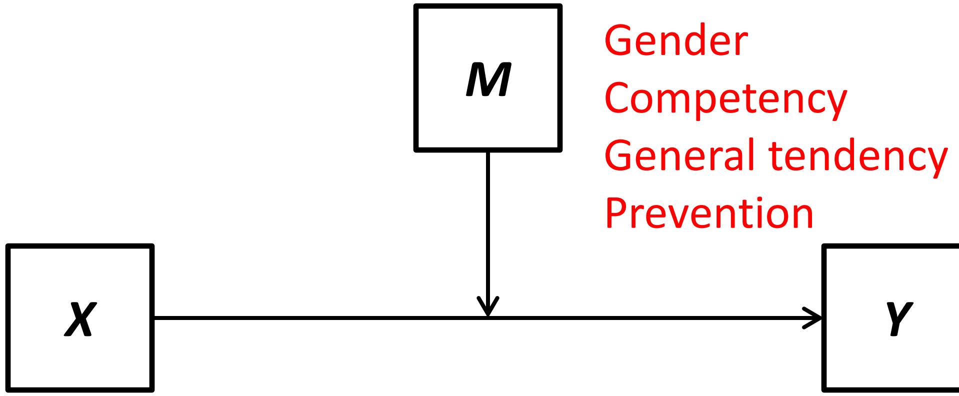
Moderation, Mediation, and Conditional Process Analysis in R

Amanda Kay Montoya
PhD Candidate
Ohio State University

University of Washington
Department of Psychology
May 17, 2017

Moderation



Gender
Competency
General tendency toward Promotion or Prevention

Need for Support

Relative Promotion vs. Prevention

Moderation is the idea that the relationship between a focal predictor (*X*) and an outcome (*Y*) may depend on some other variable (*M*).

This can be described as a *contingent relationship* or an *interaction*.

The idea being that the magnitude or the sign (or both) of the relationship between *X* and *Y* depends on *M*.

Many different kinds of variables may act as moderators: experimental condition, individual differences, demographics, physiological variables, cognitive variables, environmental variables, etc.

A quick example: Name some possible moderators!

Modeling Non-Contingent Relationships

A multiple regression model without interaction terms, fixes the relationship between the predictors and the outcomes to be the same regardless of the level of other predictors.

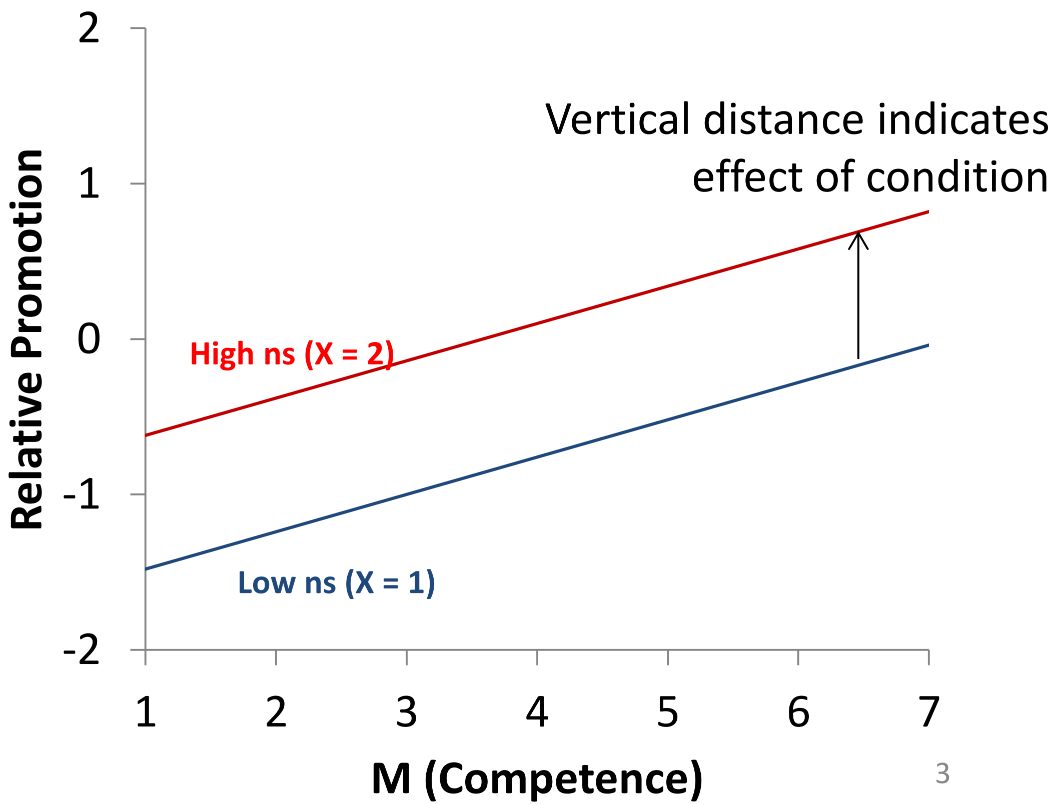
$$Y_i = b_0 + b_1 X_i + b_2 M_i$$

The effect of condition (1 unit increase in nscond) results in a .86 unit increase in relative promotion, regardless of competence.

Example:
Y: Relative Promotion(-6 – 6)
X: NS Condition (1 or 2)
M: Competence (1 – 7)

```
vaughnlm2 <-
lm(relativepromotion ~ nscond +
competence, data = vaughndata)
summary(vaughnlm2)
```

$$Y_i = -2.57 + 0.86_1 X_i + 0.24 M_i$$



Modeling Contingent Relationships

What if instead we felt that the relationship between Condition and Relative Promotion depends on Competence? Thus the relationship between need for support and relative promotion is a *function* of competence.

$$Y_i = b_0 + f(M_i)X_i + b_2M_i$$

One popular model for $f(M_i)$ is a linear model:

$$f(M_i) = b_1 + b_3M_i = \theta_{X \to Y|M}$$

This way we can rewrite the model:

$$Y_i = b_0 + \theta_{X \to Y|M}X_i + b_2M_i$$

$$Y_i = b_0 + (b_1 + b_3M_i)X_i + b_2M_i$$

$$Y_i = b_0 + b_1X_i + b_2M_i + b_3M_iX_i$$

This is a regression model which can be estimated, where the significance of b_3 reflects whether the relationship between *X* and *Y* is linearly dependent on *M.*

