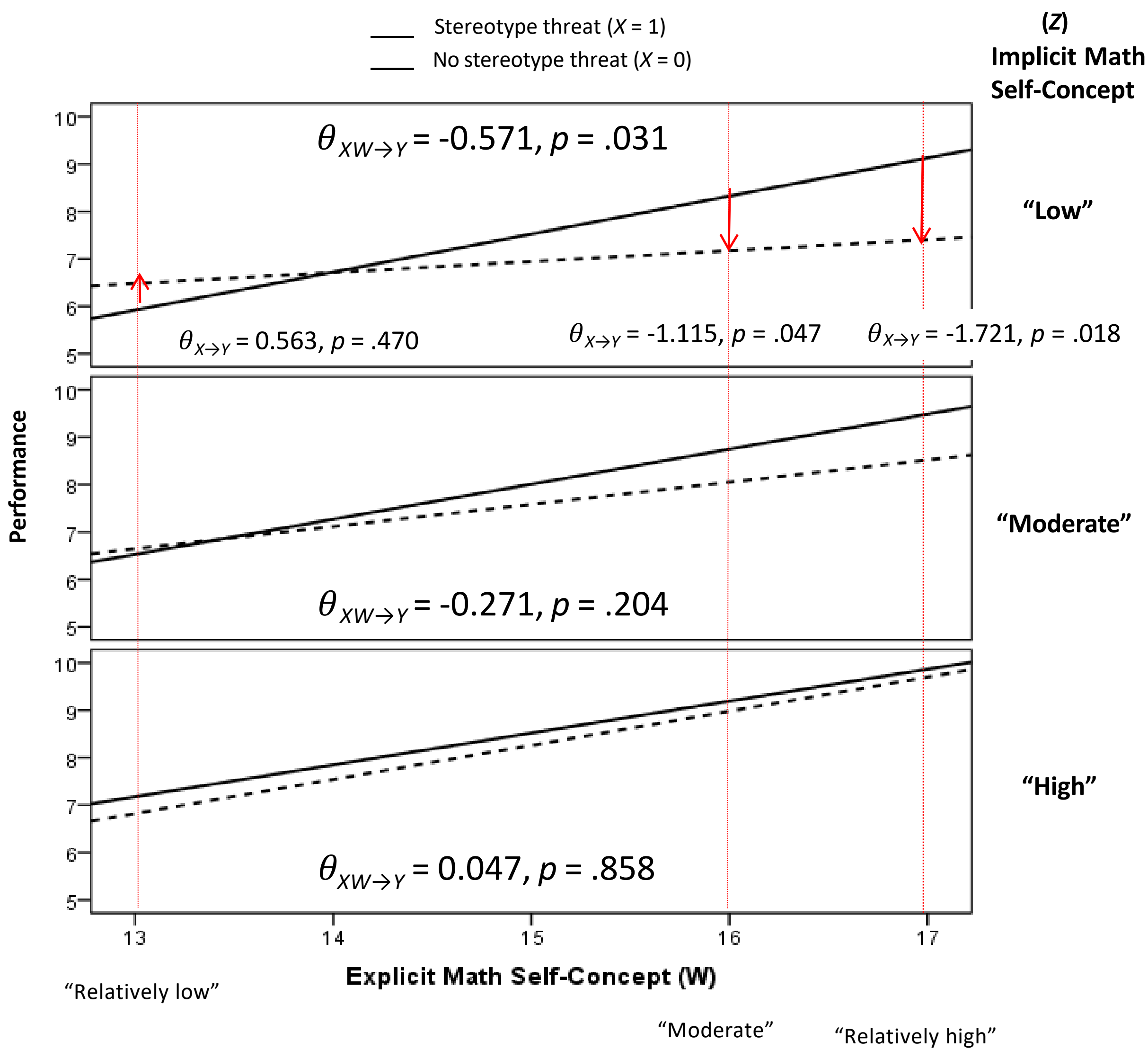
You can override these defaults using the **moments** option or the **wmodval** and **zmodval** options.

Probing the conditional two way interaction



Among those low implicit mathematical self-concept, stereotype threat significantly reduced performance only among those moderate to high in explicit math self-concept.

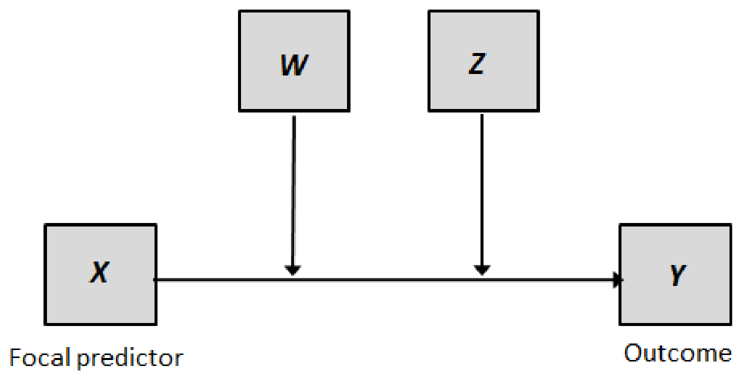
Putting it all together



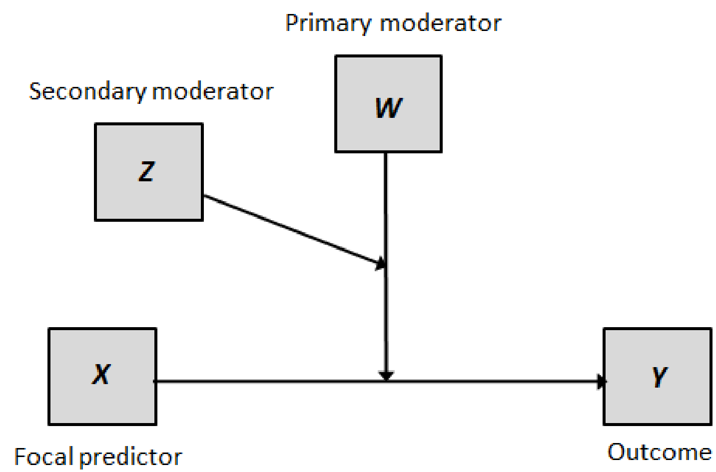
A statistically significant three way interaction between stereotype threat, explicit math self-concept, and implicit math self-concept reflects the finding that mathematical self-concept moderates the effect of stereotype threat on performance only among those relatively low in implicit mathematical self-concept. And stereotype threat reduced performance only among those with a "fragile" mathematical self-concept—those who **said** they consider themselves mathematical but *who don't really feel that way* (as revealed by a relatively lower score on the implicit math self-concept measure).

Choosing between additive and multiplicative multiple moderation

Additive multiple moderator model



Multiplicative multiple moderator model ("moderated moderation" model)

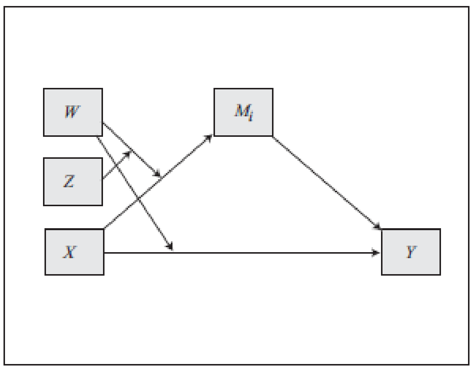


- Both models allow *X*'s effect on *Y* to depend on both *W* and *Z*
- The additive model imposes the constraint that *W*'s effect on *X*'s effect on *Y* is independent of *Z*.
- The moderated moderation model allows *W*'s effect on *X*'s effect on *Y* to depend linearly on *Z*.
- Diagram your predictions, and think about your theory. Does it predict moderation of moderation, or just that both *W* and *Z* affect *X*'s effect on *Y*?

Constructing and editing models in PROCESS

Historically, PROCESS has operated by a model number system. The model numbers and the models those numbers represent can be found in the documentation. Choose the model number that corresponds to the model you would like to estimate.

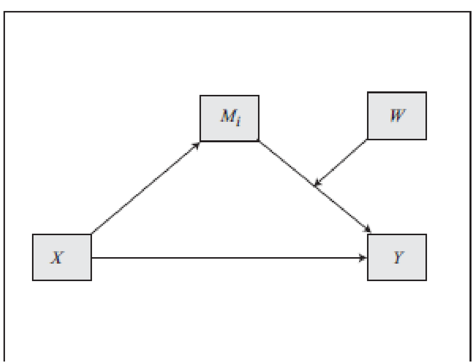
Model 13



Many of the preprogrammed numbered models you will find useful.

But what if the model you want to estimate does not correspond to any preprogrammed model represented by a model number?

Model 14



Version 2: Too bad. Nothing you can do about it (unless you know some tricks).

Version 3: Within certain constraints, you can create your own model from scratch, or edit an existing model number to make it correspond to the model you want to estimate.

The *B* matrix

The *B* matrix is the heart of the representation of a model in PROCESS. It is a matrix of 0s and 1s specifying whether (1) or not (0) the variable in the column sends an effect to the variable in the row.

Variables sending effects (i.e., arrow points away)

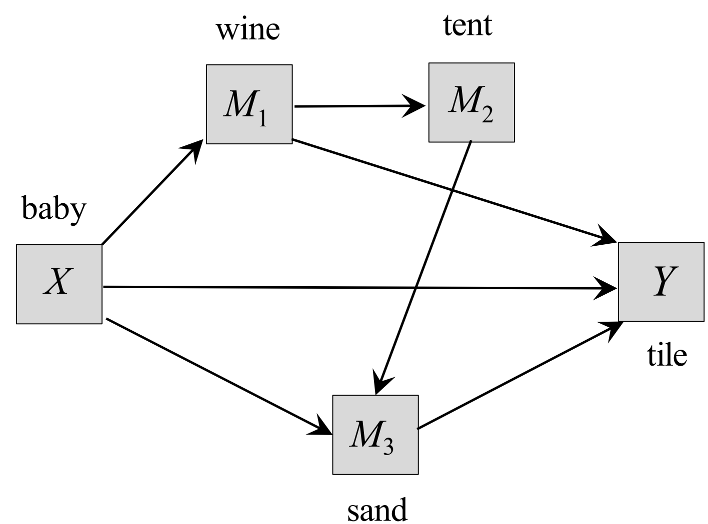
Variables receiving effects (an arrow points at)

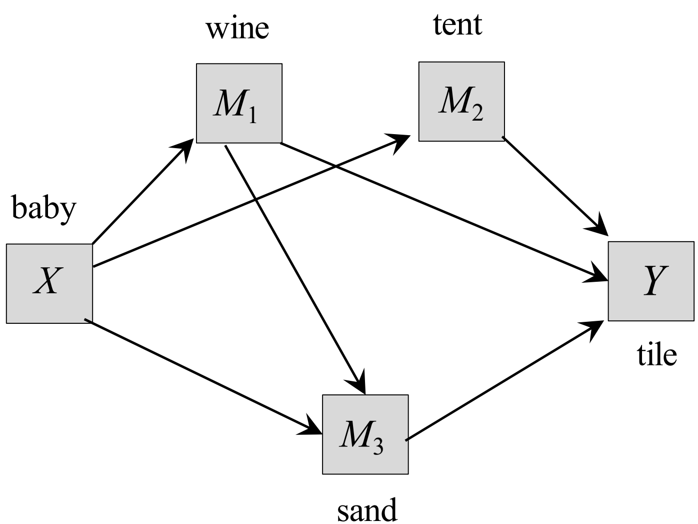
	X	M_1	M_2	$\ldots$	M_k
M_1	0/1	■	■	■	■
M_2	0/1	0/1	■	■	■
$\ldots$	0/1	0/1	0/1	■	■
M_k	0/1	0/1	0/1	0/1	■
Y	0/1	0/1	0/1	0/1	0/1

A model you program can contain one *X*, one *Y*, and up to 6 mediators. Certain cells are fixed to zero (the black squares above) to ensure the model is recursive (no feedback loops).

Examples

		X baby	M_1 wine	M_2 tent	M_3 sand
M_1	wine	1	■	■	■
M_2	tent	0	1	■	■
M_3	sand	1	0	1	■
Y	tile	1	1	0	1

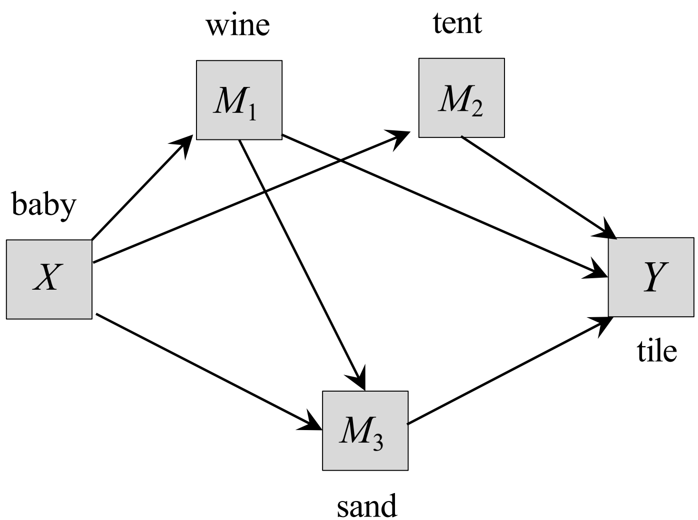




		X baby	M_1 wine	M_2 tent	M_3 sand
M_1	wine	1	■	■	■
M_2	tent	1	0	■	■
M_3	sand	1	1	0	■
Y	tile	0	1	1	1

Programming the *B* matrix

The mediation component of a model is programmed using the **bmatrix=** statement followed by a sequence of zeros and ones.



		X baby	M_1 wine	M_2 tent	M_3 sand
M_1	wine	1	■	■	■
M_2	tent	1	0	■	■
M_3	sand	1	1	0	■
Y	tile	0	1	1	1

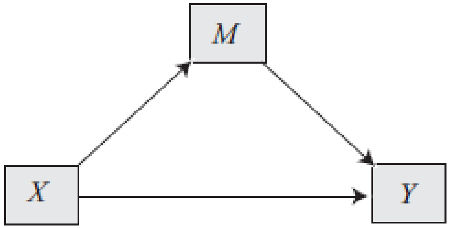
Read the *B* matrix from left to right, top to bottom, skipping the black squares, and enter the zeros and ones in the sequence as they are encountered in the B matrix. Separate with commas in SPSS and R, using the c() operator in R. No commas in SAS.

```
process y=tile/x=baby/m=wine tent sand/bmatrix=1,1,0,1,1,0,0,1,1,1.
```

```
%process (data=four,y=tile,x=baby,m=wine tent sand,bmatrix=1 1 0 1 1 0 0 1 1 1)
```

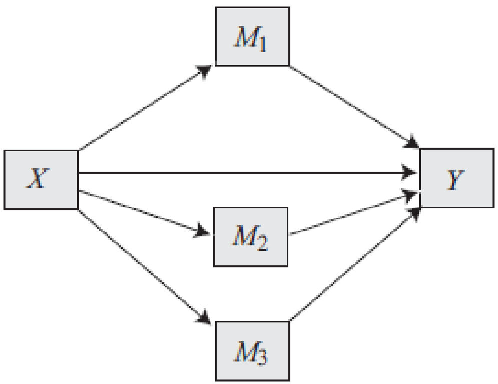
```
process(data=four,y="tile",x="baby",m=c("wine","tent","sand"),bmatrix=c(1,1,0,1,
  1,0,0,1,1,1))
```

Examples



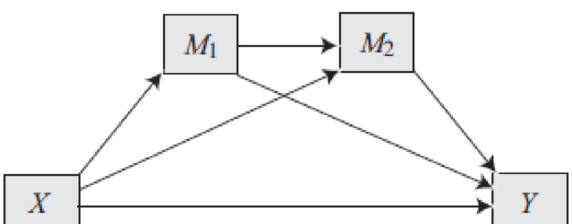
	X	M
M	1	■
Y	1	1

bmatrix=1,1,1



	X	M_1	M_2	M_3
M_1	1	■	■	■
M_2	1	0	■	■
M_3	1	0	0	■
Y	1	1	1	1

bmatrix=1,1,0,1,0,0,1,1,1,1



	X	M_1	M_2
M_1	1	■	■
M_2	1	1	■
Y	1	1	1

bmatrix=1,1,1,1,1,1

Some rules about programming models

- The model must be *recursive*, meaning no feedback loops or bidirectional cause.
- One *X*, one *Y*, and between 1 and 6 mediators.
- All variables must send at least one effect or receive at least one effect.
- All variables specified as mediators must both send and receive at least one effect (no "dangling mediators").
- Only zeros and ones in the **bmatrix=** statement.
- With k mediators, the **bmatrix=** statement must contain a sequence of $0.5(k + 1)(k + 2)$ zeros and ones.

Moderation: The *W* and *Z* matrices

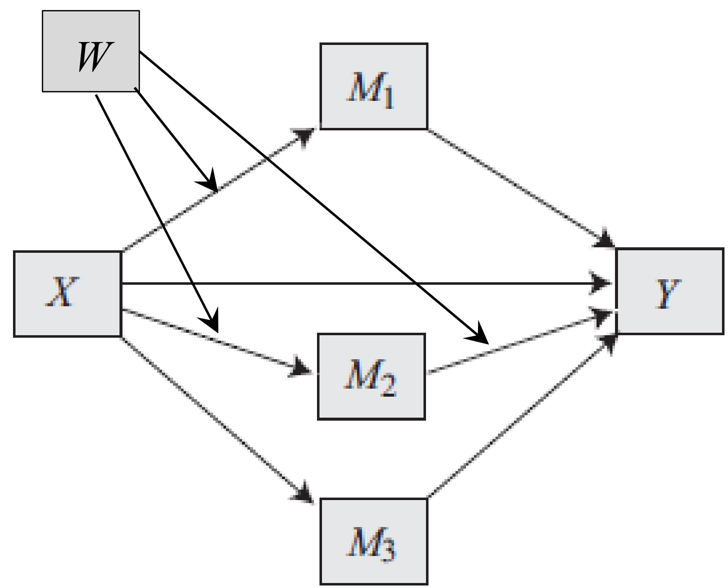
The *W* and *Z* matrices define which paths specified in the *B* matrix are moderated (1) and (0) not moderated. These matrices have the same form and size as the *B* matrix.

B matrix

	X	M_1	M_2	M_3
M_1	1	■	■	■
M_2	1	0	■	■
M_3	1	0	0	■
Y	1	1	1	1

W matrix

	X	M_1	M_2	M_3
M_1	1	■	■	■
M_2	1	0	■	■
M_3	0	0	0	■
Y	0	0	1	0



Note: A path fixed at zero in the *B* matrix cannot be moderated.

Moderation: The *W* and *Z* matrices

A second moderator *Z* can be specified. But *Z* can only be used if *W* has already been used. That is, if your model has only one moderator, it must be called *W*.

B matrix

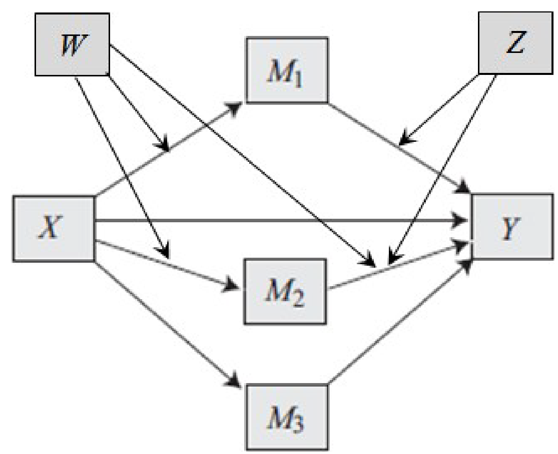
	X	M_1	M_2	M_3
M_1	1	■	■	■
M_2	1	0	■	■
M_3	1	0	0	■
Y	1	1	1	1

W matrix

	X	M_1	M_2	M_3
M_1	1	■	■	■
M_2	1	0	■	■
M_3	0	0	0	■
Y	0	0	1	0

Z matrix

	X	M_1	M_2	M_3
M_1	0	■	■	■
M_2	0	0	■	■
M_3	0	0	0	■
Y	0	1	1	0



Programming the *W* and *Z* matrices

B matrix

	X	M_1	M_2	M_3
M_1	1	■	■	■
M_2	1	0	■	■
M_3	1	0	0	■
Y	1	1	1	1

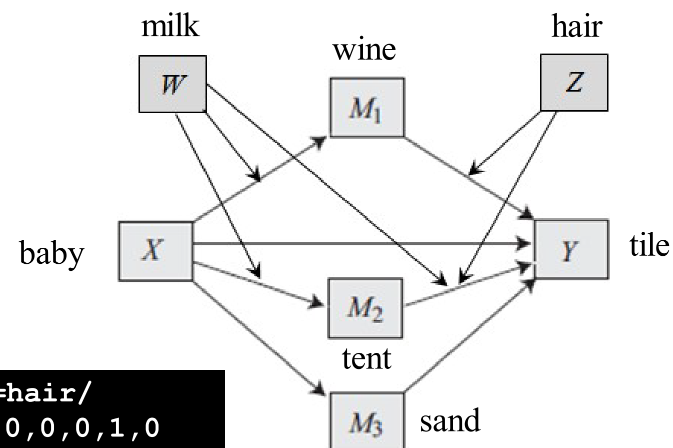
W matrix

	X	M_1	M_2	M_3
M_1	1	■	■	■
M_2	1	0	■	■
M_3	0	0	0	■
Y	0	0	1	0

Z matrix

	X	M_1	M_2	M_3
M_1	0	■	■	■
M_2	0	0	■	■
M_3	0	0	0	■
Y	0	1	1	0

The *W* and *Z* matrices are programmed with the **wmatrix=** and **zmatrix=** statements. Use the same procedure as for the *B* matrix.




```
process y=tile/x=baby/m=wine tent sand/w=milk/z=hair/
  bmatrix=1,1,0,1,0,0,1,1,1,1/wmatrix=1,1,0,0,0,0,0,0,1,0
  /zmatrix=0,0,0,0,0,0,0,1,1,0.
```

```
%process (data=four,y=tile,x=baby,m=wine tent sand,w=milk,z=hair,bmatrix=1 1 0 1 0 0 1 1 1 1,
   wmatrix=1 1 0 0 0 0 0 0 1 0,zmatrix=0 0 0 0 0 0 0 1 1 0)
```

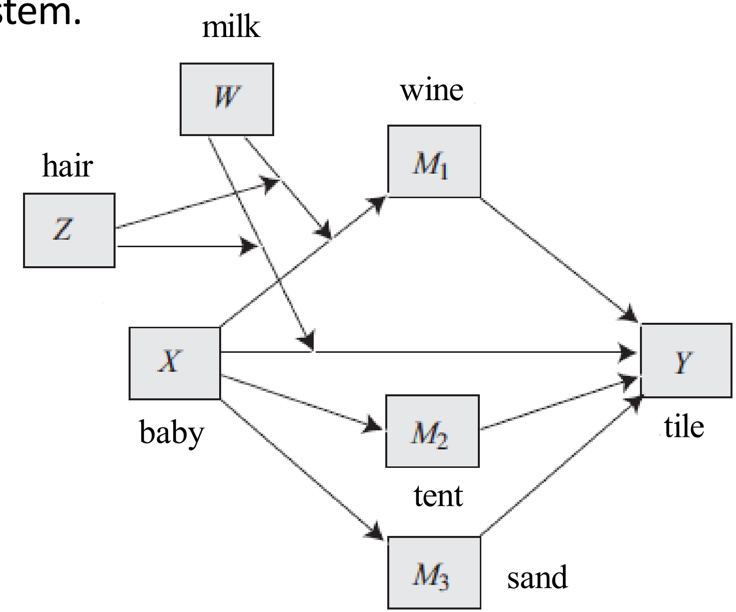
```
process  (data=four,y="tile",x="baby",m=c("wine","tent","sand"),w="milk",z="hair",
   bmatrix=c(1,1,0,1,0,0,1,1,1,1),wmatrix=c(1,1,0,0,0,0,0,0,1,0),
   zmatrix=c(0,0,0,0,0,0,0,1,1,0))
```

Moderated moderation: The *WZ* matrix

Moderated moderation, also known as three-way interaction, is held in the *WZ* matrix. *W* is the primary moderator, and *Z* is the secondary moderator. Program using the **wzmatrix=** statement, using the same 0/1 system.

WZ matrix

	X	M_1	M_2	M_3
M_1	1	■	■	■
M_2	0	0	■	■
M_3	0	0	0	■
Y	1	0	0	0




```
process y=tile/x=baby/m=wine tent sand/w=milk/z=hair/bmatrix=1,1,0,1,0,0,1,1,1,1
   /wzmatrix=1,0,0,0,0,0,1,0,0,0.
```

```
%process (data=four,y=tile,x=baby,m=wine tent sand,w=milk,z=hair,bmatrix=1 1 0 1 0 0 1 1 1 1,
   wzmatrix=1 0 0 0 0 0 1 0 0 0)
```

```
process(data="four",y="tile",x="baby",m=c("wine","tent","sand"),w="milk",z="hair",
   bmatrix=c(1,1,0,1,0,0,1,1,1,1),wzmatrix=c(1,0,0,0,0,0,1,0,0,0))
```

Note: if you don't use **wmatrix** or **zmatrix**, this implies all cells are zero

Putting it all together

This system allows for the construction of some very complex models.

B matrix

	X	M_1	M_2	M_3
M_1	1	■	■	■
M_2	1	0	■	■
M_3	1	0	0	■
Y	1	1	1	1

W matrix

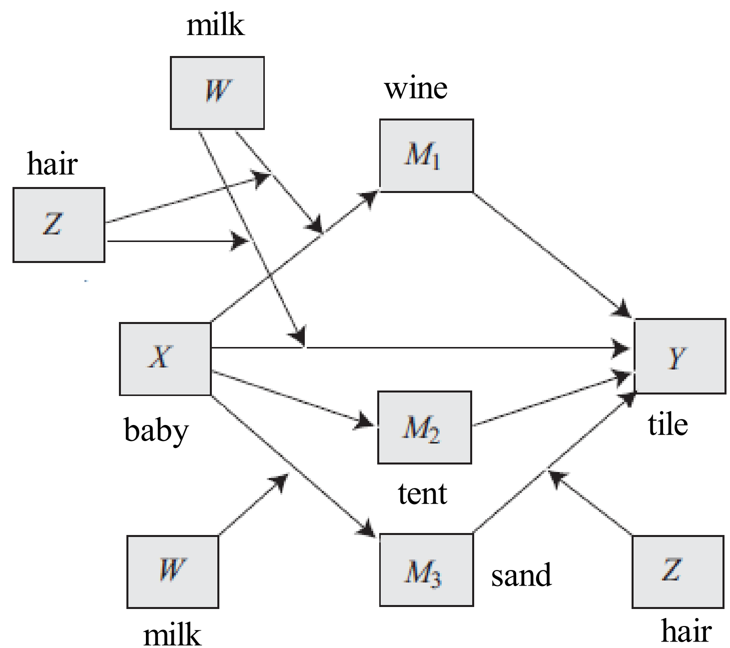
	X	M_1	M_2	M_3
M_1	0	■	■	■
M_2	0	0	■	■
M_3	1	0	0	■
Y	0	0	0	0

Z matrix

	X	M_1	M_2	M_3
M_1	0	■	■	■
M_2	0	0	■	■
M_3	0	0	0	■
Y	0	0	0	1

WZ matrix

	X	M_1	M_2	M_3
M_1	1	■	■	■
M_2	0	0	■	■
M_3	0	0	0	■
Y	1	0	0	0




```
process y=tile/x=baby/m=wine tent sand/w=milk/z=hair/bmatrix=1,1,0,1,0,0,1,1,1,1
  /wmatrix=0,0,0,1,0,0,0,0,0,0/zmatrix=0,0,0,0,0,0,0,0,0,1/wzmatrix=1,0,0,0,0,0,1,0,0,0.
```

```
%process (data=four,y=tile,x=baby,m=wine tent sand,w=milk,z=hair,bmatrix=1 1 0 1 0 0 1 1 1 1,
  wmatrix=0 0 0 1 0 0 0 0 0 0,zmatrix=0 0 0 0 0 0 0 0 0 1,wzmatrix=1 0 0 0 0 0 1 0 0 0)
```

```
process(data=four,y="tile",x="baby",m=c("wine","tent","sand"),w="milk",z="hair",
  bmatrix=c(1,1,0,1,0,0,1,1,1,1),wmatrix=c(0,0,0,1,0,0,0,0,0,0),zmatrix=c(0,0,0,0,0,0,
  0,0,0,1),wzmatrix=c(1,0,0,0,0,0,1,0,0,0))
```

Rules on the specification of moderation

- Up to two moderators, *W* and *Z*, can be specified. If only one is used, it must be *W*.

- A path fixed to zero in the *B* matrix cannot be moderated.

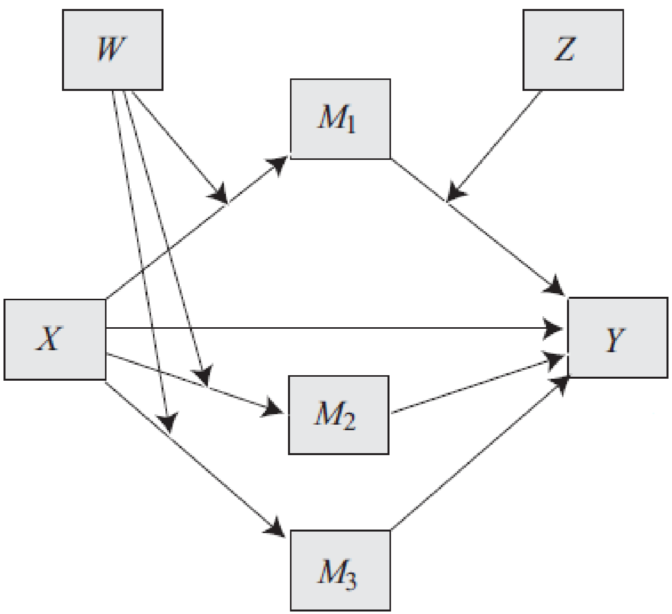
- Unless explicitly specified otherwise, all cells in the *W*, *Z*, and *WZ* matrices are automatically set to zero whenever a custom model is constructed using the **bmatrix** statement.

- Whenever a cell in the *WZ* matrix is set to 1, corresponding cells in the *W* and *Z* matrices are automatically set to 1. This is because three way interaction requires the inclusion of the two way interactions.

Editing a numbered model

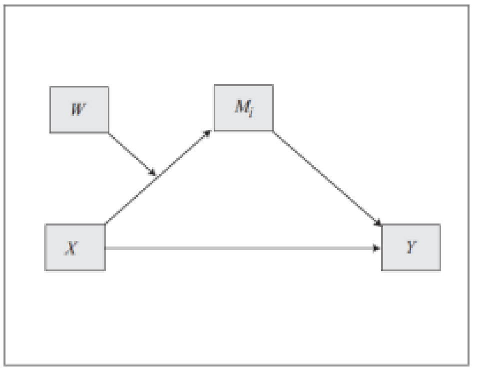
Many preprogrammed numbered models are likely to be close to the model you want to estimate. You can edit a modeled number, adding a desired interaction, or removing one you don't want. This is done by reprogramming the *W*, *Z*, and/or *WZ* matrices.

Example



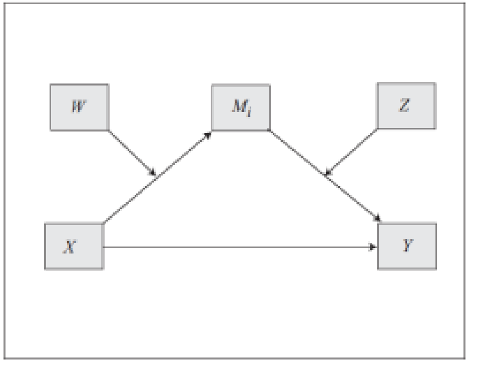
This is like model 7, except model 7 doesn't include moderation of any *M* to *Y* paths.

Model 7



This is like model 21, except model 21 would includes moderation by *Z* of the M_2 and M_3 to *Y* paths as well as the M_1 to *Y* path.

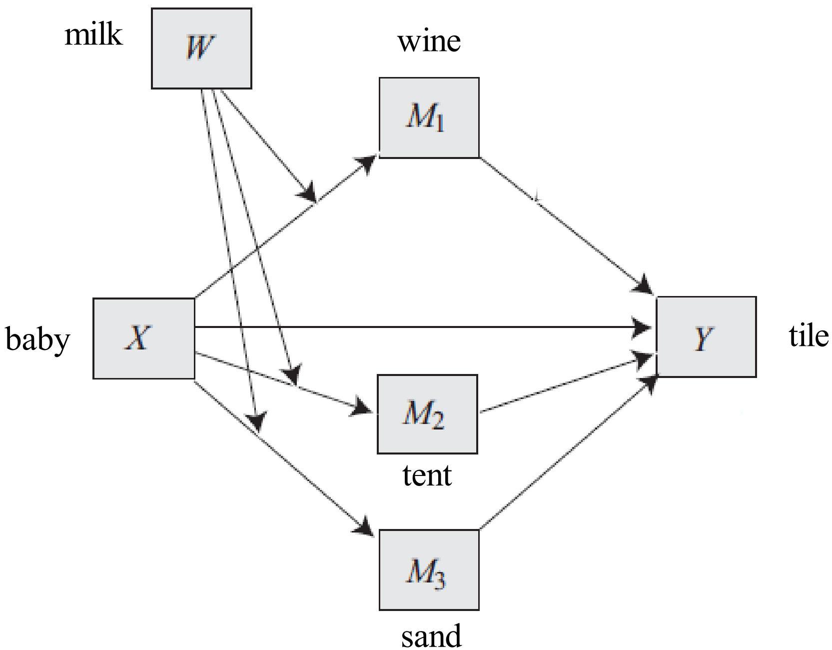
Model 21



NOTE: When using a model number, the *B* matrix cannot be edited.

Option 1: Edit model 7

In model 7, the *Z* matrix is all zeros because there is no *Z* in model 7. So program the *Z* matrix to include the moderation of the desired path by *Z*.



Preprogrammed Z matrix

	X	M_1	M_2	M_3
M_1	0	■	■	■
M_2	0	0	■	■
M_3	0	0	0	■
Y	0	0	0	0

Desired *Z* matrix

	X	M_1	M_2	M_3
M_1	0	■	■	■
M_2	0	0	■	■
M_3	0	0	0	■
Y	0	1	0	0

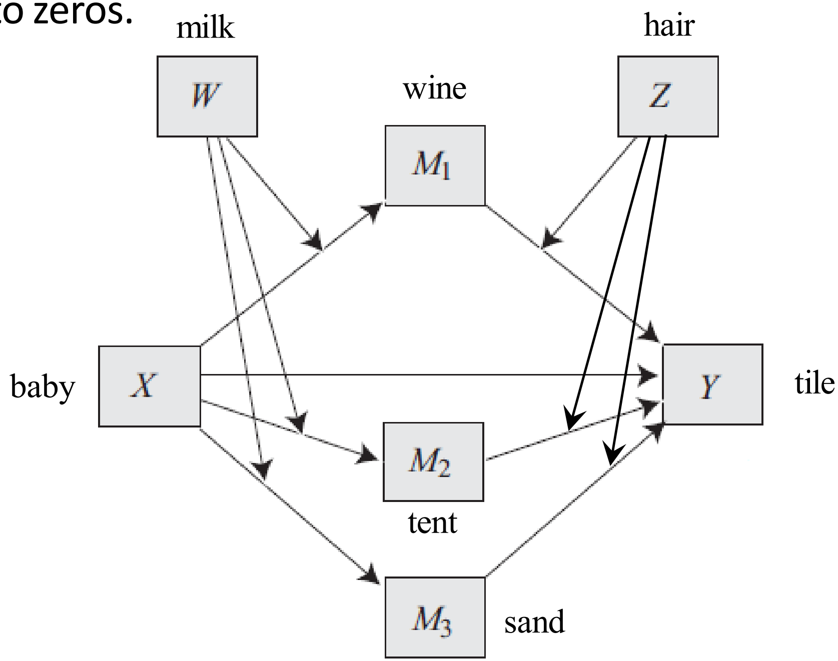
```
process y=tile/x=baby/m=wine tent sand/w=milk/z=hair/model=7/
    zmatrix=0,0,0,0,0,0,0,1,0,0.
```

```
%process (data=four,y=tile,x=baby,m=wine tent sand,w=milk,z=hair,model=7,
    zmatrix=0 0 0 0 0 0 0 1 0 0);
```

```
process (data=four,y="tile",x="baby",m=c("wine","tent","sand"),w="milk",
    z="hair",model=7,zmatrix=c(0,0,0,0,0,0,0,1,0,0))
```

Option 2: Edit model 21

In model 21, the Z matrix contains ones in certain cells that allow the M_2 and M_3 paths to Y to be moderated by Z. We can reprogram the Z matrix, turning the offensives ones into zeros.



Preprogrammed Z matrix

	X	M_1	M_2	M_3
M_1	0	■	■	■
M_2	0	0	■	■
M_3	0	0	0	■
Y	0	1	1	1

Desired Z matrix

	X	M_1	M_2	M_3
M_1	0	■	■	■
M_2	0	0	■	■
M_3	0	0	0	■
Y	0	1	0	0

```
process y=tile/x=baby/m=wine tent sand/w=milk/z=hair/model=21/
    zmatrix=0,0,0,0,0,0,0,1,0,0.
```

```
%process (data=four,y=tile,x=baby,m=wine tent sand,w=milk,z=hair,model=21,
    zmatrix=0 0 0 0 0 0 0 1 0 0);
```

```
process(data=four,y="tile",x="baby",m=c("wine","tent","sand"),w="milk",
    z="hair",model=21,zmatrix=c(0,0,0,0,0,0,0,1,0,0))
```

The MATRICES statement

If you want to check to make sure you have programmed the matrices correctly, or you want to see what the matrices of a preprogrammed model look like, add **matrices=1** to a PROCESS command.

```
process y=tile/x=baby/m=wine tent
  sand/w=milk/z=hair/model=7/
  zmatrix=0,0,0,0,0,0,0,1,0,0
  /matrices=1.
```

```
%process (data=four,y=tile,
  x=baby,m=wine tent sand,
  w=milk,z=hair,model=7,
  zmatrix=0 0 0 0 0 0 0 1 0 0,
  matrices=1)
```

```
process(data=four,y="tile",
 x="baby",m=c("wine","tent",
 "sand"),w="milk",z="hair",
 model=7,zmatrix=c(0,0,0,0,0,0,
 0,1,0,0),matrices=1)
```

Matrices that don't appear in the output have zeros in all cells. If your model includes covariates, the *C* matrix will appear here too.

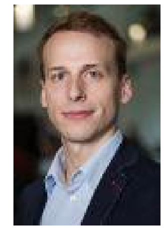
```
********** MODEL DEFINITION MATRICES **********

BMATRIX: Paths freely estimated (1) and fixed
to zero (0):
      baby wine tent sand
wine   1
tent   1    0
sand   1    0    0
tile   1    1    1    1

WMATRIX: Paths moderated (1) and not moderated
(0) by W:
      baby wine tent sand
wine   1
tent   1    0
sand   1    0    0
tile   0    0    0    0

ZMATRIX: Paths moderated (1) and not moderated
(0) by Z:
      baby wine tent sand
wine   0
tent   0    0
sand   0    0    0
tile   0    1    0    0
```

An excel tool for writing custom model PROCESS code

http://dariusfrank.com/

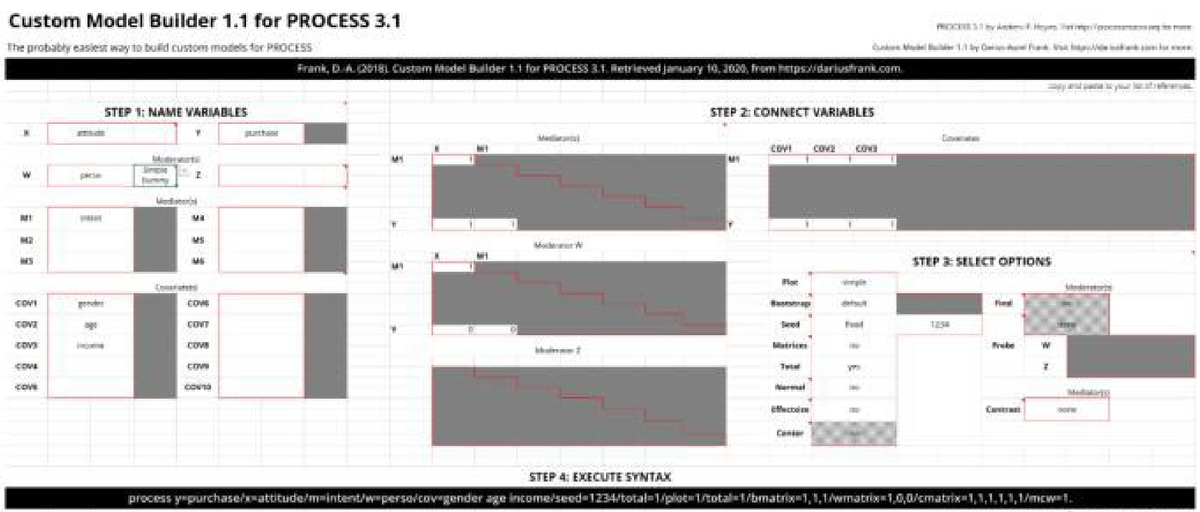
Custom Model Builder for PROCESS 3.0

by Darius-Aurel Frank



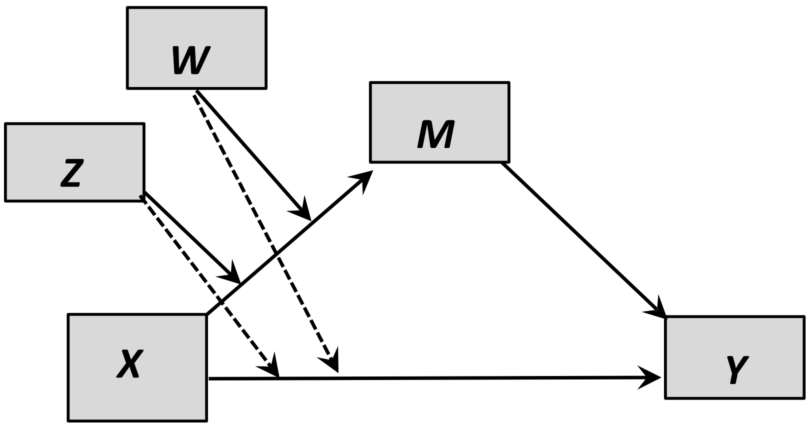
Preview

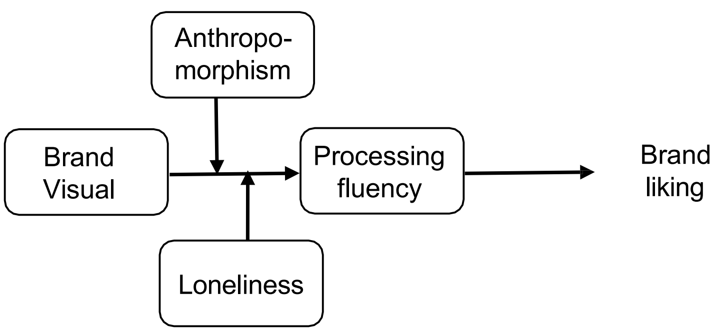
Custom Model Builder 1.1 for PROCESS



Conditional process analysis with two moderators of the same path

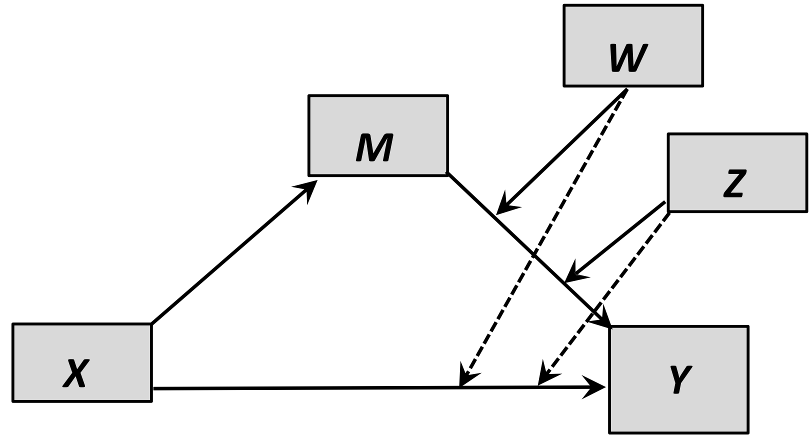
A first stage dual moderated mediation model

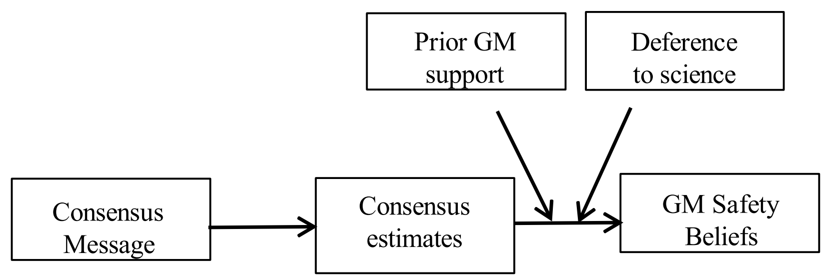




Orth, U. R., Cornwell, B., Ohlhoff, J., & Naber, C. (2017). Seeing faces: The role of brand visual processing and social connection in brand liking. ***European Journal of Social Psychology,*** *47*, 348-361.