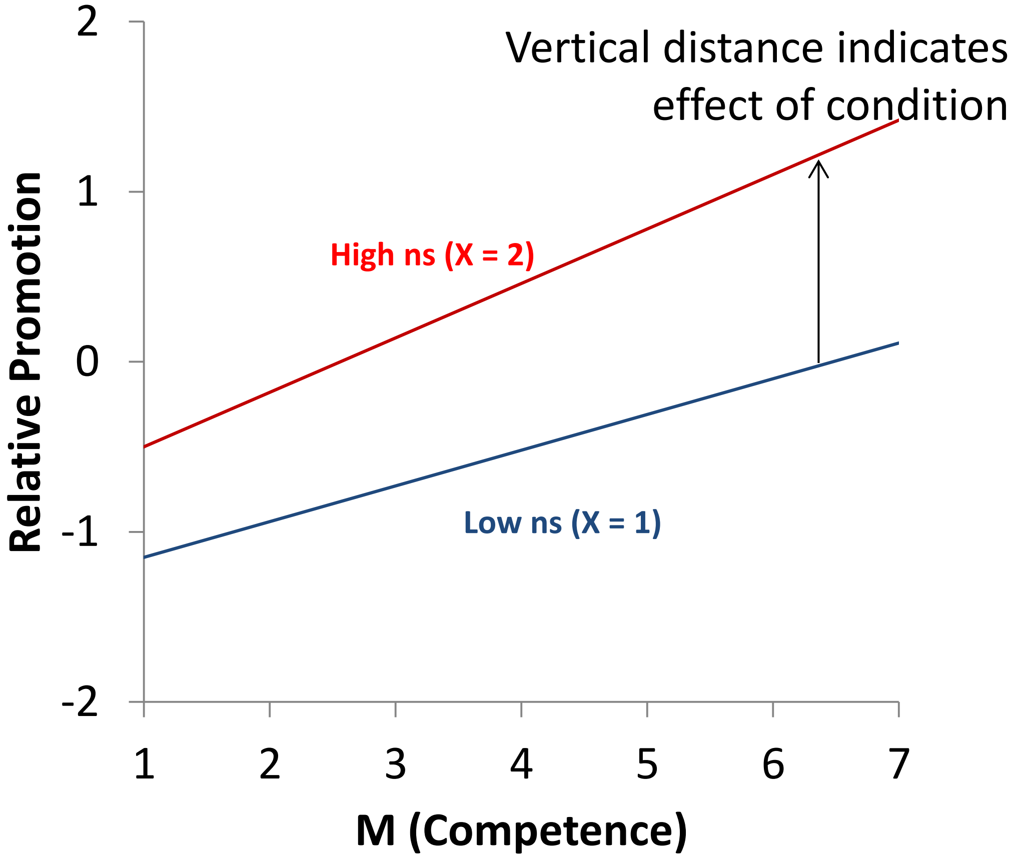
Modeling Contingent Relationships

What if instead we felt that the relationship between Condition and Relative Promotion depends on Competence?

$$Y_i = b_0 + (b_1 + b_3 M_i)X_i + b_2 M_i$$

$$Y_i = -1.90 + (0.34 + 0.11_3 M_i)X_i + 0.10 M_i$$

```
vaughnlm3 <-
lm(relativepromotion ~
nscond*competence, data =
vaughndata)
summary(vaughnlm3)
```



Interpreting Coefficients

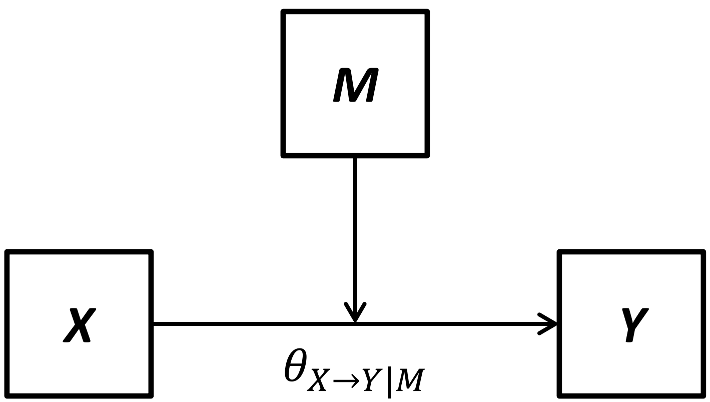
$$Y_i = b_0 + (b_1 + b_3 M_i) X_i + b_2 M_i$$

b_0 : Predicted *Y* when *X* and *M* are both zero

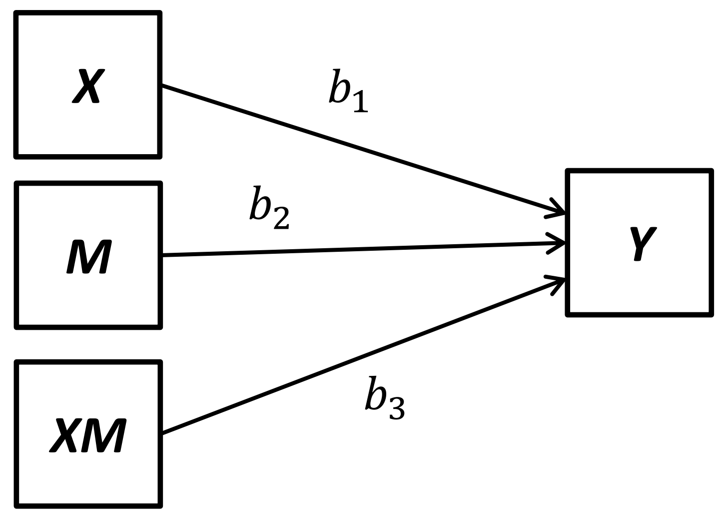
$$\hat{Y} = b_0 + (b_1 + b_3 0)0 + b_2 0 = b_0 + (b_1)0 + 0 = b_0$$

b_1 : Increase in *Y* with a one unit increase in *X* when *M* is zero

$$\hat{Y} = b_0 + (b_1 + b_3 0)X_i + b_2 0 = b_0 + (b_1)X_i + 0 = b_0 + \underbrace{b_1}_{\text{Slope of XY line when M = 0}} X_i$$



Conceptual Diagram



Statistical Diagram

Interpretting Coefficients

$$Y_i = b_0 + (b_1 + b_3 M_i)X_i + b_2 M_i$$

b_0 : Predicted *Y* when *X* and *M* are both zero

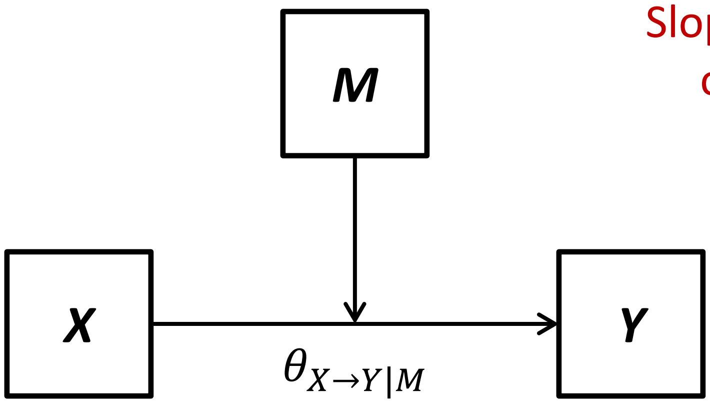
b_1 : Increase in *Y* with a one unit increase in *X* when *M* is zero