A second stage dual moderated mediation model





Dixon, G. (2016). Applying the gateway belief model to genetically modified food perceptions: New insights and additional questions. ***Journal of Communication,*** *66*, 888-908.

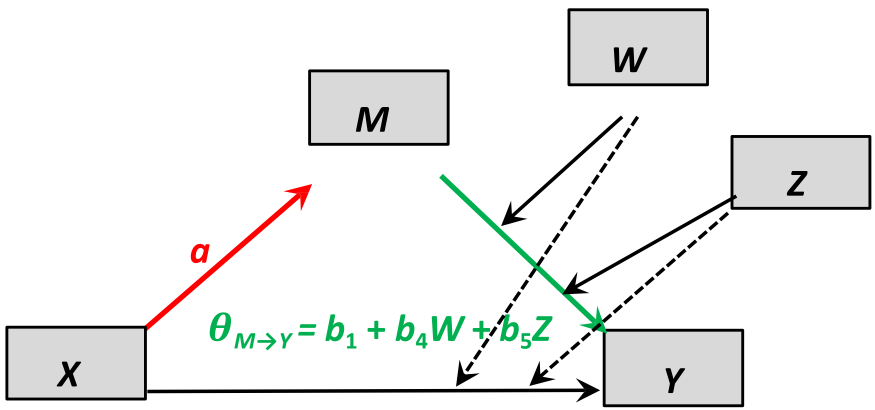
Where are models like this discussed?

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs, 85*, 5-40.

Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist, 64*, 19-54.

- Extends the "index of moderated mediation" concept to more complex models involving more than one moderator.
- Provides a means of quantifying and testing a hypothesis about whether an indirect effect is moderated when a second moderator is held fixed, and how much the moderation of the mediation of one variable's effect on another depends on a second moderator.
- Discusses visualizing indirect effects in complex multiple moderator models.

The indirect effect is an additive linear function of two moderators



Indirect effect of *X*

$a\theta_{M \to Y} =$

$a(b_1 + b_4W + b_5Z)$

or

$ab_1 + ab_4W + ab_5Z$

which is a linear function of both *W* and *Z*.

$$\widehat{M}_i = a_0 + \boxed{aX_i}$$

$$\widehat{Y}_i = c_0' + c'X_i + b_1M_i + b_2W_i + b_3Z_i + b_4M_iW_i + b_5M_iZ_i$$

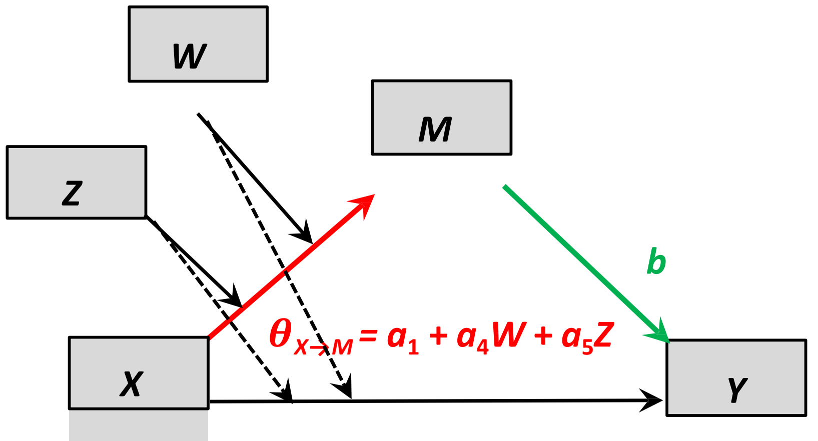
The model for *Y* can be written in equivalent form as

$$\widehat{Y}_i = c_0' + c'X_i + \boxed{(b_1 + b_4W_i + b_5Z_i)}M_i + b_2W_i + b_3Z_i$$

or $$\widehat{Y}_i = c_0' + c'X_i + \theta_{M \to Y}M_i + b_2W_i + b_3Z_i$$

where $\theta_{M \to Y} = b_1 + b_4W + b_5Z$ is the conditional effect of *M* on *Y*. It is a linear function of both *W* and *Z*.

The indirect effect is an additive linear function of two moderators



Indirect effect of X

$\theta_{X\rightarrow M}\, b$ =

$(a_1 + a_4W + a_5Z)b$

or

$a_1b + a_4bW + a_5bZ$

which is a linear function of both W and Z.

$$\widehat{M_i} = a_0 + a_1X_i + a_2W_i + a_3Z_i + a_4X_iW_i + a_5X_iZ_i$$

$$\widehat{Y_i} = c_0' + c'X_i + \boxed{bM_i}$$

The model for M can be written in equivalent form as

$$\widehat{M_i} = a_0 + \boxed{(a_1 + a_4W_i + a_5Z_i)X_i} + a_2W_i + a_3Z_i$$

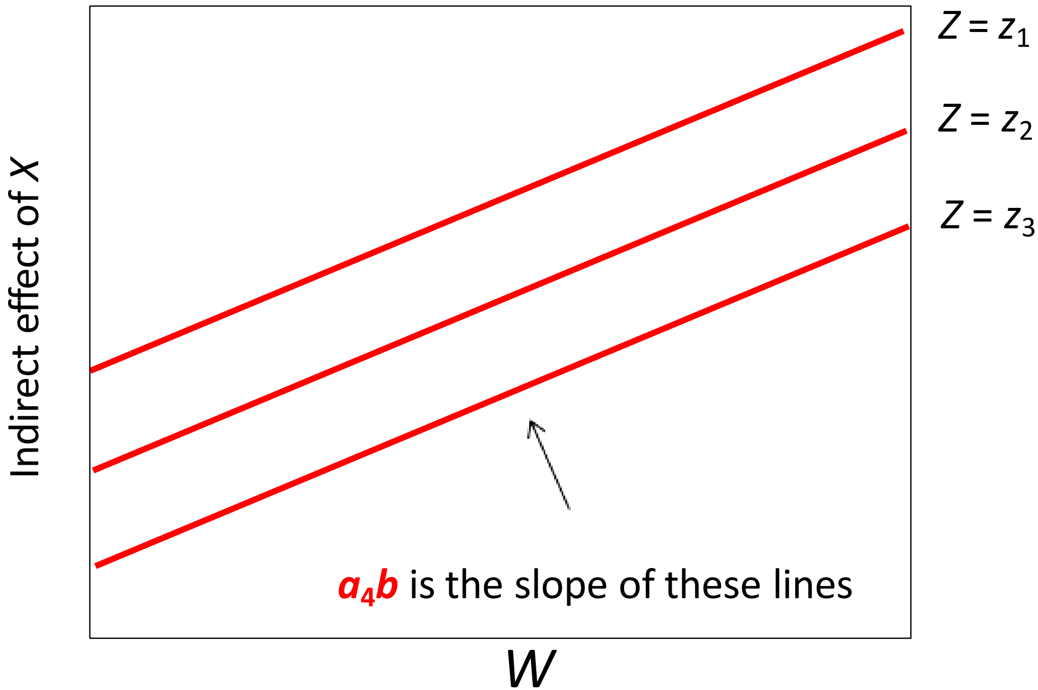
$$\widehat{M_i} = a_0 + \theta_{X\rightarrow M}X_i + a_2W_i + a_3Z_i$$

where $\theta_{X\rightarrow M} = a_1 + a_4W + a_5Z$ is the conditional effect of X on M. It is a linear function of both W and Z.

Index of partial moderated mediation

Indirect effect of X: $(a_1 + a_4W + a_5Z)b = a_1b + \underline{a_4bW} + a_5bZ$

When visualized, the model of the indirect effect of X might look something like this (depending on the estimates of the regression coefficients)



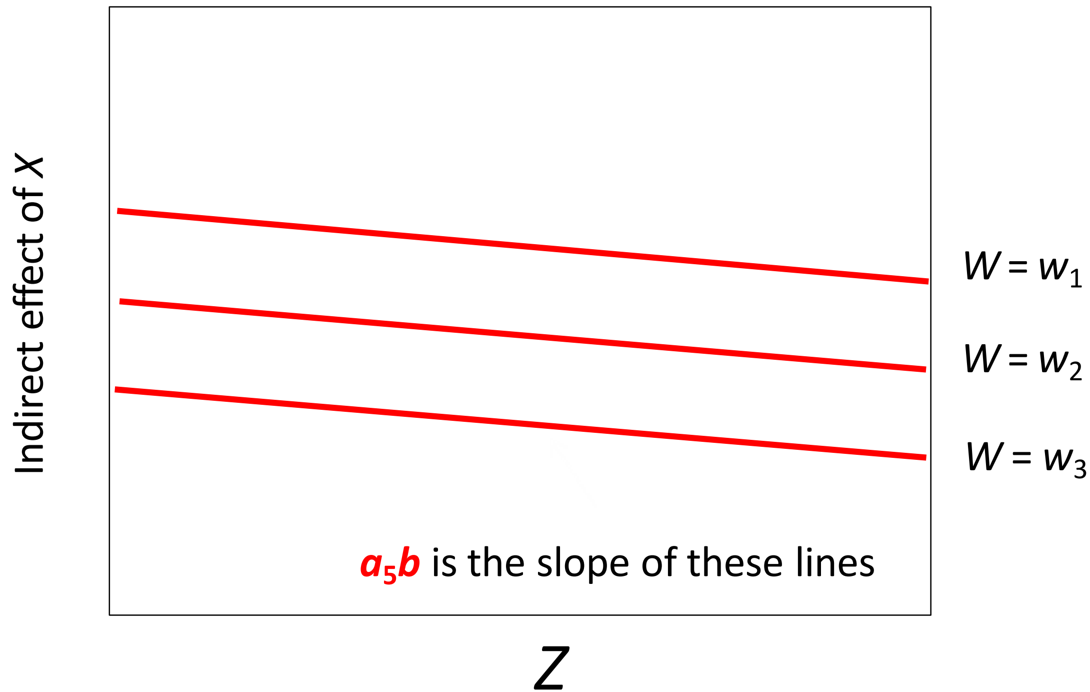
- a_4b, the "index of partial moderated mediation" by W, quantifies how the indirect effect of X changes as W changes but Z is fixed.

- Test whether this index is different from zero to test "partial moderated mediation." of X's effect on Y through M by W. PROCESS can do this using a bootstrap CI.

Index of partial moderated mediation

Indirect effect of X: $(a_1 + a_4W + a_5Z)b = a_1b + a_4bW + \underline{a_5bZ}$

When visualized, the model of the indirect effect of X might look something like this:
(depending on the estimates of the regression coefficients)



- a_5b, the "index of partial moderated mediation" by Z, quantifies how the indirect effect of X changes as Z changes but W is fixed.
- Test whether this index is different from zero to test "partial moderated mediation." of X's effect on Y through M by Z. PROCESS can do this using a bootstrap CI.

An example

Peltonen, K., Qouta, S., Sarraj, E. E., & Panamäki, R-L. (2010). Military trauma and social development: The moderating and mediating roles of peer and sibling relations in mental health. *International Journal of Behavioral Development*, *34*, 554-563.

Photo from MyriamsFotos via Pixabay

208 Palestinian children between the ages of 10 and 14 living in Gaza and measured in 2006.

TRAUMA: A count of exposure to traumatic events during the Al-Aqsa Intifada (e.g., shelling of home, being shot, losing family members, witnessing of killing). Range: 0 to 18

FRQUAL: Quality of a child's friendships as measured with the Friendship Qualities questionnaire. Eight items scaled 1 to 5 e.g., "I have friends with whom can share my secrets"

DEPRESS: Depressive symptoms measured by the Child Depression Inventory (CDI). Range 1 to 28.

AGE: Child age in years

SEX: Sex of the child (0 = female, 1 = male)

The data: gaza

gaza.sav

File Edit View Data Transform Analyze Direct Marketing Graphs Utilitie

	age	sex	frqual	trauma	depress
1	13	1	5.000	8.00	8
2	10	1	4.875	10.00	8
3	10	0	3.500	6.00	7
4	12	1	3.000	8.00	19
5	11	0	4.125	8.00	3
6	11	1	4.250	15.00	5
7	11	1	4.125	12.00	8
8	10	1	3.375	9.00	11
9	11	1	5.000	6.00	6
10	12	1	5.000	7.00	4
11	10	1	1.000	7.00	27

gaza.sas

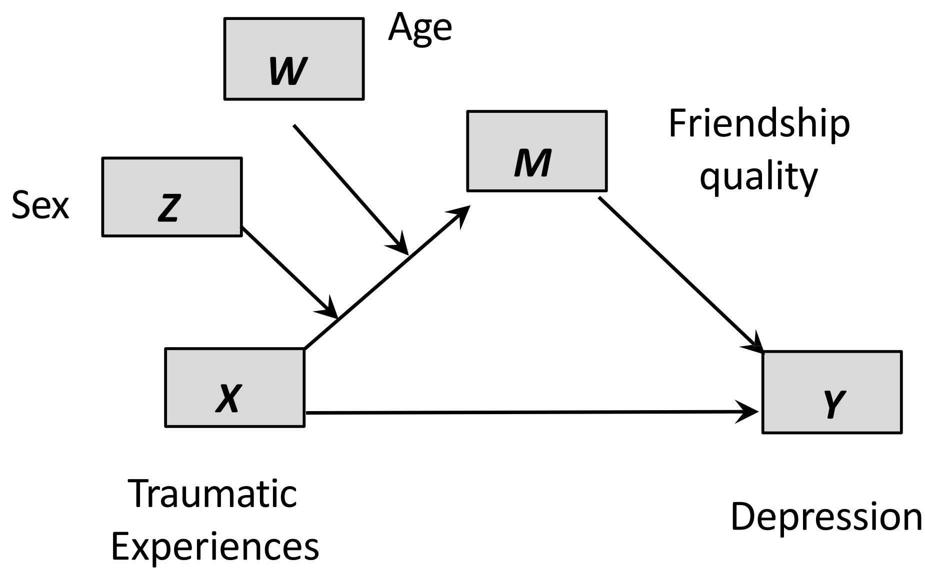
gaza

```
data gaza;
input age sex frqual trauma depress;
datalines;
13 1 5.000 8.00 8
10 1 4.875 10.00 8
10 0 3.500 6.00 7
12 1 3.000 8.00 19
11 0 4.125 8.00 3
11 1 4.250 15.00 5
11 1 4.125 12.00 8
10 1 3.375 9.00 11
11 1 5.000 6.00 6
12 1 5.000 7.00 4
```

In R: Don't forget to change the path below to where your **gaza.csv** file is located.

```
> gaza<-read.table("c:/mmcpa/gaza.csv", sep=",",header=TRUE)
> head(gaza)
  age sex frqual trauma depress
1  13   1  5.000      8       8
2  10   1  4.875     10       8
3  10   0  3.500      6       7
4  12   1  3.000      8      19
5  11   0  4.125      8       3
6  11   1  4.250     15       5
```

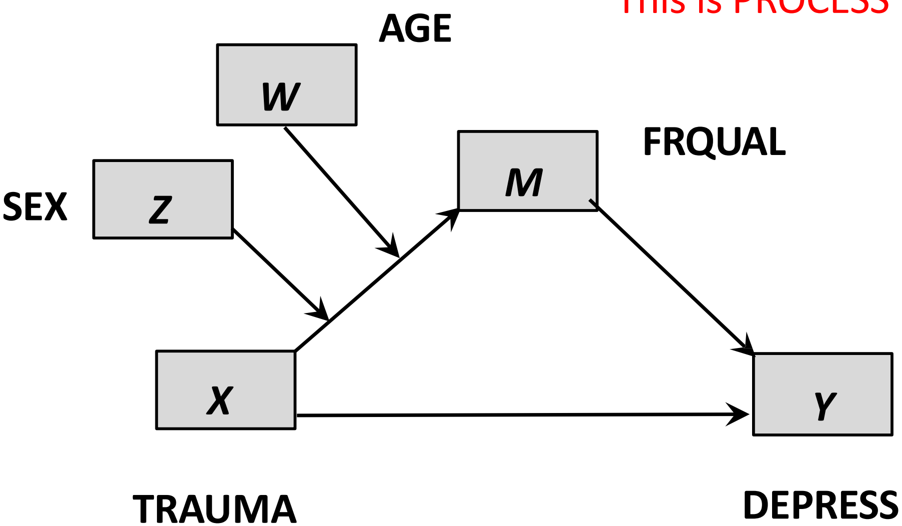
Estimation in PROCESS

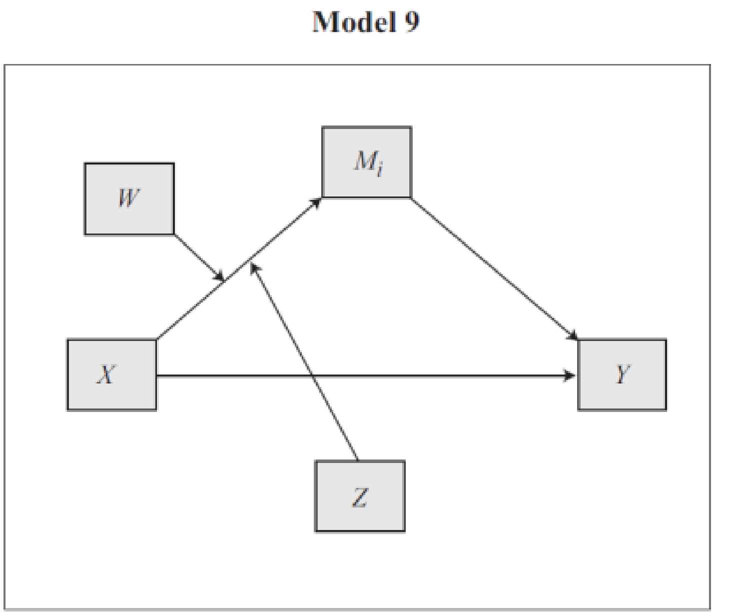


The investigators proposed a model in which friendship quality mediates the effect of trauma on depression, with that mechanism varying by both sex and age due to the effect of trauma on friendship quality varying as a function of both sex and age.

Estimation in PROCESS

This is PROCESS model 9






```
process y=depress/x=trauma/m=frqual/w=age/z=sex/moments=1/plot=1/model=9
  /seed=145.
```

```
%process (data=gaza,y=depress,x=trauma,m=frqual,w=age,z=sex,moments=1,plot=1,
  model=9,seed=145);
```

```
process(data=gaza,y="depress",x="trauma",m="frqual",w="age",z="sex",moments=1,
  plot=1,model=9,seed=145)
```

PROCESS output

```
Model  : 9
    Y.: depress
    X  : trauma
    M  : frqual
    W  : age
    Z. : sex

Sample
Size:  208

**************************************************************************
OUTCOME VARIABLE:
 frqual
```

$$\widehat{M_i} = 1.825 + 0.223X_i + 0.242W_i - 0.914Z_i - 0.029X_iW_i + 0.122X_iZ_i$$

Model Summary

R	R-sq	MSE	F	df1	df2	p
.3062	.0938	.5754	4.1811	5.0000	202.0000	.0012

Model

	coeff	se	t	p	LLCI	ULCI	
constant	1.8247	1.1547	1.5803	.1156	-.4521	4.1016	
trauma	.2232	.1220	1.8299	.0687	-.0173	.4638	← a_1
age	.2424	.1071	2.2636	.0247	.0312	.4535	
Int_1	-.0288	.0111	-2.5921	.0102	-.0507	-.0069	← a_4
sex	-.9141	.2452	-3.7276	.0003	-1.3976	-.4306	
Int_2	.1219	.0323	3.7767	.0002	.0582	.1855	← a_5

```
Product terms key:
 Int_1    :        trauma    x        age
 Int_2    :        trauma    x        sex
```

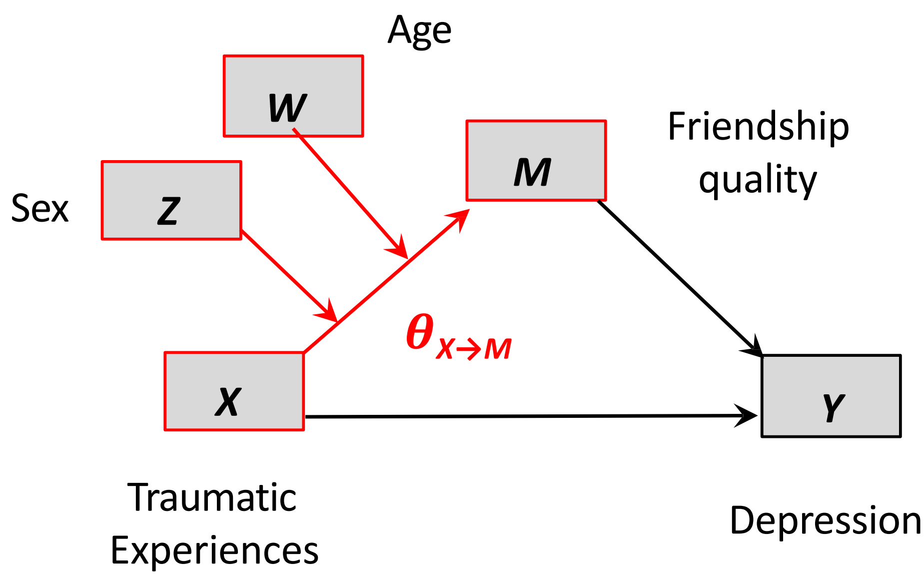
Test(s) of highest order unconditional interaction(s):

	R2-chng	F	df1	df2	p
X*W	.0301	6.7192	1.0000	202.0000	.0102
X*Z	.0640	14.2636	1.0000	202.0000	.0002
BOTH(X)	.0697	7.7643	2.0000	202.0000	.0006

\-----------------

Notice that the effect of trauma on friendship quality varies significantly by both age and sex. More on this later.

The conditional effect of *X* on *M*



$$\widehat{M}_i = a_0 + a_1X_i + a_2W_i + a_3Z_i + a_4X_iW_i + a_5X_iZ_i$$

can be written in equivalent form as

$$\widehat{M}_i = a_0 + (a_1 + a_4W_i + a_5Z_i)X_i + a_2W_i + a_3Z_i$$

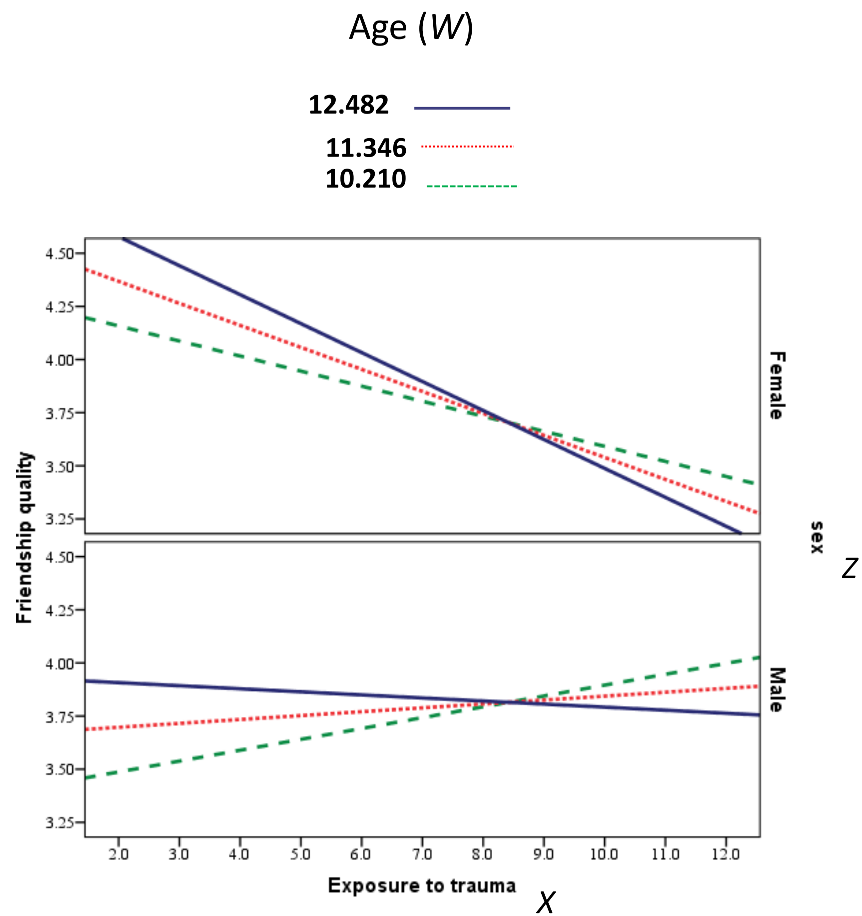
$$\widehat{M}_i = 1.825 + (0.223 - 0.029W_i + 0.122Z_i)X_i + 0.242W_i - 0.914Z_i$$

and so *X*'s effect on *M* is $\theta_{X\rightarrow M} = 0.223 - 0.029W + 0.122Z$

PROCESS provides what we need to visualize the model of *M*

In SPSS, the plot option in PROCESS produces a program you can use to produce a rough visual depiction of the model. Or use the information in the table to plot in your preferred software.

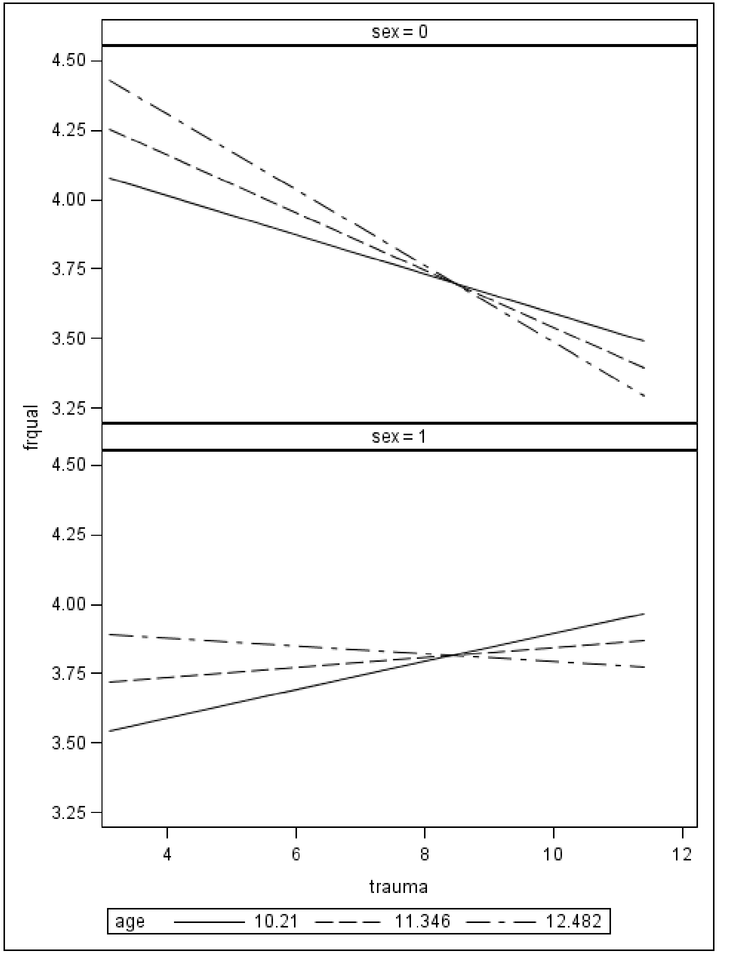
```
DATA LIST FREE/
   trauma      age         sex         frqual      .
BEGIN DATA.
     3.0987    10.2103     .0000       4.0801
     7.2548    10.2103     .0000       3.7857
     11.4109   10.2103     .0000       3.4914
     3.0987    10.2103     1.0000      3.5436
     7.2548    10.2103     1.0000      3.7557
     11.4109   10.2103     1.0000      3.9678
     3.0987    11.3462     .0000       4.2540
     7.2548    11.3462     .0000       3.8237
     11.4109   11.3462     .0000       3.3934
     3.0987    11.3462     1.0000      3.7175
     7.2548    11.3462     1.0000      3.7937
     11.4109   11.3462     1.0000      3.8698
     3.0987    12.4820     .0000       4.4280
     7.2548    12.4820     .0000       3.8617
     11.4109   12.4820     .0000       3.2954
     3.0987    12.4820     1.0000      3.8915
     7.2548    12.4820     1.0000      3.8316
     11.4109   12.4820     1.0000      3.7718
END DATA.
GRAPH/SCATTERPLOT=
 trauma    WITH      frqual    BY        age
    /PANEL   ROWVAR=  sex.
```



In SAS

In SAS, write a program around the data to produce a plot.

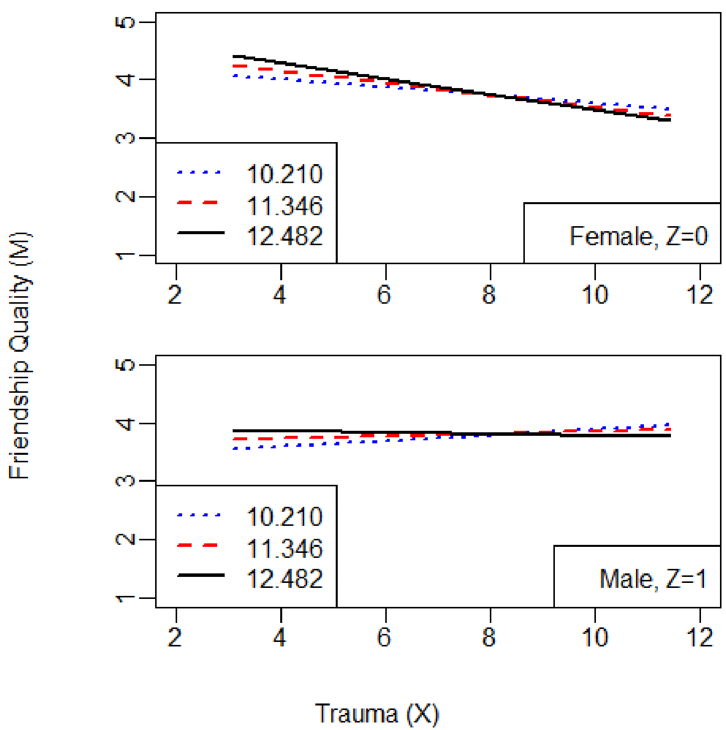
```
data gazaplot;
input trauma age sex frqual;
datalines;
      3.0987   10.2103    .0000       4.0801
      7.2548   10.2103    .0000       3.7857
     11.4109   10.2103    .0000       3.4914
      3.0987   10.2103   1.0000       3.5436
      7.2548   10.2103   1.0000       3.7557
     11.4109   10.2103   1.0000       3.9678
      3.0987   11.3462    .0000       4.2540
      7.2548   11.3462    .0000       3.8237
     11.4109   11.3462    .0000       3.3934
      3.0987   11.3462   1.0000       3.7175
      7.2548   11.3462   1.0000       3.7937
     11.4109   11.3462   1.0000       3.8698
      3.0987   12.4820    .0000       4.4280
      7.2548   12.4820    .0000       3.8617
     11.4109   12.4820    .0000       3.2954
      3.0987   12.4820   1.0000       3.8915
      7.2548   12.4820   1.0000       3.8316
     11.4109   12.4820   1.0000       3.7718

run;
proc sgpanel;
panelby sex / columns=1;
series x=trauma y=frqual/group=age lineattrs
=(color=black);run;
```



In R

```
oldp<-par(mfrow=c(2,1),mar=c(3,4,0,0),oma=c(2,2,2,2),
mgp=c(5,0.5,0))
x<-c(3.099,7.255,11.419,3.099,7.255,11.419,3.099,7.255,11.419)
w<-c(10.210,10.210,10.210,11.346,11.346,11.346,12.482,12.482,12.482)
yz0<-c(4.080,3.786,3.491,4.254,3.824,3.393,4.428,3.862,3.295)
yz1<-c(3.544,3.756,3.968,3.718,3.793,3.870,3.892,3.832,3.772)
legend.txt<-c("10.210","11.346","12.482")
for (i in 1:2){
if (i==1)
  {y<-yz0
  legend2.txt<-c("Female, Z=0")}
if (i==2)
  {y<-yz1
  legend2.txt<-c("Male, Z=1")}
plot(y=y,x=x,col="white",ylim=c(1,5),cex=1.5,
  xlim=c(2,12),tcl=-0.5)
lines(x[w==10.210],y[w==10.210],lwd=2,lty=3,col="blue")
lines(x[w==11.346],y[w==11.346],lwd=2,lty=2,col="red")
lines(x[w==12.482],y[w==12.482],lwd=2,lty=1,col="black")
legend("bottomleft", legend=legend.txt,lwd=2,
  lty=c(3,2,1),col=c("blue","red","black"))
legend("bottomright",legend=legend2.txt)
box}
mtext("Trauma (X)",side=1,outer=TRUE)
mtext("Friendship Quality (M)",side=2,outer=TRUE)
par<-oldp
```

From PROCESS **plot** output



The conditional effect of *X* on *M*

```
Focal predict: trauma (X)
      Mod var: age      (W)
      Mod var: sex      (Z)
```

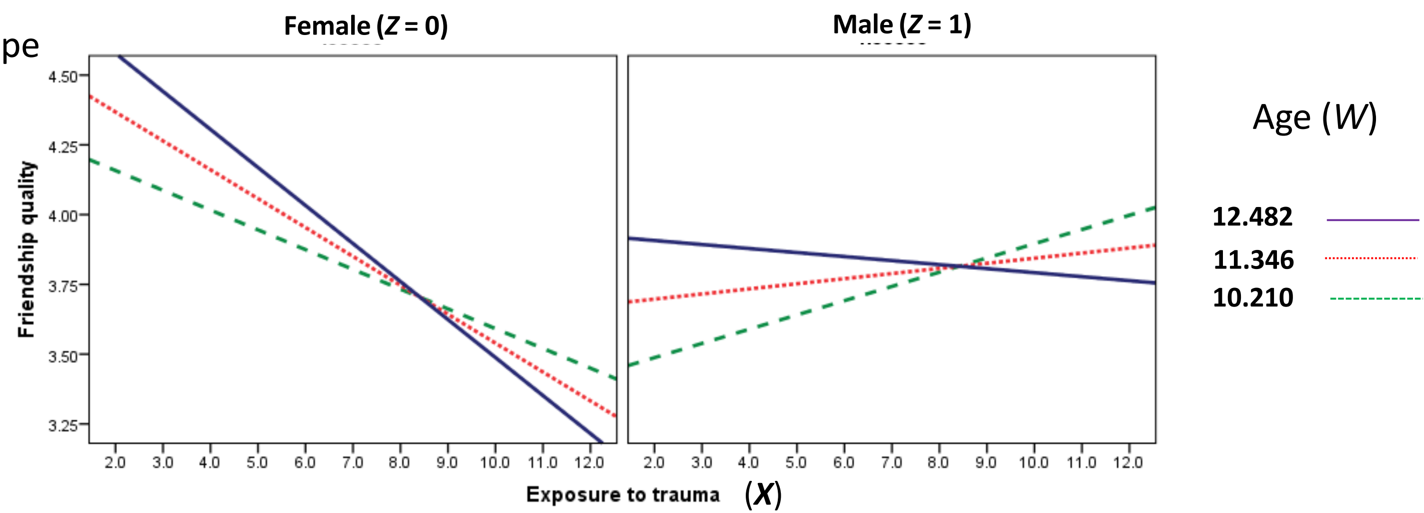
$$\theta_{X \to M} = 0.223 - 0.029W + 0.122Z$$

Conditional effects of the focal predictor at values of the moderator(s):

W age	*Z* sex	Effect	se	t	p	LLCI	ULCI
10.2103	.0000	-.0708	.0257	-2.7593	.0063	-.1214	-.0202
10.2103	1.0000	.0510	.0251	2.0335	.0433	.0015	.1005
11.3462	.0000	-.1035	.0258	-4.0078	.0001	-.1545	-.0526
11.3462	1.0000	.0183	.0173	1.0575	.2915	-.0158	.0525
12.4820	.0000	-.1363	.0315	-4.3204	.0000	-.1984	-.0741
12.4820	1.0000	-.0144	.0170	-.8470	.3980	-.0479	.0191

$\boldsymbol{\theta_{X \to M}}$ is the slope of these lines

M



The differences in slope conditioned on the second moderator are constant across values of the second moderator. These depict two-way and not “three way” interaction.

PROCESS output

$$\widehat{Y}_i = 21.927 + 0.228X_i - 3.572M_i$$

```
OUTCOME VARIABLE:
 depress

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .4637      .2150    38.0694    28.0709     2.0000   205.0000      .0000

Model
              coeff         se          t          p       LLCI       ULCI
constant    21.9272     2.3872     9.1853      .0000    17.2206    26.6338
trauma        .2883      .1043     2.7637      .0062      .0826      .4939
frqual      -3.5718      .5507    -6.4858      .0000    -4.6576    -2.4860

****************** DIRECT AND INDIRECT EFFECTS OF X ON Y *****************

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .2883      .1043     2.7637      .0062      .0826      .4939

Conditional indirect effects of X on Y:

INDIRECT EFFECT:
 trauma      ->    frqual        ->    depress

          age        sex     Effect     BootSE   BootLLCI   BootULCI
      10.2103      .0000      .2530      .1386      .0178      .5626
      10.2103     1.0000     -.1823      .1038     -.4040      .0107
      11.3462      .0000      .3698      .1378      .1377      .6789
      11.3462     1.0000     -.0654      .0693     -.2176      .0584
      12.4820      .0000      .4867      .1563      .2089      .8368
      12.4820     1.0000      .0514      .0673     -.0884      .1767
```

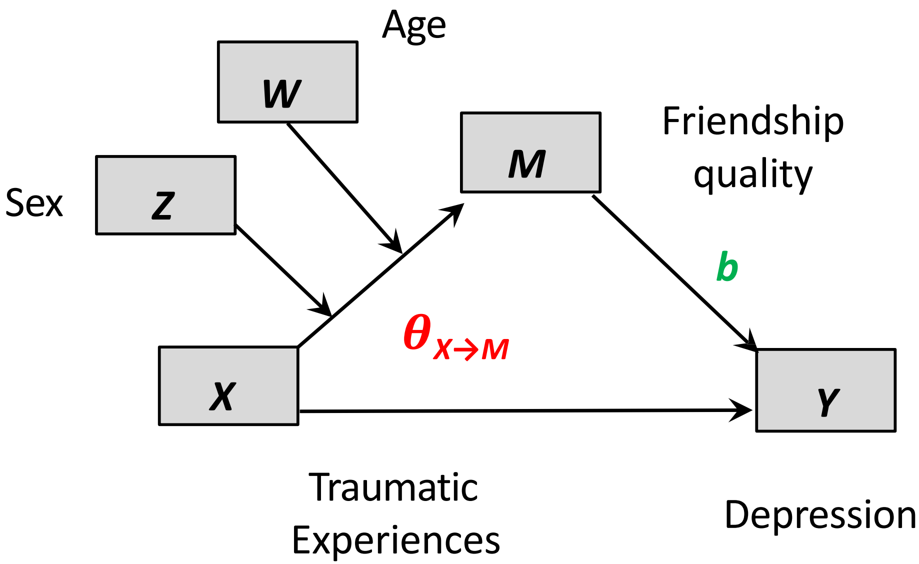
c'

b

c'

Conditional indirect effects

The conditional indirect effect



$\theta_{X \to M} = a_1 + a_4 W + a_5 Z$
$= 0.223 - 0.029W + 0.122Z$

***b* = -3.572**

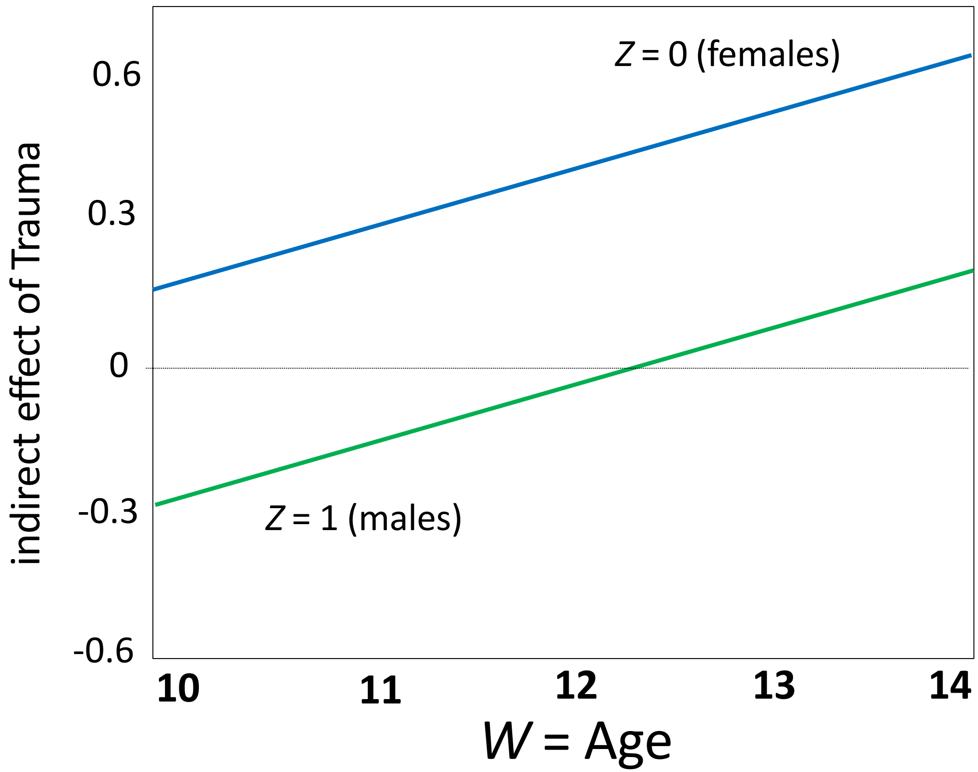
$$\widehat{M_i} = 1.825 + 0.223X_i + 0.242W_i - 0.914Z_i - 0.029X_iW_i + 0.122X_iZ_i$$

$$\widehat{Y_i} = 21.927 + 0.228X_i - 3.572M_i$$

The indirect effect of *X* on *Y* is the product of the conditional effect of *X* on *M*, which is $\theta_{X \to M} = a_1 + a_4W + a_5Z = 0.223 - 0.029W + 0.122Z$, and the effect of *M* on *Y* (***b* = -3.572**): $(a_1 + a_4W + a_5Z)b = (0.223 - 0.029W + 0.122Z)(-3.572)$, which can be equivalently written as $a_1b + a_4bW + a_5bZ =$ $\mathbf{-0.797 + 0.103}\boldsymbol{W}\mathbf{ - 0.435}\boldsymbol{Z}$. It is an additive linear function of both *W* and *Z* and *depends on* or is *conditioned on* the two moderators.