b_2 : Increase in Y with a one unit increase in *M* when *X* is zero

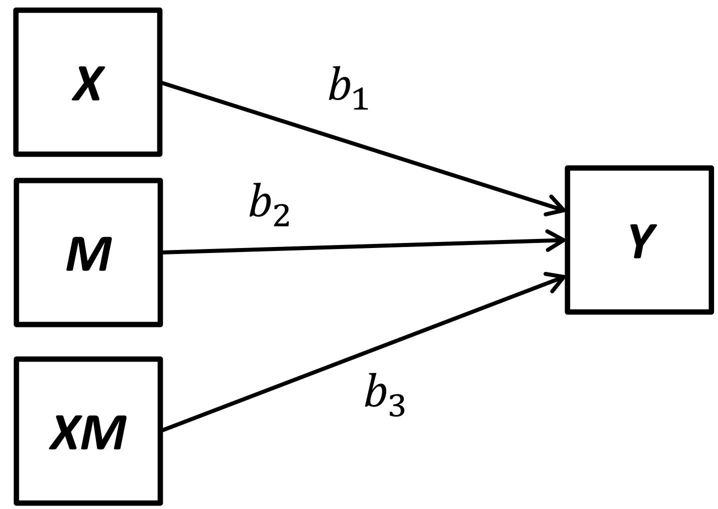
$$\hat{Y} = b_0 + (b_1 + b_3 M_i)0 + b_2 M_i = b_0 + 0 + b_2 M_i = b_0 + \overbrace{b_2}^{\text{Slope of MY line when X = 0}} M_i$$

b_3 : Increase in the relationship between *X* and *Y* with a one unit increase in *M*

$$f(M_i) = b_1 + \underbrace{b_3}_{\text{Slope of effect of X on Y}} M_i = \theta_{X \to Y|M}$$



Conceptual Diagram



Statistical Diagram

Symmetry in Moderation

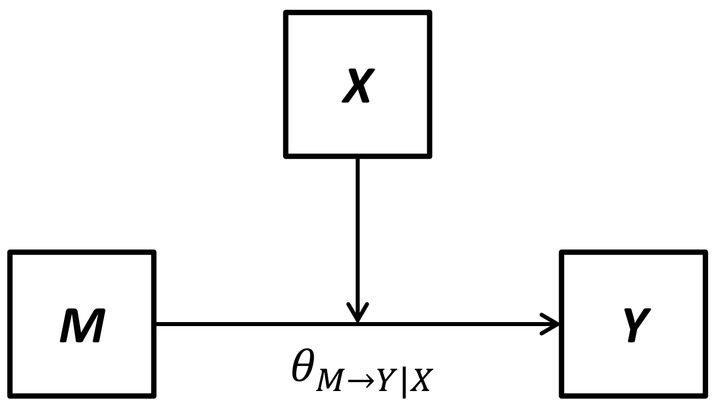
$$Y_i = b_0 + b_1 X_i + b_2 M_i + b_3 M_i X_i$$

We saw that this model can be expressed such that it is clear that *X*'s effect on *Y* depends on *M*

$$Y_i = b_0 + (b_1 + b_3 M_i) X_i + b_2 M_i$$

But it can also be equivalently expressed that *M*'s effect on *Y* depends on *X*

$$Y_i = b_0 + (b_2 + b_3 X_i) M_i + b_1 X_i$$



Here *X* moderates the effect of *M* on *Y*. *X* is the moderator, with the conditional effect of *M* on *Y* given *X* expressed as $\theta_{M \to Y|X}$. Which variable to think of as the moderator is not a mathematical concern, but rather a substantive research concern. These two models are mathematically equivalent.

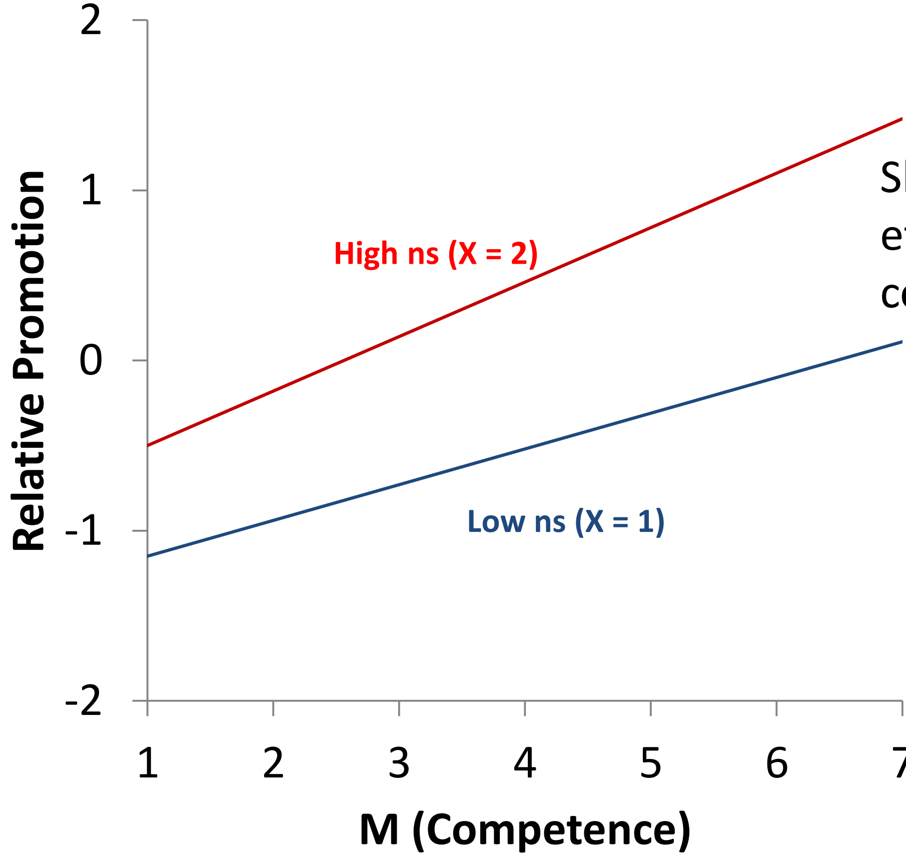
Symmetry in Moderation

What if instead we felt that the relationship between **competence** and **relative promotion** depends on **condition**?

$$Y_i = b_0 + (b_2 + b_3 X_i) M_i + b_1 X_i$$

$$Y_i = -1.90 + (0.10_2 + 0.11 X_i) M_i + 0.34 X_i$$

The effect of competence on perceived relative promotion is b_3 units higher in the high need support condition than the low need support condition.



Slope of the line is effect of competence

Estimation in R

```
vaughnlm.gend <- lm(relativepromotion ~ DichGend*nscond,
data = vaughndata)
summary(vaughnlm.gend)
```

Probing an Interaction: The "Pick-a-Point" Approach

$$Y_i = b_0 + b_1 X_i + b_2 M_i + b_3 M_i X_i$$

Select a value of the moderator (*M*) at which you'd like to have an estimate of the focal predictor variable's (*X*) effect on *Y*. Then derive its standard error. The ratio of the effect to its standard error is distributed as $t(df_{residual})$ under the null hypothesis that the effect of the focal predictor is zero at that moderator value.