A visual representation

$$(a_1 + a_4W + a_5Z)b = a_1b + a_4bW + a_5bZ = -0.797 + 0.103W - 0.435Z$$



The indirect varies by age with sex fixed, and by sex with age fixed. This phenomenon is "partial moderated mediation"---the moderation of the indirect effect of *X* by one moderator when the second is held constant. But this is merely a description of the data.

Partial moderation by *W* of the mediation

Indirect effect of *X*

$$(a_1 + a_4W + a_5Z)b$$

or

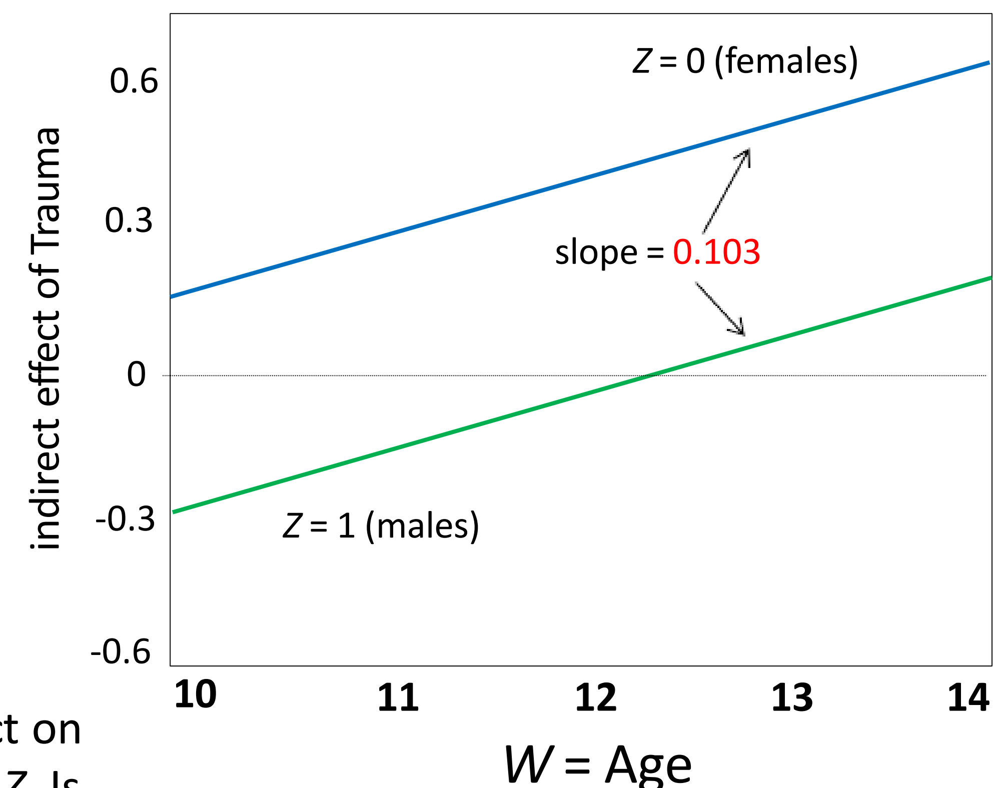
$$a_1b + a_4bW + a_5bZ$$

$$-0.797 + 0.103W - 0.435Z$$

$$a_4b = \mathbf{0.103}$$

This is the "index of partial moderated mediation" of *X's* effect on *Y* through *M* by *W*, controlling for *Z*. Is this statistically significant? If so, this implies **partial moderated mediation.**

a_4b is the common slope of these lines



Inference for the index of partial moderated mediation

A bootstrap confidence interval can be used for inference about the partial moderation of the indirect effect of *X* on *Y* through *M* by *W*. PROCESS does this for you.

Indices of partial moderated mediation:

	Index	BootSE	BootLLCI	BootULCI
age	.1029	.0470	.0087	.1934
sex	-.4352	.1577	-.7878	-.1625

We can say that independent of any moderation of trauma's indirect effect on depression by sex, age moderates the indirect effect of trauma. The indirect effect is "larger" (i.e., further to the right on the number line) among older kids.

Partial moderation by Z of the mediation

Indirect effect of X

$$(a_1 + a_4W + a_5Z)b$$

or

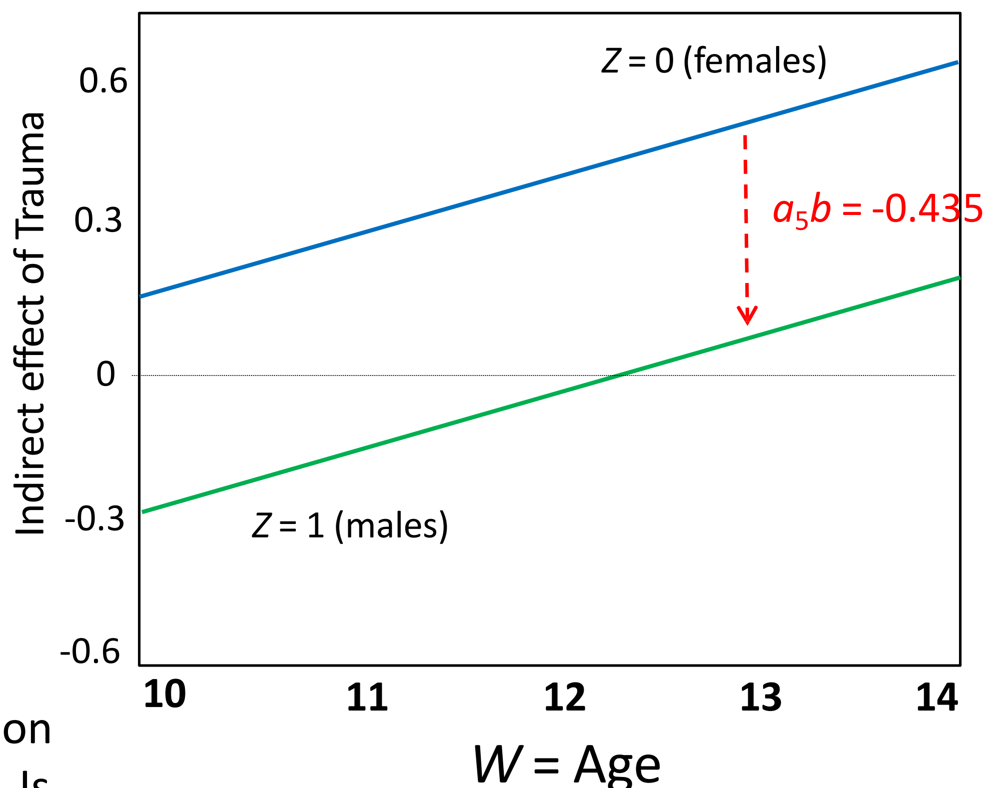
$$a_1b + a_4bW + a_5bZ$$

$$-0.797 + 0.103W - 0.435Z$$

$$a_5b = -0.435$$

This is the "index of partial moderated mediation" of X's effect on Y through M by Z, controlling for W. Is this statistically significant? If so, this implies **partial moderated mediation.**

a_5b is the distance between these lines



Inference for the index of partial moderated mediation

A bootstrap confidence interval can be used for inference about the partial moderation of the indirect effect of *X* on *Y* through *M* by *Z*. PROCESS does this for you.

Indices of partial moderated mediation:

	Index	BootSE	BootLLCI	BootULCI
age	.1029	.0470	.0087	.1934
sex	-.4352	.1577	-.7878	-.1625

We can say that independent of any moderation of trauma's indirect effect on depression by age, the indirect effect of trauma differs between males and females. The indirect effect is "smaller" (i.e., more to the left on the number line) among males than among females.

Conditional indirect effects

PROCESS generates the conditional indirect effect of *X* at various combinations of *W* and *Z*, with bootstrap confidence intervals

$(a_1 + a_4W + a_5Z)b$

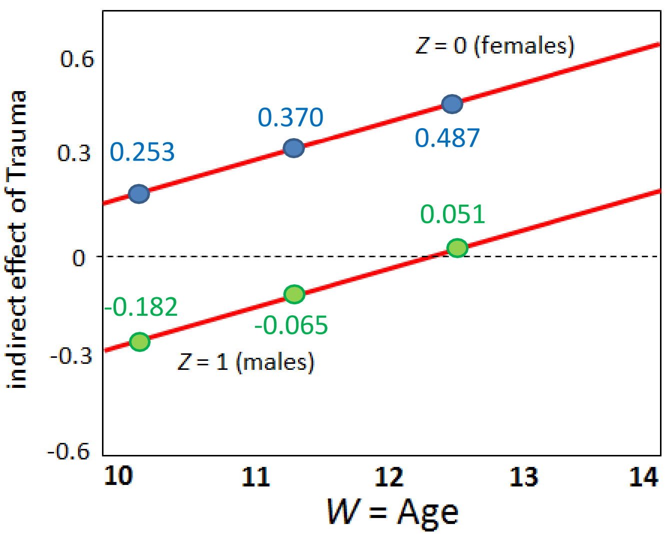
$-0.797 + 0.103W - 0.435Z$

```
Conditional indirect effects of X on Y:

INDIRECT EFFECT:
 trauma        ->   frqual        ->    depress

      age        sex       Effect     BootSE  BootLLCI    BootULCI
  10.2103      .0000        .2530      .1386     .0178       .5626
  10.2103     1.0000       -.1823      .1038    -.4040       .0107
  11.3462      .0000        .3698      .1378     .1377       .6789
  11.3462     1.0000       -.0654      .0693    -.2176       .0584
  12.4820      .0000        .4867      .1563     .2089       .8368
  12.4820     1.0000        .0514      .0673    -.0884       .1767
```

Age (*W*)	Sex (*Z*)	$\theta_{X \to M}$	b	$\theta_{X \to M}\, b$
10.210	0	-0.071	-3.572	0.253
10.210	1	0.051	-3.572	-0.182
11.346	0	-0.104	-3.572	0.370
11.346	1	0.018	-3.572	-0.065
12.482	0	-0.136	-3.572	0.487
12.482	1	-0.014	-3.572	0.051

This model has a major set of constraints

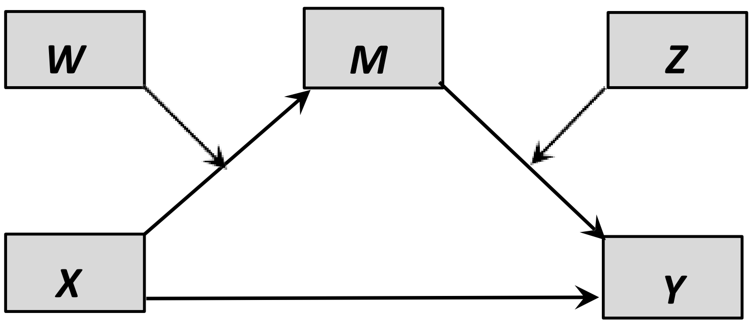
```
Conditional indirect effects of X on Y:

INDIRECT EFFECT:
 trauma        ->     frqual        ->     depress

       age          sex       Effect       BootSE     BootLLCI     BootULCI
   10.2103        .0000        .2530        .1386        .0178        .5626
   10.2103       1.0000       -.1823        .1038       -.4040        .0107
   11.3462        .0000        .3698        .1378        .1377        .6789
   11.3462       1.0000       -.0654        .0693       -.2176        .0584
   12.4820        .0000        .4867        .1563        .2089        .8368
   12.4820       1.0000        .0514        .0673       -.0884        .1767
```

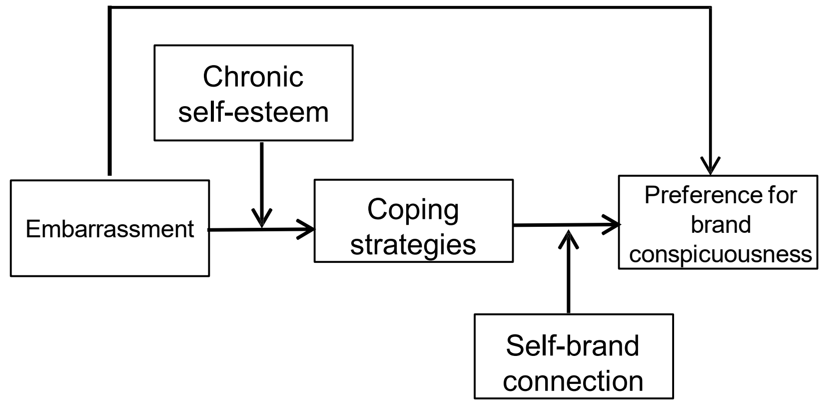


These results suggest that the indirect effect of traumatic experiences on depression through friendship quality is positive among females regardless of age, but among males, the indirect effect is not different from zero regardless of age. **But this does NOT mean that relationship between age and the size of the indirect effect of trauma differs between males and females, or that the sex difference in the indirect effect varies by age. Our model doesn't allow that, regardless of what the data look like or the reality we are modeling.**

When the first and second stages are moderated by different variables

A first and second stage dual moderated mediation model

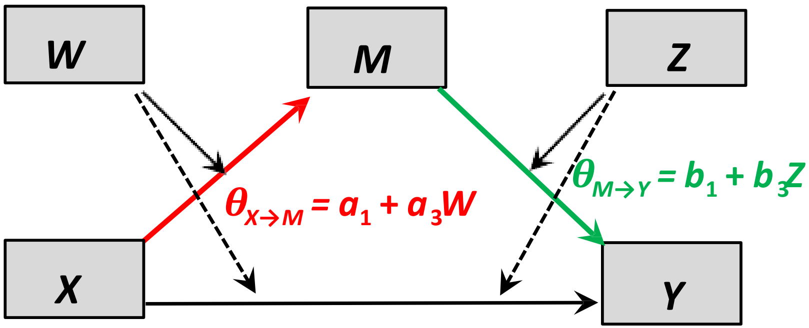




Song, X., Huang, F., & Li, X. (2017). The effect of embarrassment on preferences for brand conspicuousness: The roles of self-esteem and self-brand connection**. *Journal of Consumer Psychology*,** *27*, 69-83.

In this model, the moderation of the indirect effect of *X* on *Y* through *M* by one moderator can vary as a function of the second moderator. We can test if it does with an *index of moderated moderated mediation.* And we can estimate and test how much the indirect effect varies as a function of one moderator when we condition on a value of the second moderator using an *index of conditional moderated mediation.*

The indirect effect as a multiplicative function of two moderators



Indirect effect of X

$$\theta_{X\to M}\theta_{M\to Y} = (a_1 + a_3W)(b_1 + b_3Z)$$

or

$$a_1b_1 + a_3b_1W + a_1b_3Z + a_3b_3WZ$$

$$\widehat{M}_i = a_0 + a_1X_i + a_2W_i + a_3X_iW_i$$

$$\widehat{Y}_i = c_0' + c'X_i + b_1M_i + b_2Z_i + b_3M_iZ_i$$

These equations can be written in equivalent form as

$$\widehat{M}_i = a_0 + \theta_{X\to M}X_i + a_2W_i \quad \text{where} \quad \theta_{X\to M} = a_1 + a_3W$$

$$\widehat{Y}_i = c_0' + c'X_i + \theta_{M\to Y}M_i + b_2Z_i \quad \text{where} \quad \theta_{M\to Y} = b_1 + b_3Z$$

The indirect effect as a function of one moderator is a function of a second

Indirect effect of *X:* $(a_1 + a_3W)(b_1 + b_3Z)$

$$= a_1b_1 + a_3b_1W + a_1b_3Z + a_3b_3WZ$$

This equation for the indirect effect of *X* can be written in two alternative forms:

$$a_1b_1 + (a_3b_1 + a_3b_3Z)W + a_1b_3Z$$

and so the indirect of *X* on *Y* through *M* depends on *W*, with that dependency being a linear function of *Z*: $a_3b_1 + a_3b_3Z$

OR

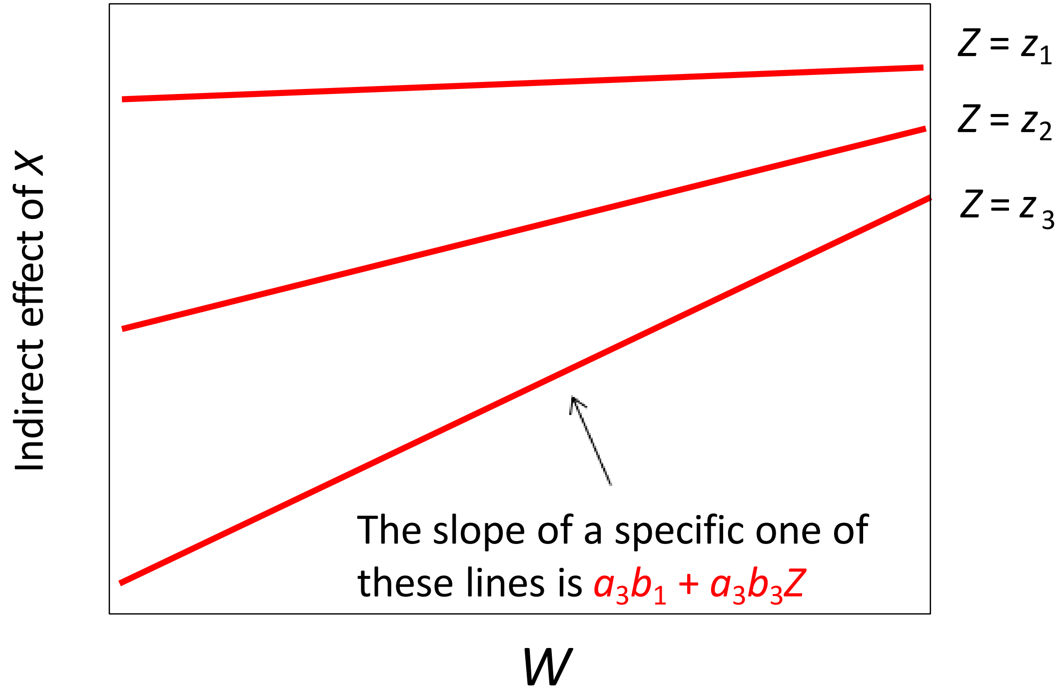
$$a_1b_1 + (a_1b_3 + a_3b_3W)Z + a_3b_1W$$

and so the indirect of *X* on *Y* through *M* depends on *Z*, with that dependency being a linear function of *W*: $a_1b_3 + a_3b_3W$

Index of conditional moderated mediation

Indirect effect of X: $a_1b_1 + (a_3b_1 + a_3b_3Z)W + a_1b_3Z$

When visualized, the model of the indirect effect of X might look something like this:

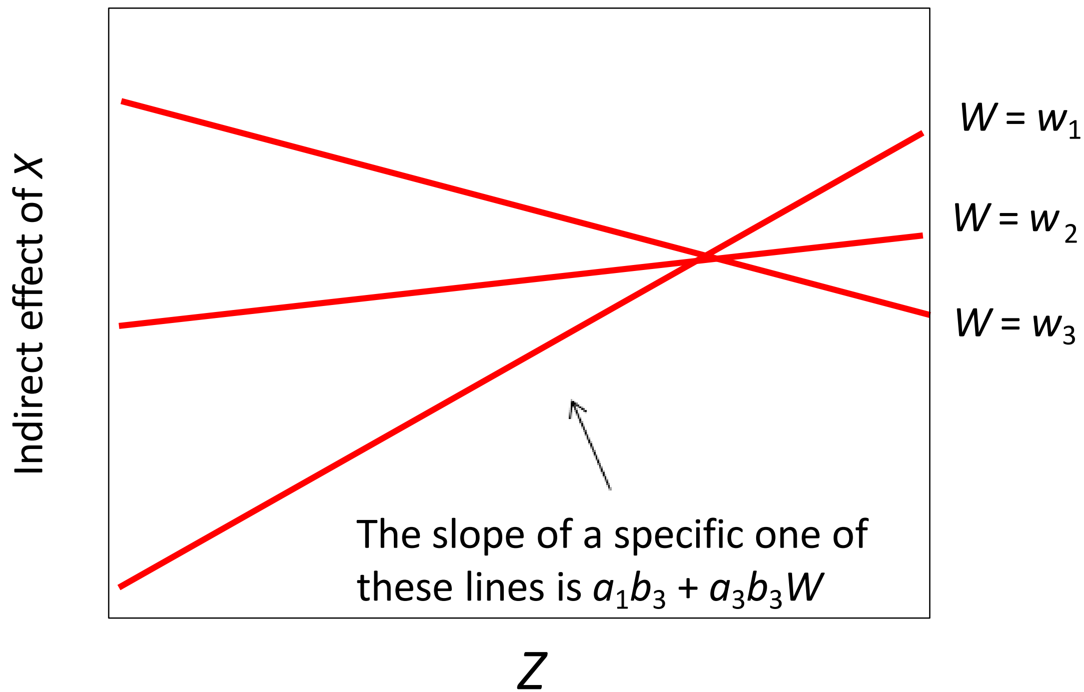


- $a_3b_1 + a_3b_3Z$, the "index of conditional moderated mediation" for W, quantifies how the indirect effect of X changes as W changes but Z is set to a specific value $Z = z$.
- Test whether this index is different from zero to test "conditional moderated mediation" of X's effect on Y through M by W when $Z = z$. PROCESS can do this.

Index of conditional moderated mediation

Indirect effect of X: $a_1b_1 + (\underline{a_1b_3 + a_3b_3W})Z + a_3b_1W$

When visualized, the model of the indirect effect of X might look something like this:



- $a_1b_3 + a_3b_3W$, the "index of conditional moderated mediation" for Z, quantifies how the indirect effect of X changes as Z changes but W is set to a specific value $W = w$.
- Test whether this index is different from zero to test "conditional moderated mediation" of X's effect on Y through M by Z when $W = w$. PROCESS can do this

Primary and secondary moderators

Indirect effect of *X:* $(a_1 + a_3W)(b_1 + b_3Z)$

$= a_1b_1 + a_3b_1W + a_1b_3Z + a_3b_3WZ$

This equation for the indirect effect of *X* can be written in two alternative forms:

$$a_1b_1 + (a_3b_1 + a_3b_3Z)W + a_1b_3Z$$

and so the indirect of *X* on *Y* through *M* depends on *W* (the **primary moderator**), with that dependency being a linear function of *Z* (the **secondary moderator**)

OR

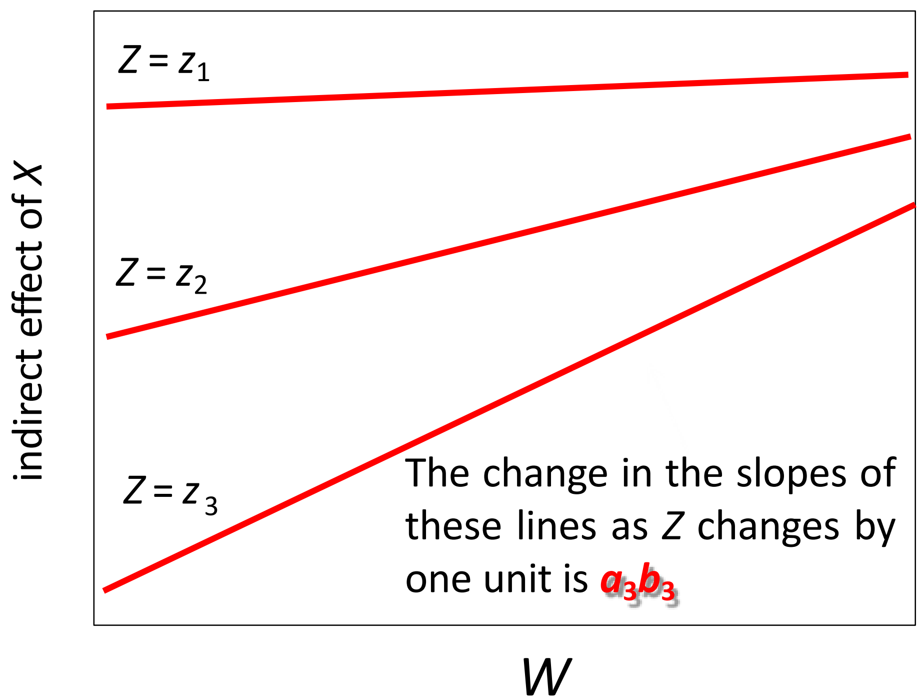
$$a_1b_1 + (a_1b_3 + a_3b_3W)Z + a_3b_1W$$

and so the indirect of *X* on *Y* through *M* depends on *Z* (the **primary moderator**) with that dependency being a linear function of *W* (the **secondary moderator**)

A test of moderated moderated mediation

Why bother conditioning on a second moderator if the moderation of indirect effect by the first moderator doesn't vary with the second moderator? We can test whether the moderation of mediation is itself moderated.

Indirect effect of X: $a_1b_1 + (a_3b_1 + \boxed{a_3b_3}Z)W + a_1b_3Z$



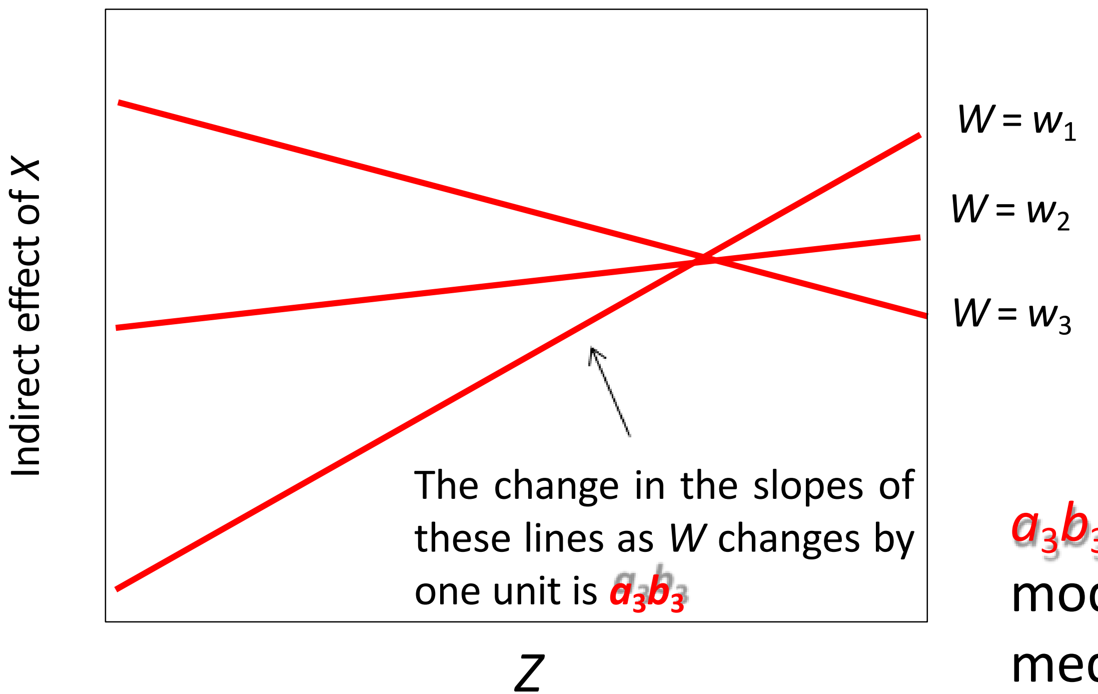
a_3b_3 is the "index of moderated moderated mediation"

A hypothesis test of confidence interval for the index of moderated moderation provides an inference for moderation of moderated mediation. PROCESS can do this.

Moderation of moderated mediation is symmetrical

We can switch the roles of moderators without changing the index of moderated moderated mediation.

$$\text{Indirect effect of } X\text{: } a_1b_1 + (a_1b_3 + \boxed{a_3b_3}W)Z + a_3b_1W$$



a_3b_3 is still the "index of moderated moderated mediation"

An example, inspired by…

Halbesleben, J. R. B. (2010). The role of exhaustion and workaround in predicting occupational injuries: A cross-lagged panel study of health care professionals. ***Journal of Occupational Health Psychology, 15***, 1-16.

Artwork from OpenClipart-Vectors via Pixabay

300 health care professions in a university hospital

EXHAUST : A self-report measure of physical and emotional exhaustion in the last month (scaled 1 to 7).

SAFETY: A self-report measure of the use of time-saving "workarounds" of safety protocols (scaled 1 to 7)

INJURY: A measure of occupational injury frequency and seriousness based on personnel incident reports (scaled 1 to 6) 6 months after assessment.

INJURYB: Occupational injury frequency (as measured for INJURY) at baseline.

TENURE: Job experience = years of employment

SEX: 0 = female, 1 = male

Numerical variables scaled such that higher = more.

The data: injuries

injuries.sav

injuries.sas

*injuries.sav [DataSet3] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Wind

22 :

	exhaust	safety	injury	tenure	serious	injuryb	latency	sex
1	3.50	5.50	2.10	5	.00	1.90	4.60	.00
2	2.60	3.50	1.00	8	.00	1.10	4.70	.00
3	3.40	3.70	2.30	9	.00	2.50	3.30	.00
4	4.30	4.40	2.30	8	.00	3.20	5.70	.00
5	3.90	4.60	3.00	12	1.00	1.00	3.60	.00
6	2.80	3.50	2.00	11	.00	2.60	5.80	.00
7	2.20	3.10	1.60	3	.00	2.30	3.00	1.00
8	2.50	3.90	2.40	2	.00	2.90	5.50	1.00
9	3.50	5.70	3.40	2	1.00	4.50	6.50	1.00
10	3.80	4.80	2.90	8	1.00	3.50	4.50	.00
11	3.80	2.50	2.10	4	.00	3.10	4.50	1.00

injuries

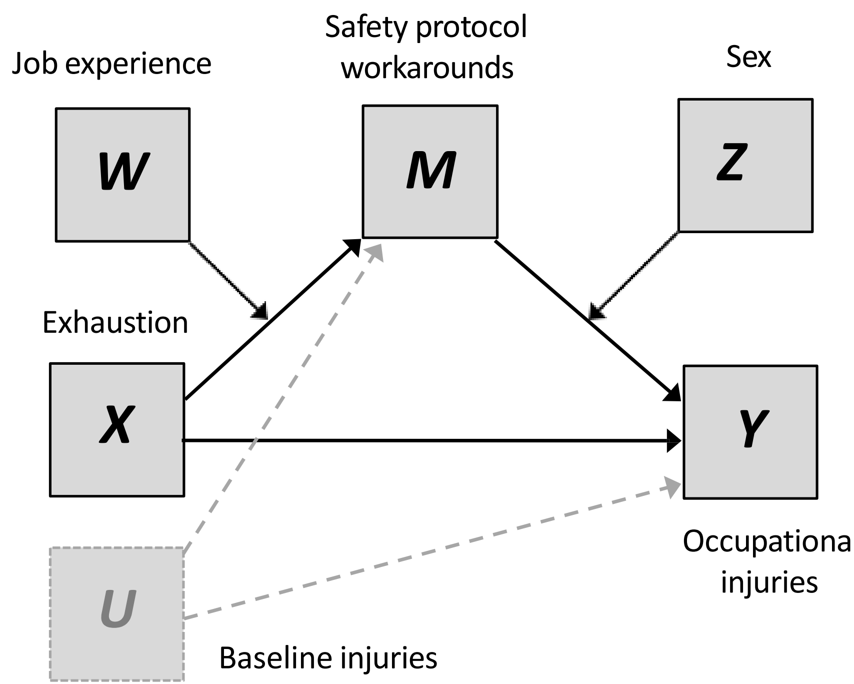
```
data injuries;
input exhaust safety injury tenure serious injuryb latency sex;
datalines;
     3.50     5.50     2.10        5      .00     1.90     4.60      .00
     2.60     3.50     1.00        8      .00     1.10     4.70      .00
     3.40     3.70     2.30        9      .00     2.50     3.30      .00
     4.30     4.40     2.30        8      .00     3.20     5.70      .00
     3.90     4.60     3.00       12     1.00     1.00     3.60      .00
     2.80     3.50     2.00       11      .00     2.60     5.80      .00
     2.20     3.10     1.60        3      .00     2.30     3.00     1.00
     2.50     3.90     2.40        2      .00     2.90     5.50     1.00
     3.50     5.70     3.40        2     1.00     4.50     6.50     1.00
     3.80     4.80     2.90        8     1.00     3.50     4.50      .00
     3.80     2.50     2.10        4      .00     3.10     4.50     1.00
     1.40     2.50     2.50        3      .00     2.30     4.60     1.00
```

In R: Don't forget to change the path below to where your **injuries.csv** file is located.

```
> injuries<-read.table("c:/mmcpa/injuries.csv", sep=",",header=TRUE)
> head(injuries)
  exhaust safety injury tenure serious injuryb latency sex
1     3.5    5.5    2.1      5       0     1.9     4.6   0
2     2.6    3.5    1.0      8       0     1.1     4.7   0
3     3.4    3.7    2.3      9       0     2.5     3.3   0
4     4.3    4.4    2.3      8       0     3.2     5.7   0
5     3.9    4.6    3.0     12       1     1.0     3.6   0
6     2.8    3.5    2.0     11       0     2.6     5.8   0
```

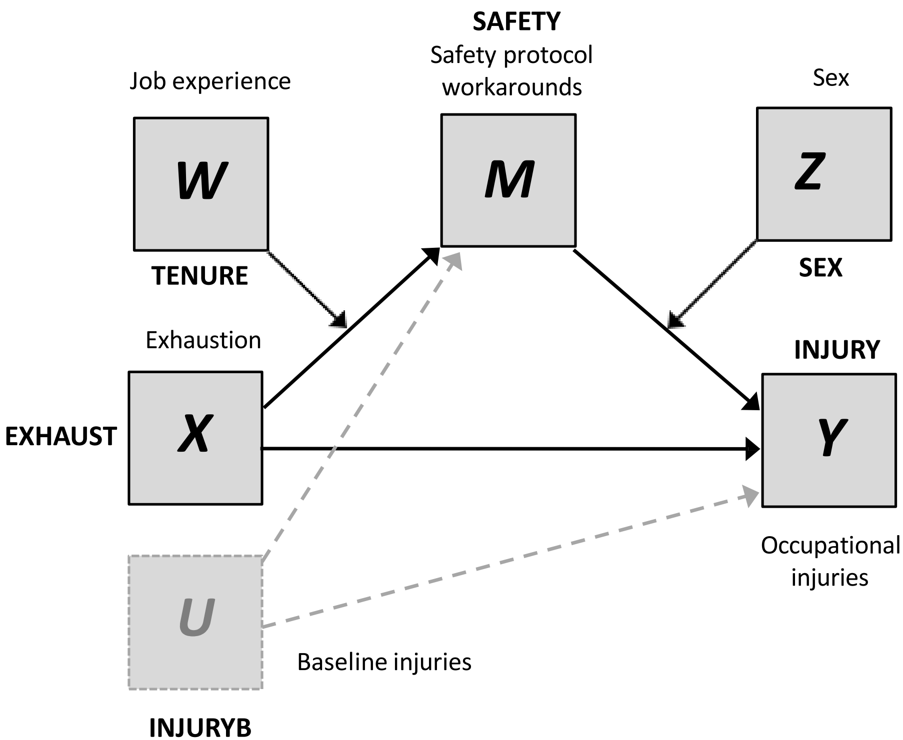
These are not the actual data from this study.

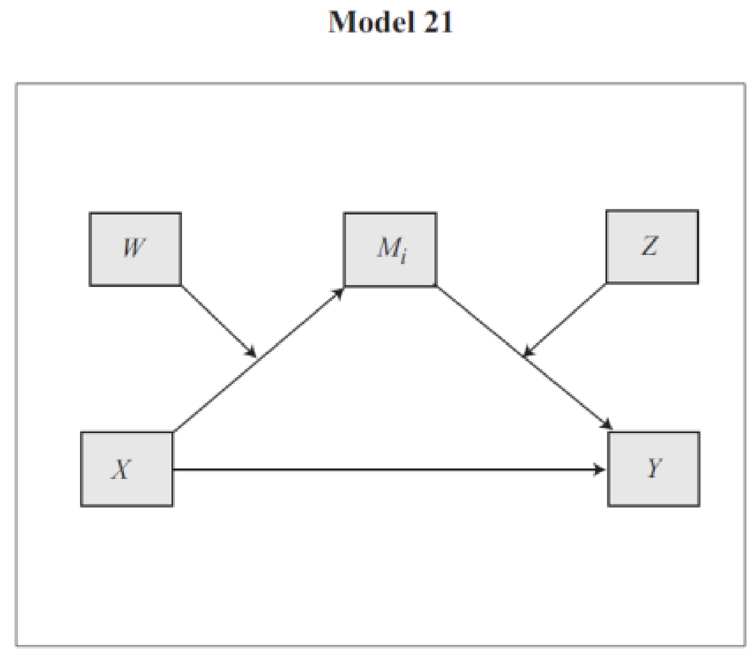
The model



Emotional and physical exhaustion is proposed to affect occupational injuries by increasing the use of workarounds to safety protocols, which enhance the likelihood of injury. But the strength of this mechanism should vary with job experience and sex, with job experience influencing the effect of exhaustion on use of safety-protocol workarounds, and sex moderating the effect of use of these workarounds on injuries.

The model and estimation in PROCESS






```
process  y=injury/x=exhaust/m=safety/w=tenure/z=sex/cov=injuryb/moments=1/plot=1/model=21/seed=465.
```

```
%process  (data=injuries,y=injury,x=exhaust,m=safety,w=tenure,z=sex,cov=injuryb,moments=1,plot=1,
   model=21,seed=456);
```

```
process(data=injuries,y="injury",x="exhaust",m="safety",w="tenure",z="sex",cov="injuryb",
   moments=1,plot=1,model=21,seed=456)
```

PROCESS output

```
Model  : 21
    Y  : injury
    X  : exhaust
    M  : safety
    W  : tenure
    Z  : sex

Covariates:
 injuryb

Sample
Size:  300

**************************************************************************
OUTCOME VARIABLE:
 safety

Model Summary
```

$$\widehat{M_i} = -2.491 + 0.626X_i + 0.077W_i - 0.061X_iW_i + 0.080U_i$$

```
Model
          R      R-sq       MSE          F        df1        df2          p
      .4930     .2431     .8937    23.6840     4.0000   295.0000      .0000
               coeff         se          t          p       LLCI       ULCI
Constant      2.4913      .3978     6.2631      .0000     1.7084     3.2741
Exhaust        .6256      .1092     5.7298      .0000      .4108      .8405
Tenure         .0773      .0540     1.4314      .1534     -.0290      .1836
Int_1         -.0607      .0162    -3.7399      .0002     -.0926     -.0287
Injuryb        .0800      .0550     1.4548      .1468     -.0282      .1882

Product terms key:
 Int_1     :        exhaust   x        tenure

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*W      .0359    13.9872     1.0000   295.0000      .0002
```

← a_1

← a_3

Using the plot option, PROCESS provides what we need to visualize the model of *M*

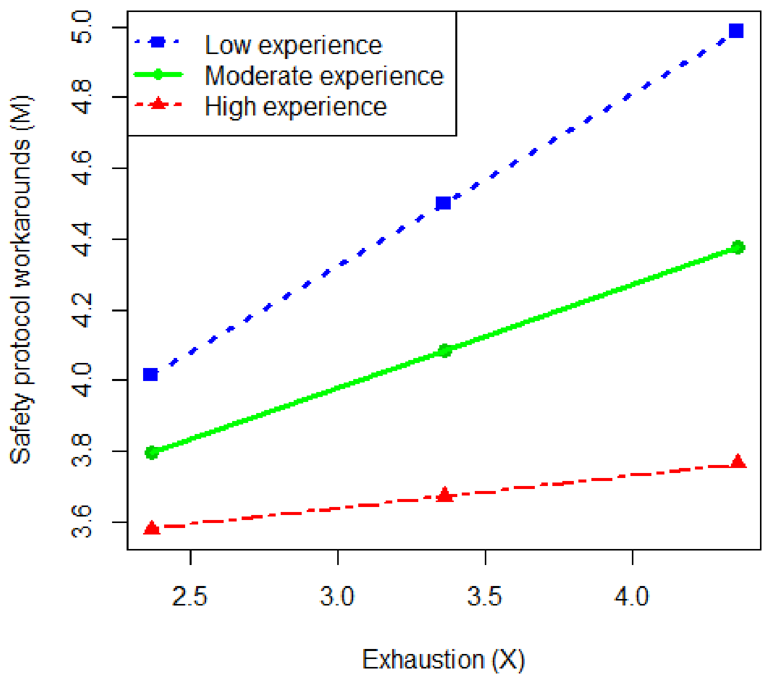
```
DATA LIST FREE/
   exhaust  tenure        safety      .
BEGIN DATA.
           2.36932.2404      4.0147
           3.36232.2404      4.5010
           4.35532.2404      4.9873
           2.36935.5133      3.7972
           3.36235.5133      4.0863
           4.35535.5133      4.3755
           2.36938.7863      3.5798
           3.36238.7863      3.6717
           4.35538.7863      3.7637
END DATA.
GRAPH/SCATTERPLOT=
 exhaustWITH                       safety BY     tenure .
```

```
data;
input exhaust tenure safety;
datalines;
   2.36932.24044.0147
   3.36232.24044.5010
   4.35532.24044.9873
   2.36935.51333.7972
   3.36235.51334.0863
   4.35535.51334.3755
   2.36938.78633.5798
   3.36238.78633.6717
   4.35538.78633.7637
run;
proc sgplot;reg x=exhaust y=safety/group=tenure;run;
```