We already know that

$$\theta_{X \rightarrow Y|M} = (b_1 + b_3 M_i)$$

The estimated standard error of θ is

$$s_{\theta_{X \rightarrow Y|M}} = \sqrt{(s_{b_1}^2 + 2M s_{b_1 b_3} + M^2 s_{b_3}^2)}$$

Squared standard error of b_1

Covariance of b_1 and b_3

Squared standard error of b_3

```
vcov(vaughlm3)
```

Variance Covariance Matrix for Regression Coefficients

```
vcov(vaughlm3)
```

	(Intercept)	nscond	competence	nscond:competence
(Intercept)	0.855474	-0.63048	-0.17984	0.124743
nscond	-0.63048	0.517984	0.124743	-0.09719
competence	-0.17984	0.124743	0.041779	-0.02701
nscond:competence	0.124743	-0.09719	-0.02701	0.019632

$$s_{\theta_{X\rightarrow Y|M}} = \sqrt{(s_{b_1}^2 + 2M s_{b_1 b_3} + M^2 s_{b_3}^2)}$$

Squared standard error of b_1

Covariance of b_1 and b_3

Squared standard error of b_3

Probing an Interaction: The "Pick-a-Point" Approach

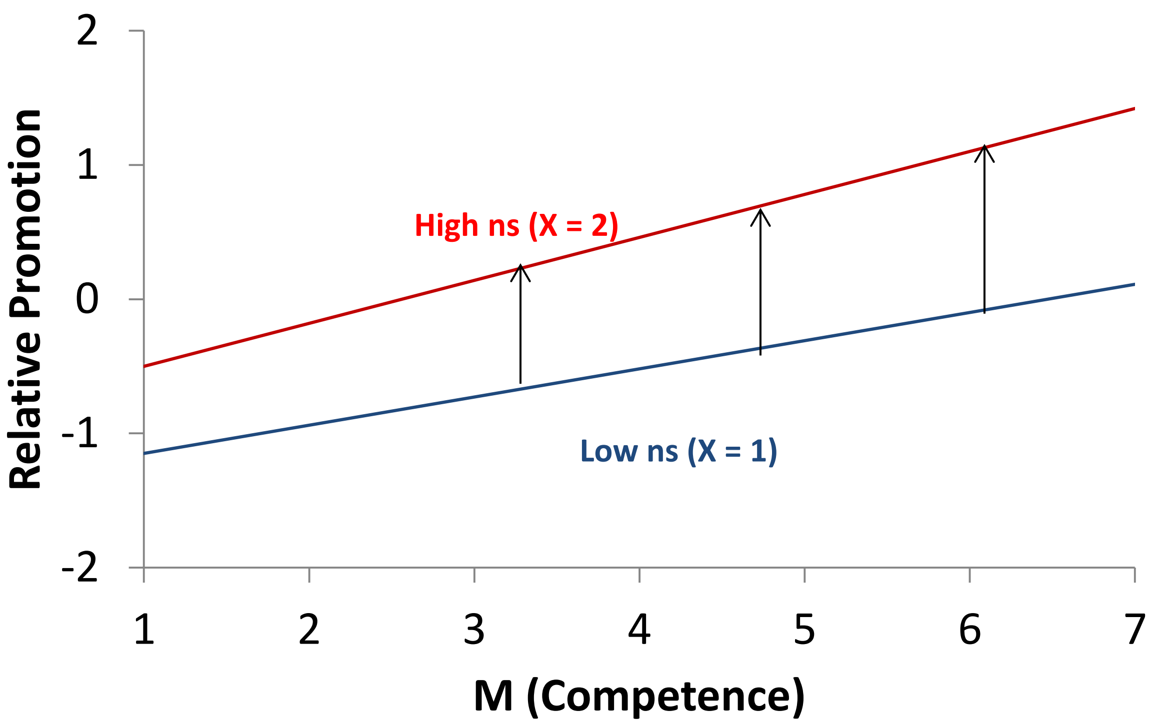
You must choose the points along the moderator to "probe" the effect of *X* on *Y*. There are some conventions for choosing to do so:

If *M* is dichotomous, choose the two coded values of *M*

If *M* is continuous, choose the Mean ± 1 SD

Let's look at an example with our protest data:

$$Y_i = -1.90 + (0.34 + 0.11_3 M_i)X_i + 0.10M_i$$



M	$\boldsymbol{\theta_{X \to Y \mid M}}$	$\boldsymbol{s_{\theta_{X \to Y \mid M}}}$	***p***
3.30	0.69	0.30	.02
4.72	0.84	0.19	<.001
6.13	0.99	0.25	<.001

Participants reported higher promotion relative to prevention in the high need support condition, at relatively low, relatively moderate, and relatively high values of competence.

Pick-a-point in R

```
# Probing Interactions in R
comp.mod.coef <- coef(vaughnlm3)
probe.contrast <- c(0, 1, 0, 3.3040)
theta.est.3.3 <- comp.mod.coef%*%probe.contrast
vcov(vaughnlm3)

se.theta.3.3 <-
sqrt(probe.contrast%*%vcov(vaughnlm3)%*%probe.contrast)
t.theta.3.3 <- theta.est.3.3/se.theta.3.3
vaughnlm3$df.residual

2*(1 - pt(t.theta.3.3, df = vaughnlm3$df.residual))
```

Practice in R

1. Data clean up: Make a new variable DichGender, based on the Gender variable. In the original variable 1 = male, 2 = female, 3 = Other, 4 = Prefer not to answer. In the new variable include all male and female respondents using 1 and 2 codes, recode 3 and 4 as missing.
2. Test if the effect condition depends on Gender.
3. Probe the effect of condition at reported gender values, and probe the effect of gender at the reported condition values.
4. Make some visualizations that clearly convey the results.

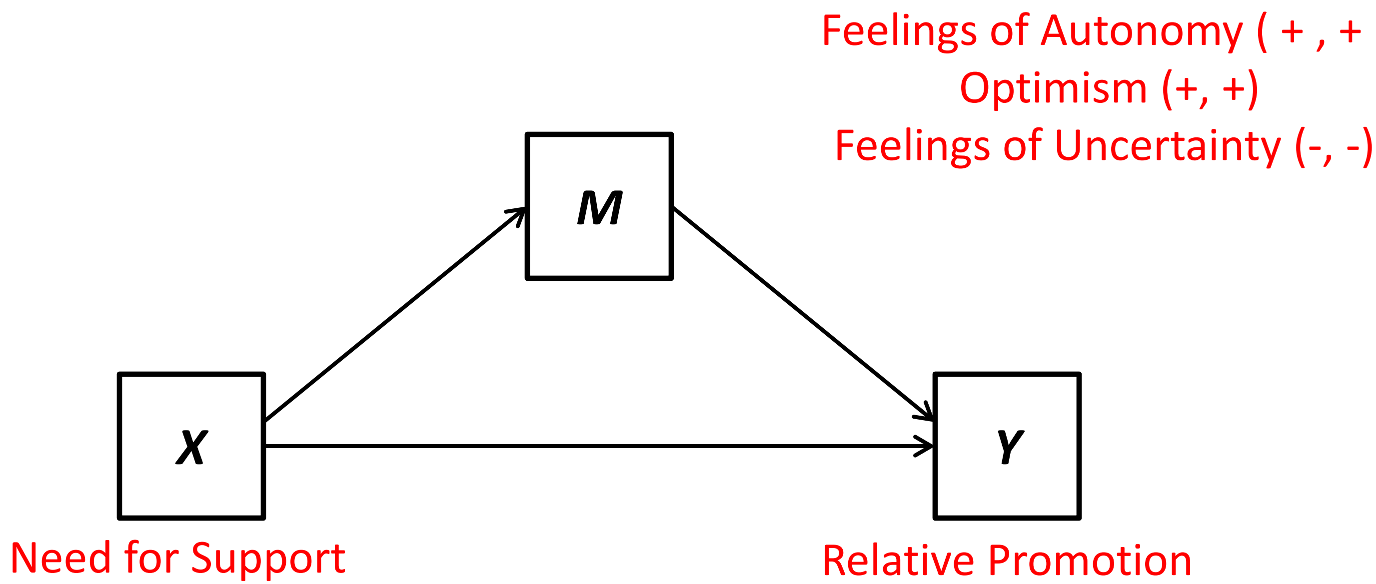
   If you can't get all the results on one plot, can you use the subset function to make two comparable plots?