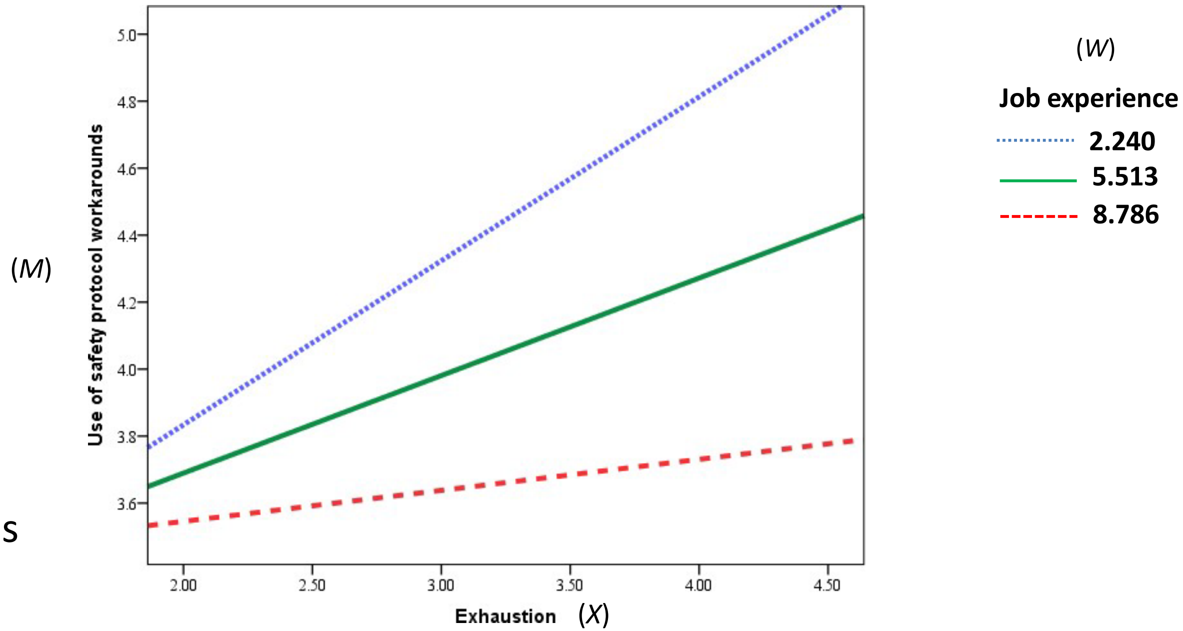
In R

```
x<-c(2.369,3.362,4.355,2.369,3.362,4.355,2.369,3.362,4.355)
w<-c(2.240,2.240,2.240,5.513,5.513,5.513,8.786,8.786,8.786)
m<-c(4.015,4.501,4.987,3.797,4.086,4.376,3.580,3.672,3.764)
wmarker<-c(15,15,15,16,16,16,17,17,17)
plot(y=m,x=x,cex=1.2,pch=wmarker,
xlab="Exhaustion (X)",
ylab="Safety protocol workarounds (M)",col=c(4,4,4,3,3,3,2,2,2))
legend.txt<-c("Low experience",
"Moderate experience", "High experience")
legend("topleft", legend = legend.txt,cex=1,
lty=c(3,1,6),lwd=c(2,3,2),pch=c(15,16,17),
col=c("blue","green","red"))
lines(x[w==2.240],m[w==2.240],lwd=3,lty=3,col="blue")
lines(x[w==5.513],m[w==5.513],lwd=3,col="green")
lines(x[w==8.786],m[w==8.786],lwd=2,lty=6,col="red")
```

From PROCESS plot output



PROCESS output: The conditional effects of *X* on *M*



$\theta_{X \to M}$ is the slopes of these lines.

```
Focal predict: exhaust (X)
      Mod var: tenure   (W)
```

$$\theta_{X \to M} = a_1 + a_3 W = 0.626 - 0.061W$$

```
Conditional effects of the focal predictor at values of the moderator(s):
```

tenure	Effect	se	t	p	LLCI	ULCI
2.2404	.4897	.0800	6.1224	.0000	.3323	.6472
5.5133	.2912	.0555	5.2504	.0000	.1820	.4003
8.7863	.0926	.0734	1.2619	.2080	-.0518	.2371

PROCESS output

$$\widehat{Y}_i = -0.593 + 0.089X_i + 0.491M_i + 1.618Z_i - 0.415M_iZ_i + 0.276U_i$$

```
OUTCOME VARIABLE:
 injury

 Model Summary
          R       R-sq        MSE          F        df1        df2          p
      .4611      .2126      .7580    15.8748     5.0000   294.0000      .0000

Model
                 coeff         se          t          p       LLCI       ULCI
constant        -.5928      .3208    -1.8478      .0656    -1.2242      .0386
exhaust          .0892      .0536     1.6625      .0975     -.0164      .1948
safety           .4192      .0674     6.2216      .0000      .2866      .5518
sex             1.6176      .4236     3.8183      .0002      .7838     2.4513
Int_1           -.4152      .0994    -4.1782      .0000     -.6107     -.2196
injuryb          .2757      .0508     5.4237      .0000      .1756      .3757

 Product terms key:
  Int_1    :        safety   x        sex

 Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
 M*Z     .0468    17.4576     1.0000   294.0000      .0000
```

$\leftarrow b_1$

$\leftarrow b_3$

Using the plot option, PROCESS provides what we need to visualize the model of *Y*

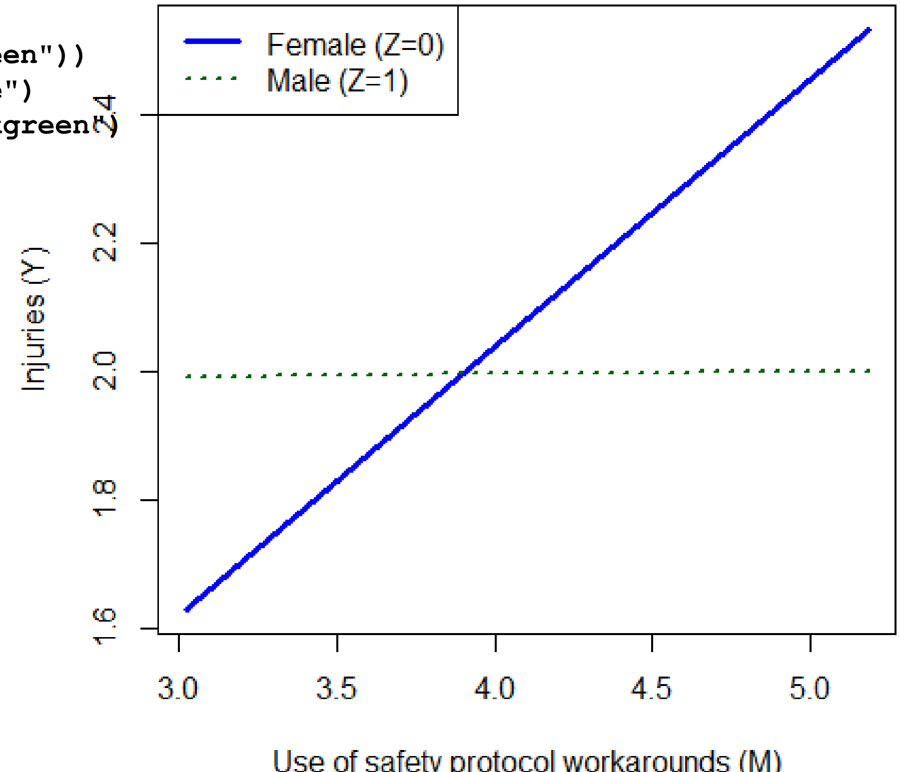
```
DATA LIST FREE/
  safety sex                    injury      .
BEGIN DATA.
   3.0227 .0000                  1.6285
   4.1020 .0000                  2.0809
   5.1813 .0000                  2.5334
   3.0227 1.0000                 1.9912
   4.1020 1.0000                 1.9955
   5.1813 1.0000                 1.9999
END DATA.
GRAPH/SCATTERPLOT=
 safety     WITH     injury    BY         sex

.
```

```
data;
input safety sex injury;
datalines;
        3.0227       .0000      1.6285
        4.1020       .0000      2.0809
        5.1813       .0000      2.5334
        3.0227      1.0000      1.9912
        4.1020      1.0000      1.9955
        5.1813      1.0000      1.9999
run;
proc sgplot;reg x=safety
y=injury/group=sex;run;
```

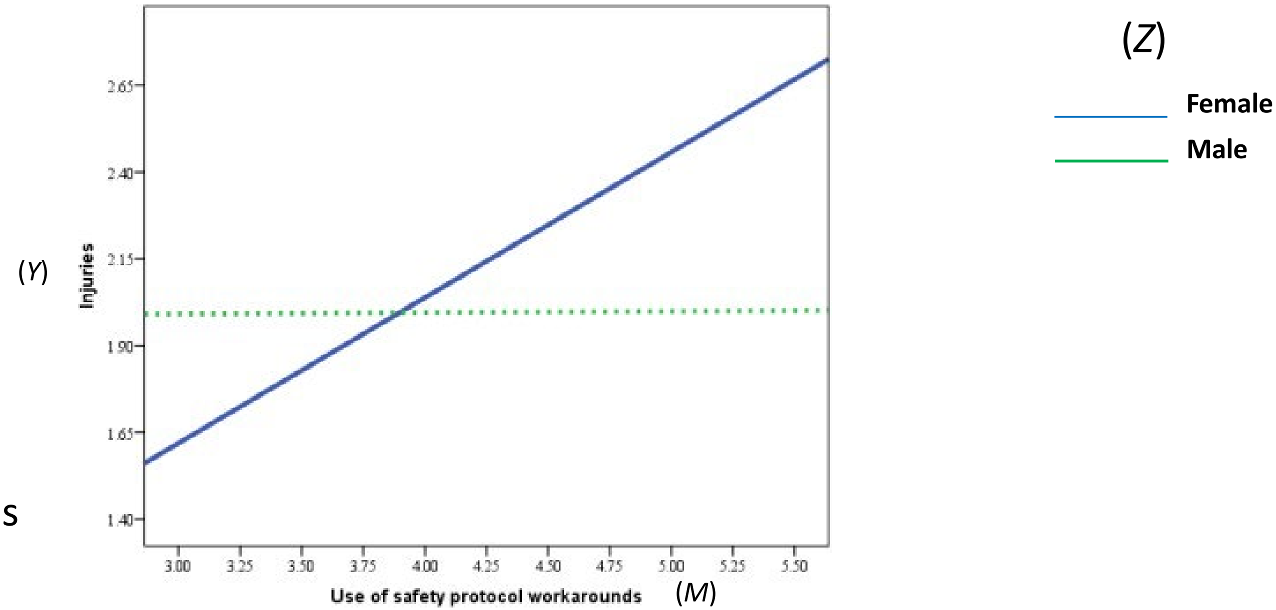
In R

```
z<-c(0,0,0,1,1,1)
m<-c(3.023,4.102,5.181,3.023,4.102,5.181)
y<-c(1.629,2.081,2.533,1.991,1.996,2.000)
plot(y=y,x=m,pch=15,col="white",
xlab="Use of safety protocol workarounds (M)",
ylab="Injuries (Y)")
legend.txt<-c("Female (Z=0)",
"Male (Z=1)")
legend("topleft",legend=legend.txt,
lty=c(1,3),lwd=c(3,2),col=c("blue","darkgreen"))
lines(m[z==0],y[z==0],lwd=3,lty=1,col="blue")
lines(m[z==1],y[z==1],lwd=2,lty=3,col="darkgreen")
```

From PROCESS plot output



PROCESS output: The conditional effects of *M* on *Y*



$\theta_{M \to Y}$ is the slopes of these lines.

```
Focal predict: safety    (M)
      Mod var: sex       (Z)
```

$$\theta_{M \to Y} = b_1 + b_3 Z = 0.419 - 0.415Z$$

```
Conditional effects of the focal predictor at values of the moderator(s):

      sex     Effect          se           t           p        LLCI       ULCI
    .0000      .4192       .0674      6.2216       .0000       .2866      .5518
   1.0000      .0040       .0753       .0536       .9573      -.1442      .1523
```

PROCESS output

The summary provides the direct effect with *p*-value and confidence interval, as well as conditional indirect effects with bootstrap confidence intervals.

```
****************** DIRECT AND INDIRECT EFFECTS OF X ON Y *****************

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .0892      .0536     1.6625      .0975     -.0164      .1948     <— c'

Conditional indirect effects of X on Y:

INDIRECT EFFECT:
 exhaust      ->      safety       ->     injury
      W            Z
     tenure          sex     Effect     BootSE   BootLLCI   BootULCI
     2.2404        .0000      .2053      .0428      .1267      .2953
     2.2404       1.0000      .0020      .0429     -.0809      .0877
     5.5133        .0000      .1221      .0260      .0739      .1749
     5.5133       1.0000      .0012      .0253     -.0484      .0520
     8.7863        .0000      .0388      .0313     -.0286      .0962
     8.7863       1.0000      .0004      .0098     -.0214      .0213
```

$$\theta_{X\to M}\theta_{M\to Y} = (a_1 + a_3W)(b_1 + b_3Z)$$
$$= a_1b_1 + a_3b_1W + a_1b_3Z + a_3b_3WZ$$
$$= 0.2623 - 0.0254W - 0.2597Z + 0.0252WZ$$

$a_1 = 0.626$
$a_3 = -0.061$
$b_1 = 0.419$
$b_3 = -0.415$

The conditional indirect effects

```
Conditional indirect effects of X on Y:

INDIRECT EFFECT:
 exhaust       ->      safety       ->     injury
         W             Z
     tenure          sex       Effect      BootSE    BootLLCI    BootULCI
     2.2404        .0000        .2053       .0428      .1267       .2953
     2.2404       1.0000        .0020       .0429     -.0809       .0877
     5.5133        .0000        .1221       .0260      .0739       .1749
     5.5133       1.0000        .0012       .0253     -.0484       .0520
     8.7863        .0000        .0388       .0313     -.0286       .0962
     8.7863       1.0000        .0004       .0098     -.0214       .0213
```

Tenure (W)	Sex (Z)	$\theta_{X \to M}$	$\theta_{M \to Y}$	$\theta_{X \to M}\theta_{M \to Y}$
2.240	0	0.490	0.419	0.205
2.240	1	0.490	0.001	0.002
5.513	0	0.291	0.419	0.122
5.513	1	0.291	0.001	0.001
8.786	0	0.093	0.419	0.039
8.786	1	0.093	0.001	0.000

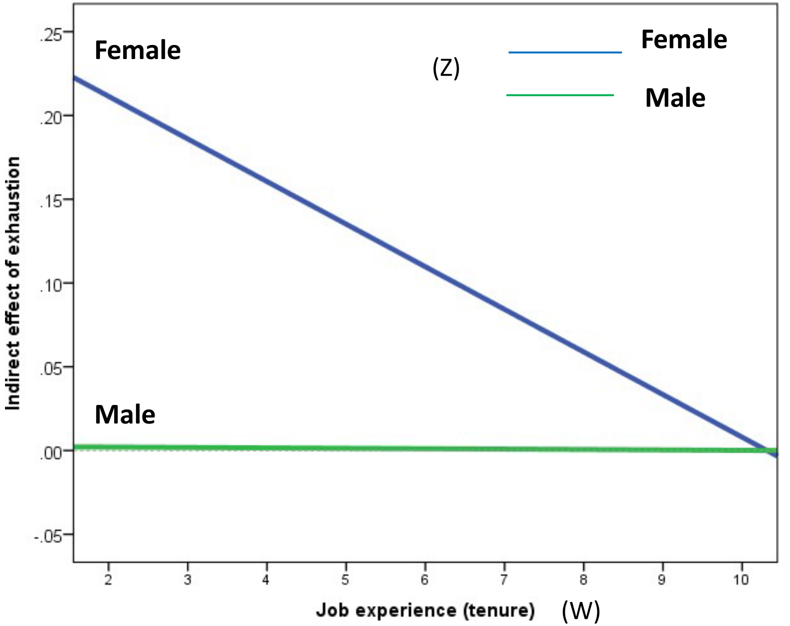
The effect of exhaustion on the use of safety work-arounds

The effect of the use of safety work-arounds on injuries.

Two visual depictions of the indirect effect

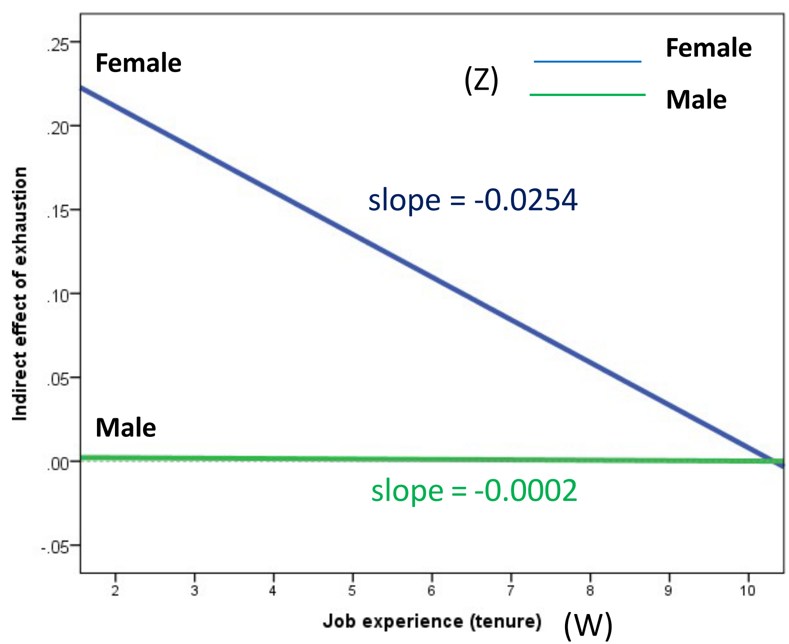
$$(a_1 + a_3W)(b_1 + b_3Z) = a_1b_1 + a_3b_1W + a_1b_3Z + a_3b_3WZ$$

$$= 0.2623 - 0.0254W - 0.2597Z + 0.0252WZ$$



The relationship between one moderator and the size of the indirect effect depends on the second moderator. The function defining the indirect effect can be used to quantify the moderation of mediation by one moderator at a level of the second moderator.

Two visual depictions of the indirect effect



$$a_1b_1 + a_3b_1W + a_1b_3Z + a_3b_3WZ$$
$$0.2623 - 0.0254W - 0.2597Z + 0.0252WZ$$

can be written in the form

$$a_1b_1 + (a_3b_1 + a_3b_3Z)\ W + a_1b_3Z$$
$$0.2623 + (-0.0254 + 0.0252Z)W - 0.2597Z$$

which shows that the relationship between W and the size of the indirect effect of X depends on Z:

$$a_3b_1 + a_3b_3Z$$
$$-0.0254 + 0.0252Z$$

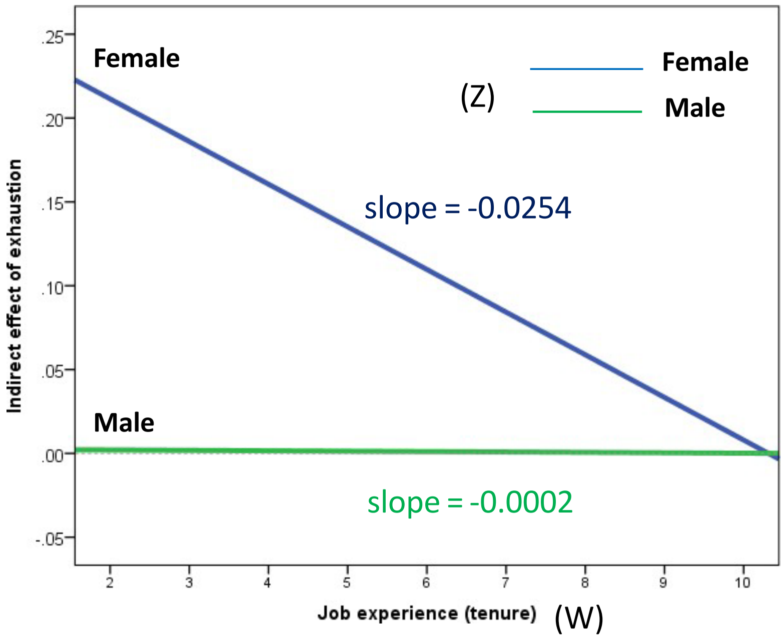
This function yields two values, one for females ($Z = 0$), and one for males ($Z = 1$)

Females: $-0.0254 + 0.0252Z = -0.0254 + 0.0252(0) = -0.0254$

Males: $-0.0254 + 0.0252Z = -0.0254 + 0.0252(1) = -0.0002$

These are the slopes of the lines above. Each line depicts how the indirect effect of X is moderated by W. Do these lines differ? If so, then the moderation of mediation of X's effect by W is moderated by Z: *moderated moderated mediation*.

PROCESS output



As Z changes by one unit, the slope of the line relating W to the indirect effect of X changes by a_3b_3 units. This is the index of *moderated moderated mediation*.

$$[a_3b_1 + a_3b_3(Z+1)] - (a_3b_1 + a_3b_3Z)$$

$$= a_3b_1 + a_3b_3Z + a_3b_3 - a_3b_1 - a_3b_3Z$$

$$= a_3b_3$$

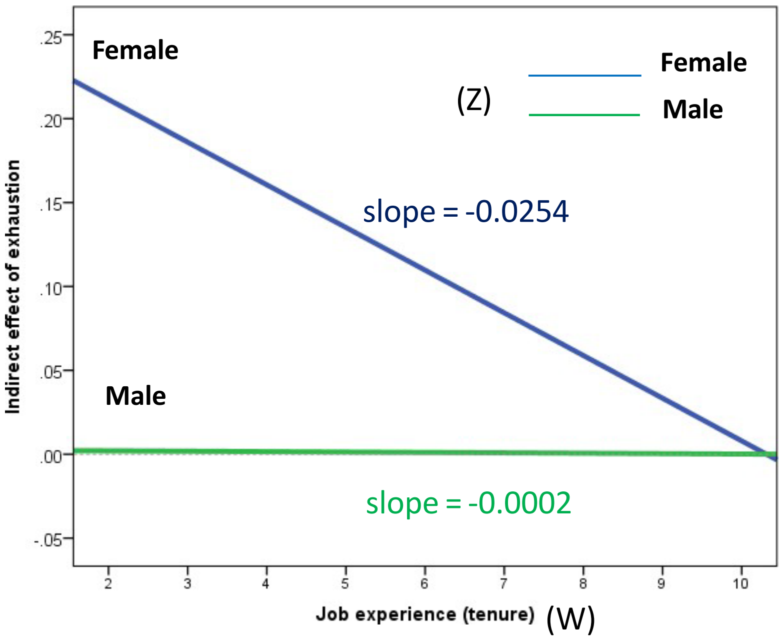
$$a_3b_1 + a_3b_3Z$$

$$-0.0254 + 0.0252Z$$

If the index of moderated mediation is different from zero, we can say the moderation of the indirect effect of X by W depends on Z. A bootstrap confidence interval provides the test. PROCESS does this:

```
Index of moderated moderated mediation
     Index      BootSE     BootLLCI    BootULCI
     .0252       .0097        .0099       .0475
```

Conditional moderated mediation



With evidence of moderation by Z of the moderation of the indirect effect of X by W, we can "probe" this moderation of moderated mediation. The function defining the moderation of the indirect effect of X by W is

$$a_3b_1 + a_3b_3Z$$

$$-0.0254 + 0.0252Z$$

which I call the "index of conditional moderated mediation". Recall that it generates the two slopes here. We can test each slope with a bootstrap confidence interval.

```
Indices of conditional moderated mediation by W
      sex      Index     BootSE   BootLLCI   BootULCI
    .0000     -.0254      .0083     -.0443     -.0116
   1.0000     -.0002      .0055     -.0119      .0104
```

The indirect effect of exhaustion is moderated by job tenure among women, but not among men.