5. Challenge: Can you repeat the above exercises but use the original Gender variable and include everyone instead?

   Hints: You may need to use the factor(), the lm function will create a coding scheme for your categorical variables can you figure out what scheme it uses?

Questions on Moderation?

Mediation



A simple mediation model connects an assumed causal variable (*X*) to an assumed outcome variable (*Y*), through some mechanism (*M*). The effect of *X* on *Y* can move completely through *M* or only partially through *M*.

M is frequently referred to as a *mediator* or *intermediary variable*.

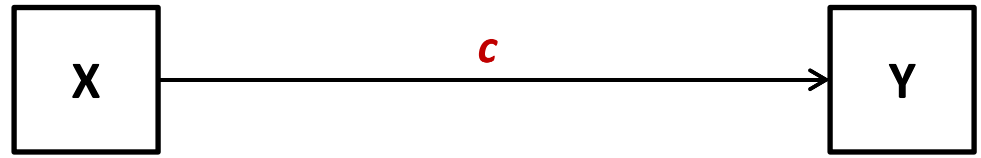
Many different kind of variables may act as mediators. Cognitive variables, individual level variables, personality variables, environmental variables, etc.

A quick example: Name some possible mediators!

Mediation: Path Analysis

Consider *a*, *b*, *c*, and *c′* to be measures of the causal effect of the variables in the mediation model.

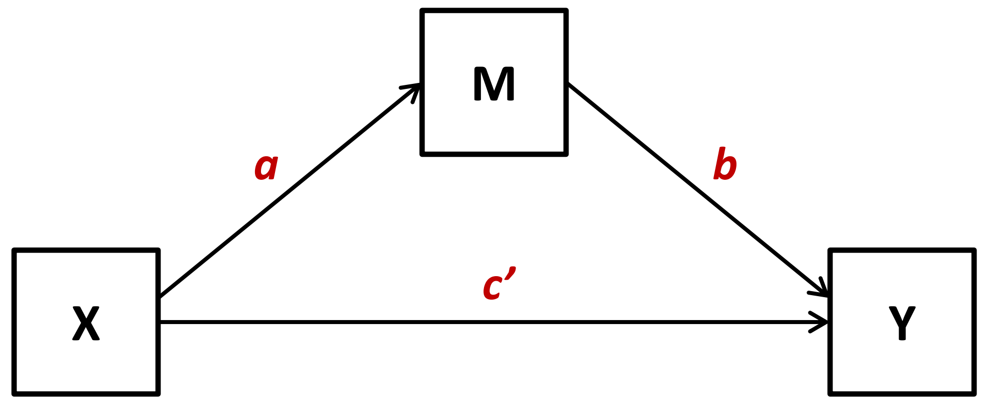
These could be measured using regression coefficients from OLS or path estimates in a structural equation model using maximum likelihood estimation.



$$Y_i = i_{Y^*} + cX_i + e_{Y_i^*}$$

$$M_i = i_M + aX_i + e_{M_i}$$

$$Y_i = i_Y + c'X_i + bM_i + e_{Y_i}$$



Direct effect of *X* on *Y* (not through *M*) = *c′*

Indirect effect of *X* on *Y* (through *M*) = *a* × *b*

Total effect = direct effect + indirect effect

c = *c′* + *a* × *b*

Indirect effect = total effect - direct effect

a × *b* = *c* - *c′*

Decomposition of effect of X on Y

$$Y_i = i_Y + c' X_i + b M_i + e_{Y_i}$$

$$M_i = i_M + a X_i + e_{M_i}$$

Insert the model for *M* into the equation for *Y*

$$Y_i = i_Y + c' X_i + b(i_M + a X_i + e_{M_i}) + e_{Y_i}$$

Group like terms

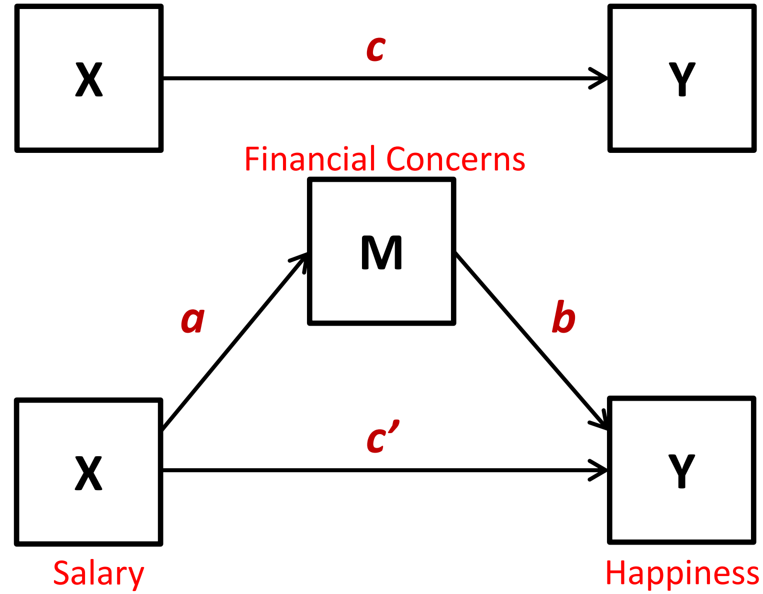
$$Y_i = \underbrace{(i_Y + b i_M)}_{i_{Y^*}} + \underbrace{(c' + ab)}_{c} X_i + \underbrace{(b e_{M_i} + e_{Y_i})}_{e_{Y_i^*}}$$

$$Y_i = i_{Y^*} + c X_i + e_{Y_i^*}$$

Total effect = direct effect + indirect effect

c = c' + $a \times b$

Path Analysis: Exercise Example



Suppose the true state of the world is such, and salary is measured in thousands of dollars per year (i.e. a one unit increase in salary corresponds to a $1000 increase in salary/year): An increase in salary of **$2,000/year** is associated with an overall increase in happiness of **3**. Suppose also that an increase in salary of **$1,000/year** is associated with a decrease in financial concerns by **2**. It is known that increasing financial concerns by **1** decreases happiness by **.5** when controlling for salary.

$$Y_i = i_{Y*} + cX_i + e_{Y_i^*}$$

$$M_i = i_M + aX_i + e_{M_i}$$

$$Y_i = i_Y + c'X_i + bM_i + e_{Y_i}$$

What are the values for:

a = ____-2____ total = ____1.5____

b = ____-.5____ direct = ____.5____

c = ____1.5____ indirect = _(-2)×(-.5) = 1____

c' = ____.5____

Direct effect of X on Y (not through M) = *c*

Indirect effect of X on Y (through M) = *a* × *b*

Total effect = direct effect + indirect effect

c = *c'* + *a* × *b*

Indirect effect = total effect - direct effect

a × *b* = *c* - *c'*

Estimation in R

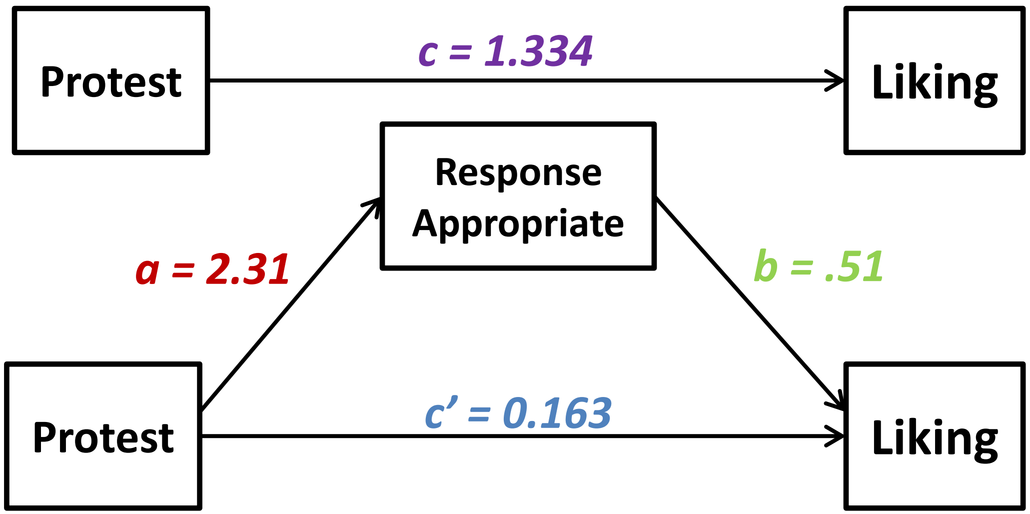
```
vaughnlm.cpath <- lm(relativepromotion ~ nscond, data =
vaughndata)
coef(vaughnlm.cpath)
cpath <- coef(vaughnlm.cpath)[2]

vaughnlm.apath <- lm(autonomy ~ nscond, data = vaughndata)
coef(vaughnlm.apath)
apath <- coef(vaughnlm.apath)[2]

vaughnlm.bpath <- lm(relativepromotion ~ nscond+autonomy,
data = vaughndata)
coef(vaughnlm.bpath)
cprimepath <- coef(vaughnlm.bpath)[2]
bpath <- coef(vaughnlm.bpath)[3]

ind.effect <- apath*bpath
```

Estimation with Vaughn Data

Indirect Effect

$a \times b = 2.31 \times 0.51 = 1.172$

Need for support increased relative promotion by .677 units indirectly through feelings of autonomy. Where need for support increased autonomy by 1.33 units, and a one unit increase in autonomy resulted in a .51 unit increase in relative promotion.



Direct Effect

$c' = 0.163$

Need for support increased relative promotion by .163 units directly (not through feelings of autonomy).

Total Effect

$c = 1.334$

Need for support increases liking by 1.334 units in total.

Inference for the direct and total effects can be drawn from the regression results because these are based on a single regression parameter.

$p = .202$

$p < .001$

Inference about the Indirect Effect

- How to make proper inference about the indirect effect may be the most active area of research in mediation analysis
- Some methods you may have heard of
  - Causal Steps / Baron and Kenny Method / Baron and Kenny Steps
  - Test of Joint Significance
  - Sobel Test / Multivariate Delta Method
  - Monte Carlo Confidence Intervals
  - Distribution of the Product Method
  - Bootstrap Confidence Intervals
    - Percentile Bootstrap
    - Bias-Corrected Bootstrap
    - Bias Corrected and Accelerated Bootstrap

  I’m going to talk about a few, but not all.

- Why is this so hard?

  We’re talking about a combination of two paths. There are many instances where the indirect effect could be zero (either *a* or *b* could be zero, or both could be zero).

Causal Steps Method

Method

1. Test if there is a significant total effect ($c \neq 0$).
2. Test if there is a significant effect of X on M ($a \neq 0$).
3. Test if there is a significant effect of M on Y controlling for X ($b \neq 0$).
4. If all three steps are confirmed, test for partial vs. complete mediation.
    1. If X still has an effect on Y controlling for M ($c' \neq 0$), this is partial mediation
    2. If X does not have a significant effect on Y controlling for M, complete mediation

Appeal

- Easy to do, just need regression
- Intuitive

What's wrong with it?

- No estimate of the indirect effect
- No quantification of uncertainty about conclusion
    - *p*-value
    - Confidence Interval
- Requirement that the total effect is significant before looking for indirect effect
- Multiple testing problem
- Issues with *complete* and *partial* mediation

Criticisms of Causal Steps

No estimation of the indirect effect.

Rationale

If each of the paths (*a* and *b*) are different from zero, then the indirect effect is different from zero, so we don't need an estimate.

Why isn't this a reasonable requirement?

- Individual tests are fallible. It may be that both paths are *actually* different from zero, but the test on the individual paths does not detect it (e.g. a small but non zero *a*-path, and a very large *b*-path means large indirect effect)
- Without a quantification of the indirect effect we cannot compare indirect effects to each other (multiple mediator models), or model how the size of the indirect effect varies (conditional process analysis).

No estimation of uncertainty

Rationale

Most of the time we'll make the right decision.

Why isn't this a reasonable requirement?

- Without confidence intervals it is difficult to discern if two studies potentially agree or disagree.

Criticisms of Causal Steps

Requirement that the total effect is significant before looking for indirect effect

Original Rationale

There must be an overall effect of *X* on *Y* in order for there to be an indirect effect of *X* on *Y* through *M*.

Why isn't this a reasonable requirement?

- Supression: Direct and indirect effects could be large but of opposite sign. $c' + ab = c$; so c will be close to zero
- Power: When $c' = 0$, and therefore $ab = c$ power to detect ab is higher than power to detect c. So why rely on a lower power test first?

Multiple Testing Problem

Each individual test has a probability of making an error. Using the causal steps logic we rely on 3 – 4 tests, each of which has its own error rate. The upper bound of the combined error rate is known but the true error rate is unknown. The error rate of a single test of the indirect effect is known.

Criticisms of Causal Steps

Complete vs. Partial Mediation

Original Rationale

If the *whole* relationship between *X* and *Y* is explained by *M* then *c'* should not be notably different than zero.

Why isn't this a reasonable requirement?

- The direct effect being zero (or close) does not indicate "complete" mediation: The direct effect is the sum of all effects not through *M*, there may be 2 or more mediators which explain the relationship between *X* and *Y* which cancel out to be zero.
- Sample size & Power: Complete mediation is a **strong claim** but researchers with smaller sample sizes and lower power will be less likely to detect *c'* and therefore more likely to claim complete mediation. **We are accepting the null hypothesis.**
- What if *c* is not significant? The concepts of complete and partial mediation mean very little in this context.
- All relationships are mediated: "partial mediation" just means there are other unmeasured mediators.

Joint Significance Test

Method

1. Test if there is a significant effect of X on M ($a \neq 0$).
2. Test if there is a significant effect of M on Y controlling for X ($b \neq 0$).

Appeal

- Easy to do, just need regression
- Intuitive

What's wrong with it?