Moderation of moderated mediation is symmetrical

$$a_1b_1 + a_3b_1W + a_1b_3Z + a_3b_3WZ$$
$$0.2623 - 0.0254W - 0.2597Z + 0.0252WZ$$

can be written in the form

$$a_1b_1 + (a_1b_3 + a_3b_3W)Z + a_3b_1W$$
$$0.2623 + (-0.2597 + 0.0252W)Z - 0.0254W$$

which shows that the relationship between Z and the size of the indirect effect of X depends on W:

$$a_1b_3 + a_3b_3W$$
$$-0.2597 + 0.0252W$$

Pick some values of W and do the math:

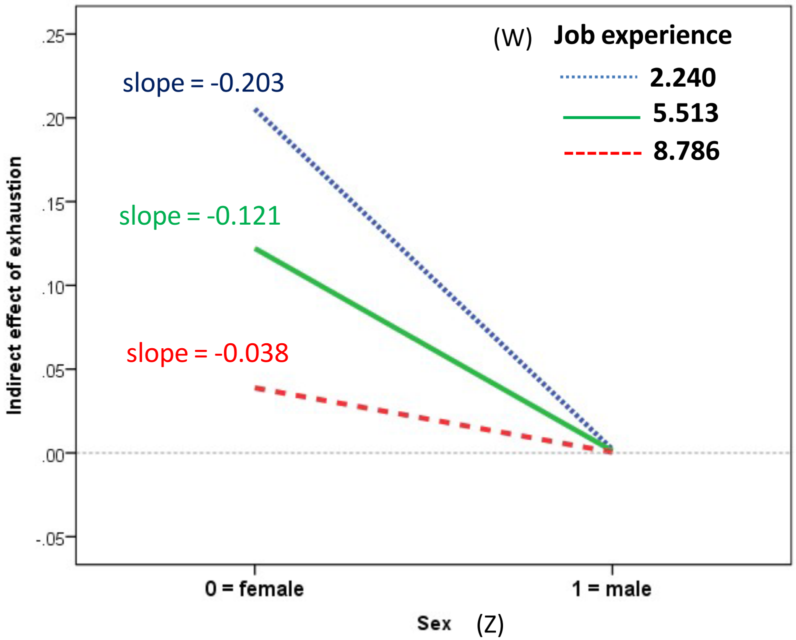
Relatively less experience: $-0.2597 + 0.0252W = -0.2597 + 0.0252(2.240) = -0.203$

Moderate experience: $-0.2597 + 0.0252W = -0.2597 + 0.0252(5.513) = -0.121$

Relatively high experience: $-0.2597 + 0.0252W = -0.2597 + 0.0252(8.876) = -0.038$

These are the slopes of the lines above. Each line depicts how the indirect effect of X is moderated by Z (i.e., differs between men and women) at a specific value of W. Do these lines differ? If so, then the moderation of mediation of X's effect by Z is moderated by W.

PROCESS output



As W changes by one unit, the slope of the line relating Z to the indirect effect of X changes by a_3b_3 units. This is the index of *moderated moderated mediation*.

$$[a_1b_3 + a_3b_3(W+1)] - (a_1b_3 + a_3b_3W)$$

$$= a_1b_3 + a_3b_3W + a_3b_3 - a_1b_3 - a_3b_3W$$

$$= a_3b_3$$

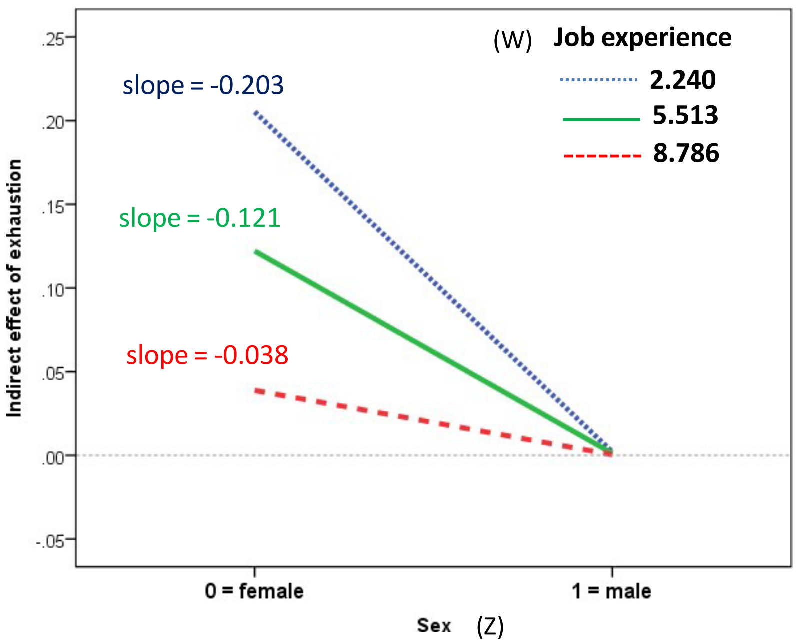
$$a_1b_3 \quad + \quad a_3b_3W$$

$$-0.2597 + 0.0252W$$

If the index of moderated mediation is different from zero, we can say the moderation of the indirect effect of X by Z depends on W. A bootstrap confidence interval provides the test. PROCESS does this. But we already know the answer.

```
Index of moderated moderated mediation
     Index      BootSE    BootLLCI    BootULCI
     .0252       .0097       .0099       .0475
```

Conditional moderated mediation



With evidence of moderation by *W* of the moderation of the indirect effect of *X* by *Z*, we can "probe" this moderation of moderated mediation. The function defining the moderation of the indirect effect of *X* by *Z* is

$$a_1 b_3 + a_3 b_3 W$$

$$-0.2597 + 0.0252W$$

which I call the "index of conditional moderated mediation". Pick some values of *W*, construct the index, and conduct an inference with a bootstrap confidence interval. PROCESS does this.

```
Indices of conditional moderated mediation by W
```

Indices of conditional moderated mediation by ~~W~~ Z

tenure	Index	BootSE	BootLLCI	BootULCI
2.2404	-.2033	.0587	-.3313	-.1013
5.5133	-.1209	.0351	-.1950	-.0583
8.7863	-.0385	.0320	-.1014	.0263

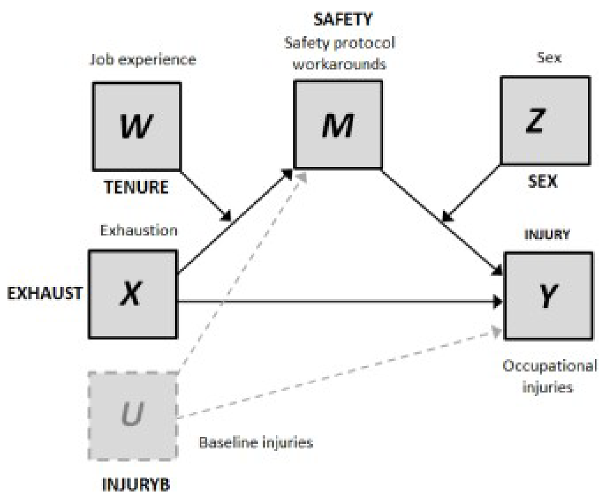
The indirect effect of exhaustion differs between men and women except among those with relatively more job experience.

THIS IS **NOT** IN THE PROCESS OUTPUT.

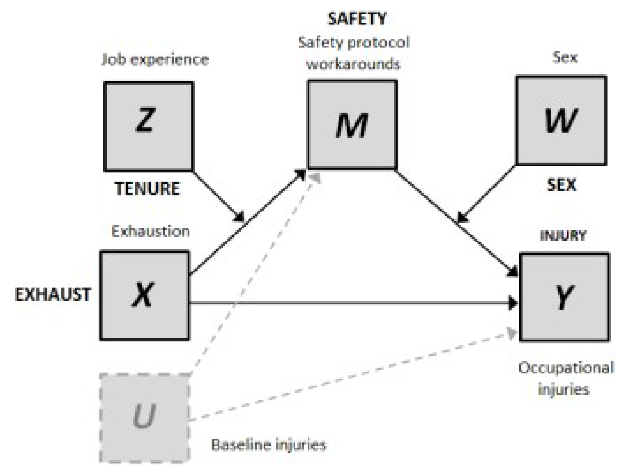
Conditional moderated mediation

PROCESS generates the indices of conditional moderated mediation of the effect of *X* on *Y* by *W* at values of *Z*. If you want the indices of conditional moderated mediation of the effect of *X* on *Y* by *Z* at values of *W*, you need to reverse *W* and *Z* in the model.

This is model 21



This model is not preprogrammed.



We need to reprogram/edit model 21, reversing what we call *W* and what we call *Z*:

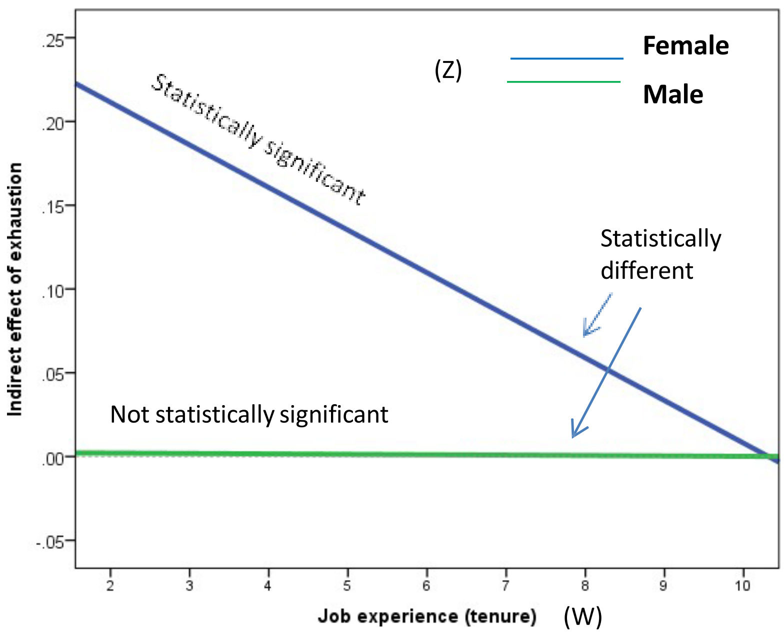
```
process y=injury/x=exhaust/m=safety/z=tenure/w=sex/cov=injuryb/moments=1/model=21/seed=456/
   wmatrix=0,0,1/zmatrix=1,0,0.
```

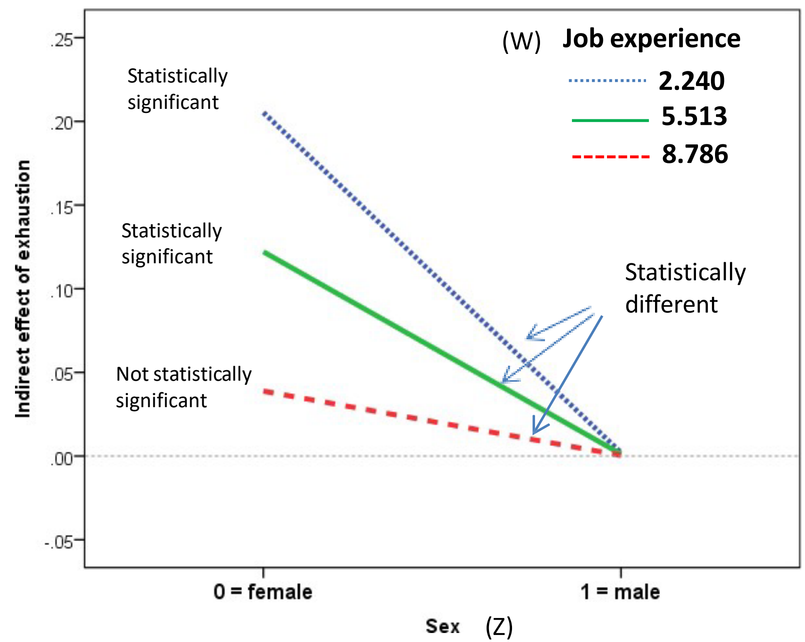
```
%process (data=injuries,y=injury,x=exhaust,m=safety,z=tenure,w=sex,cov=injuryb,moments=1,seed=456,
   model=21,wmatrix=0 0 1,zmatrix=1 0 0);
```

```
process(data=injuries,y="injury",x="exhaust",m="safety",z="tenure",w="sex",cov="injuryb",moments=1,
   seed=456,model=21,wmatrix=c(0,0,1),zmatrix=c(1,0,0))
```

Bringing it all together

```
***************** DIRECT AND INDIRECT EFFECTS OF X ON Y *****************

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
      .0892      .0536     1.6625      .0975     -.0164      .1948
Conditional indirect effects of X on Y:

INDIRECT EFFECT:
 exhaust     ->    safety     ->     injury

    tenure          sex     Effect     BootSE   BootLLCI   BootULCI
      2.2404      .0000      .2053      .0428      .1267      .2953
      2.2404     1.0000      .0020      .0429     -.0809      .0877
      5.5133      .0000      .1221      .0260      .0739      .1749
      5.5133     1.0000      .0012      .0253     -.0484      .0520
      8.7863      .0000      .0388      .0313     -.0286      .0962
      8.7863     1.0000      .0004      .0098     -.0214      .0213
```

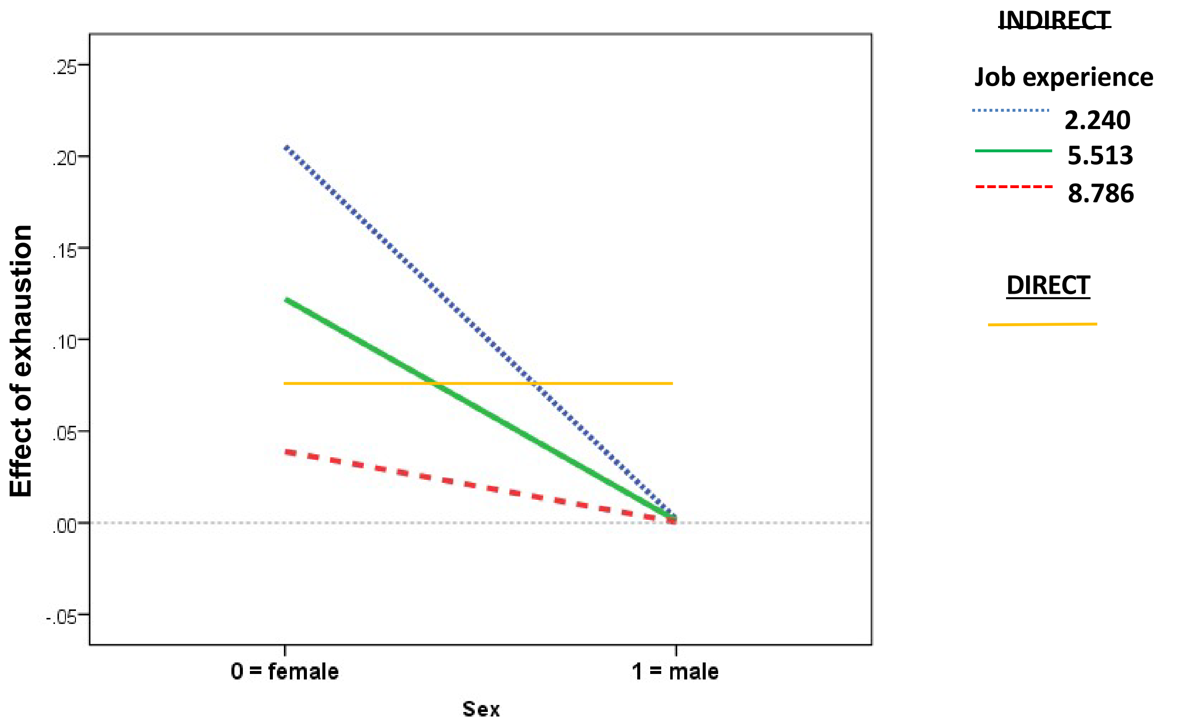




Bringing it all together

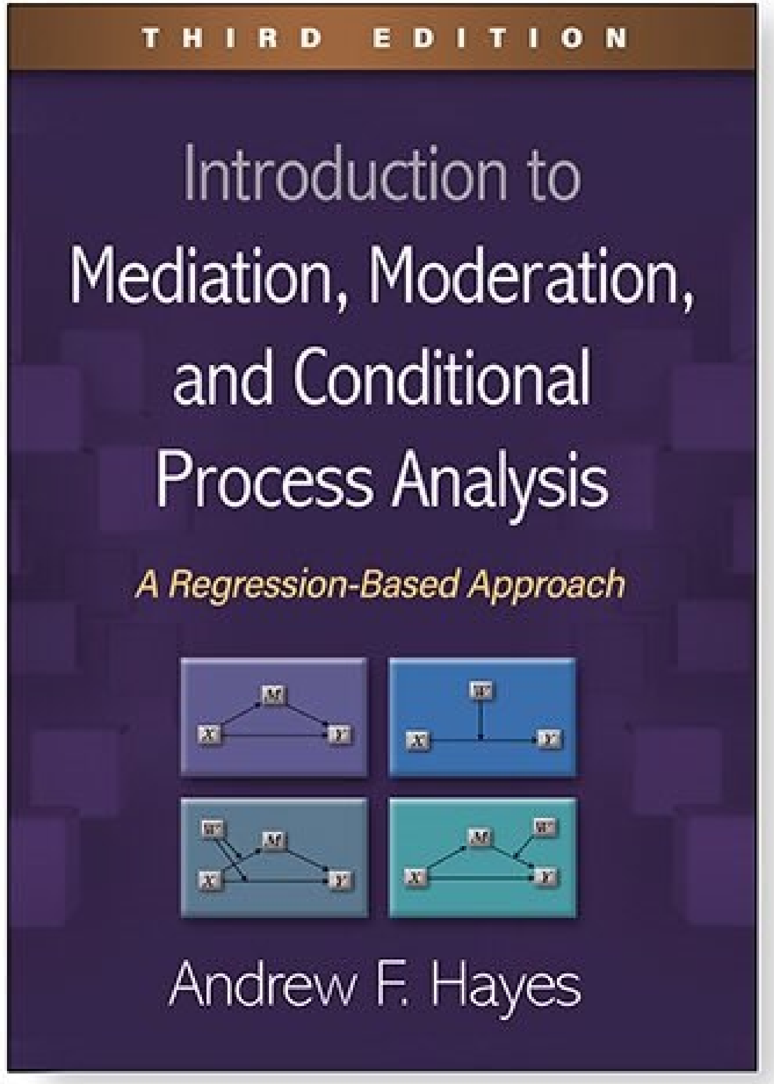
Exhaustion significantly increases injuries by increasing the use of safety-protocol workarounds which in turn increases injuries, but only among women. Among women the magnitude of this effect is more pronounced among those with less experience on the job. Among men, as well as women with more experience, this process does not appear to be operating. Independent of this process, exhaustion does not significantly affect workplace injury.

Bringing it all together



Workplace exhaustion increases injuries by increasing the use of safety-protocol workarounds which in turn increases injuries, but only among women. There is a statistically significant sex difference in the magnitude of this mechanism, with the difference being larger among workers with relatively less job experience. Among workers with relatively more experience, there is no sex difference in this indirect effect. Independent of this process, there is no additional effect of exhaustion on injuries.

Where to learn more…



Pertinent publications

Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist, 64*, 19-54.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd Ed). New York: Guilford Press.

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs, 85*, 4-40.

Hayes, A. F., & Rockwood, N. J. (2017). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour Research and Therapy, 98*, 39-57.

Hayes, A. F., Montoya, A. K., & Rockwood, N. J. (2017). The analysis of mechanisms and their contingencies: PROCESS versus structural equation modeling. *Australasian Marketing Journal, 25,* 76-81.

Montoya, A. K., & Hayes, A. F. (2017). Two condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods, 22*, 6-27.

Hayes, A. F. (2015). *An index and test of linear moderated mediation*. *Multivariate Behavioral Research, 50,* 1-22.

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.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science, 24*, 1918-1927.

Pertinent publications

Hayes, A. F., & Preacher, K. J. (2013). Conditional process modeling: Using structural equation modeling to examine contingent causal processes. In G. R. Hancock & R. O. Mueller (Eds.) *Structural equation modeling: A second course* (2nd Ed). Greenwich, CT: Information Age Publishing.

Hayes, A. F., Glynn, C. J., & Huge, M. E. (2012). Cautions regarding the interpretation of regression coefficients and hypothesis tests in linear models with interactions. *Communication Methods and Measures, 6,* 1-11.

Hayes, A. F., Preacher, K. J., & Myers, T. A. (2011). Mediation and the estimation of indirect effects in political communication research. In E. P. Bucy & R. L. Holbert (Eds), *Sourcebook for political communication research: Methods, measures, and analytical techniques.* (p. 434-465). *New York:* Routledge.

Hayes, A. F., & Preacher, K. J. (2010). Estimating and testing indirect effects in simple mediation models when the constituent paths are nonlinear. *Multivariate Behavioral Research, 45,* 627-660.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millenium. *Communication Monographs, 76,* 408-420.

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

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

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