- No estimate of the indirect effect
- No quantification of uncertainty about conclusion
    - *p*-value
    - Confidence Interval
- ~~Requirement that the total effect is significant before looking for indirect effect~~
- Multiple testing problem
- ~~Issues with *complete* and *partial* mediation~~

Sobel Test /Multivariate Delta Method

Assume the sampling distribution of the indirect effect (*ab*) is normally distributed, and derive the asymptotic variance of this normal distribution

$$Z = \frac{ab}{\sqrt{b^2 s_a^2 + a^2 s_b^2 + s_a^2 s_b^2}}$$

Under the null hypothesis that *ab* = 0, *Z* is normally distributed with mean 0 and variance 1. Can use traditional *Z* tables to compute a *p*-value associated with the observed value of *Z*.

Appeal

- Fairly easy to understand
- Provides point estimate of indirect effect
- Can calculate confidence intervals

What's wrong with it?

- The assumption that *ab* is normally distributed is incorrect
    - The product of two normal distributions is not necessarily a normal distribution
- Sobel test is conservative (too low type I error and too low power)

Writing your Own Functions

In R you can write your own functions so you don't have to repeat code over and over again.

```
add.custom <- function(x,y) x + y

add.custom(5,12)
```

For multiple step functions, include all the steps inside curly brackets

```
add.custom2 <- function(x,y) {
step1 <- x + y
step2 <- step1 + 2}
```

Setting defaults

```
add.custom3 <- function(x,y = 4) x + y
```

These are really basic examples, but writing your own functions really opens up the opportunities for you!

Strategies: Writing your Own Functions

- Start with simple test values
- Test with values where you know the answer
- Check each step as you go
- Do not try too much at once
- Print print print!
- Be careful about naming (scoping in R)

Practice

1. Build a function which given the independent, mediator, and outcome variable calculates the indirect effect.
2. Building on your previous function add a part of the output which gives a TRUE/FALSE value for the Causal Steps Test and one for the Joint Significance test.
3. Building on your previous function calculate the standard error of the indirect effect, the Z-value, and p-value for the Sobel test. Check out the norm function (dnorm, pnorm, qnorm, rnorm)

$$Z = \frac{ab}{\sqrt{b^2 s_a^2 + a^2 s_b^2 + s_a^2 s_b^2}}$$

Don’t try to do these all at once. BUILD!

Bootstrap Confidence Intervals
(Percentile)

Empirically estimate sampling distribution of the indirect effect. From this distribution compute confidence intervals which can be used for estimation and hypothesis testing.

Bootstrapping the Indirect Effect

1. Randomly sample *n* cases from your dataset with replacement.
2. Estimate the indirect effect using resampled dataset, call this $ab^{(1)}$
3. Repeat steps 1 and 2 a total of *K* times where *K* is many (10,000 recommended), each time calculated $ab^{(k)}$.
4. The sampling distribution of the $ab^{(i)}$'s can be used as an estimate of the sampling distribution of the indirect effect.
5. For a 95% confidence interval the lower and upper bounds will be the 2.5th and 97.5th percentiles of the *K* estimates of the indirect effect.

Appeal

- No assumptions about the sampling distribution of the indirect effect
- Provides point estimate of indirect effect
- Can calculate confidence intervals
- Most simulation work suggests this is a very good method which balances Type I Error and Power

What's wrong with it?

- Most software does not have this functionality built in
- Requires original data

Bootstrap Confidence Intervals

Original Data

X	M	Y
-0.35	-0.58	0.25
0.31	-0.50	1.89
-0.19	2.61	2.08
-1.30	-1.49	-0.54
0.59	1.14	1.74
-0.29	-0.29	1.04
1.80	0.08	1.23
-0.01	1.20	1.30
0.30	1.35	1.31
-0.98	0.90	-0.76

$a = .2931$ $b = .3099$

$ab = .0908$

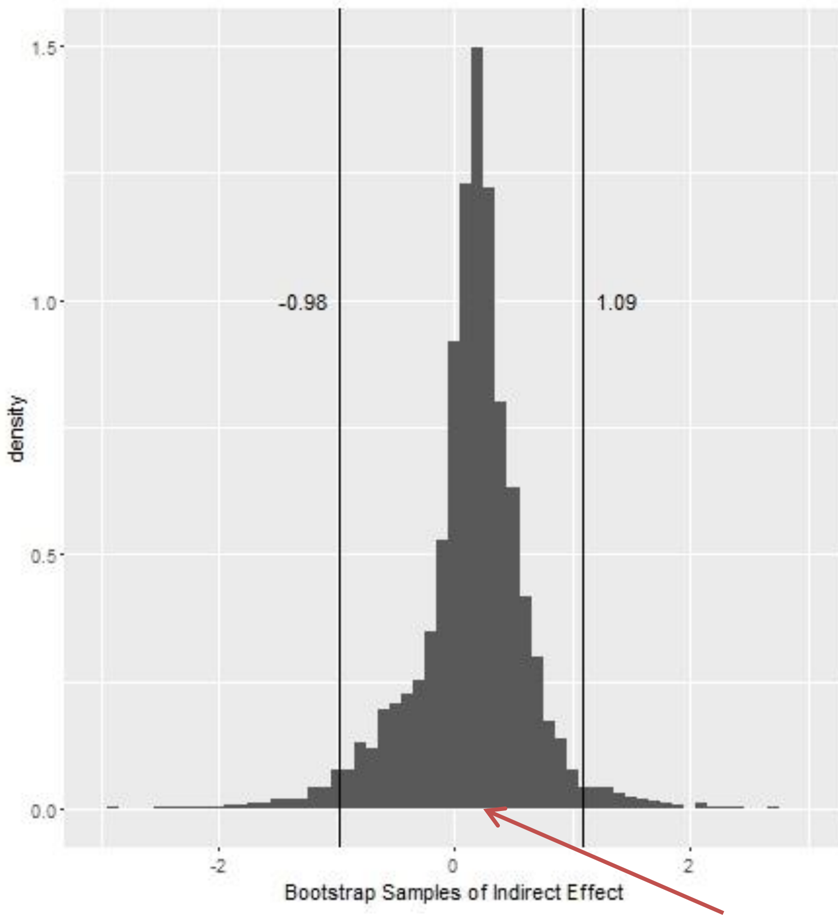
Bootstrap Sample

X	M	Y
-0.35	-0.58	0.25
-0.19	2.61	2.08
0.30	1.35	1.31
0.59	1.14	1.74
0.31	-0.50	1.89
-0.01	1.20	1.30
0.30	1.35	1.31
0.31	-0.50	1.89
0.30	1.35	1.31
-0.01	1.20	1.30

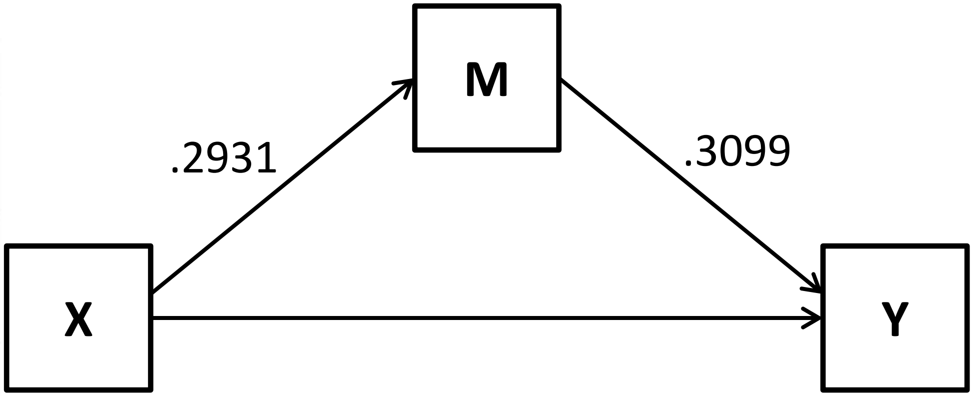
$a = -.1035$ $b = .1495$

$ab = -.0155$

Bootstrap Confidence Intervals



ab = .0908



ab = .0908

Zero is contained in the confidence interval so we **cannot** claim an indirect effect different from zero with 95% confidence. This is the same as failing to reject the null hypothesis at $\alpha = .05$.

Other Kinds of Bootstrap Confidence Intervals

All bootstrap confidence intervals use the same basic sampling technique, just use different methods for choosing the end points of the confidence intervals

Bias-Corrected Confidence Interval

- Percentile bootstrapping assumes that your sample estimate (*ab*) is unbiased in estimating the population indirect effect
- Bias-corrected reduces this assumption to assuming that the bias of *ab* is a constant (i.e. as N goes to infinity *ab* will go to the population indirect effect plus some constant)
- Bias-corrected confidence intervals estimate the bias of *ab* then adjust edges of confidence interval to be "bias-corrected" (i.e. centered not around your original estimate of *ab*), but around the point based on the bias estimation.

Bias-Corrected and Accelerated

- Same principles as BC regarding bias correction
- Acceleration allows for the assumption that the standard error of the indirect effect depends on the population value of the indirect effect
- Acceleration parameter, which is used to adjust the ends of the confidence interval is estimated using leave-one-out estimates of skew of the estimates of the indirect effect.

Looping Functions

In R you can loop through a function or set of functions multiple times.

```
i <- 1

for(j in 1:100){
  i <- i+1
  print(i)
}
```

Indexing variable (e.g. j) is key, can be used for a many purposes

```
i <- 1
results <- vector(length = 100)

for(j in 1:100){
  i <- i+1
  results[j] <- i
}
```

Again, this is really basic, but it can be used in a sophisticated way.

Sample Function

Sample function can be used to select a specific **size** sample from a vector **x** either with or without **replacement**.

```
sample(x = 1:100, size = 20, replace = TRUE)
```

The sample function can be used to randomly select rows of a dataframe

```
randsamp <- sample(x = 1:nrow(vaughndata),
       size = nrow(vaughndata), replace = TRUE)
head(randsamp)
newvaughn <- vaughndata[randsamp, ]
head(newvaughn)
```

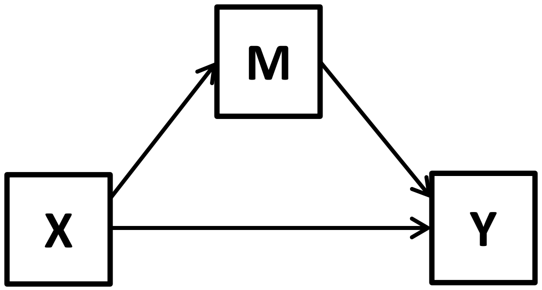
Practice: Bootstrap

1. Build a function takes a random sample with replacement from a dataset and calculates the indirect effect.
2. Use a loop function to calculate a large number of bootstrapped indirect effects (e.g. 1000)
3. Add to the function to calculate a 95% confidence interval using the percentiles of the bootstrapped indirect effects.
4. Can you combine this function with the last so you'll get the results of the Causal Steps, Join Significance, Sobel, and Bootstrap methods all together?

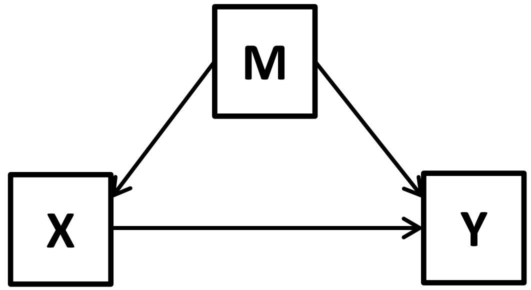
Don't try to do these all at once. BUILD!

A Brief Caution on Causality

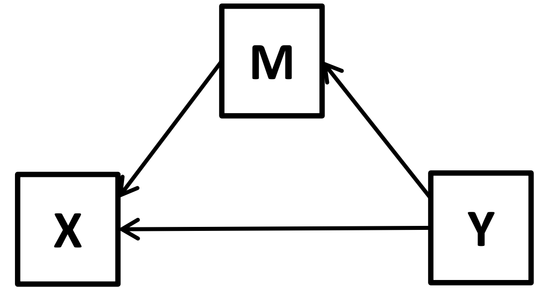
There are a number of alternative causal processes that may be occurring when a *statistical indirect effect* is present



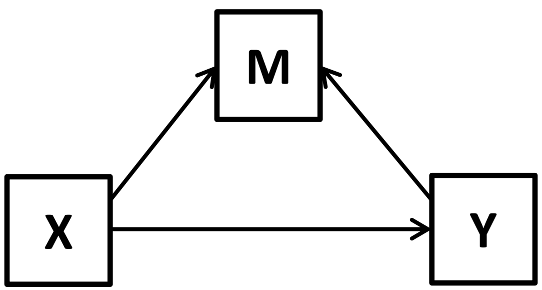
X affects *Y* through *M*



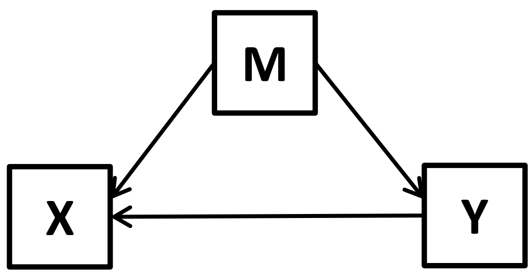
M affects *Y* through *X*



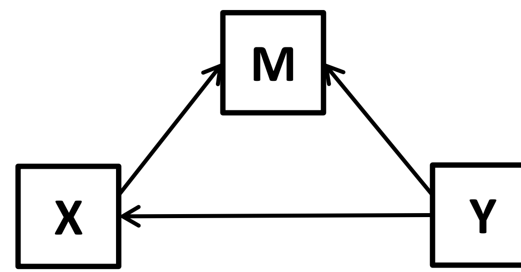
Y affects *X* through *M*



X affects *M* through *Y*



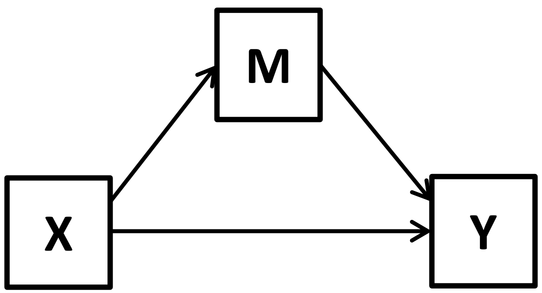
M affects *X* through *Y*



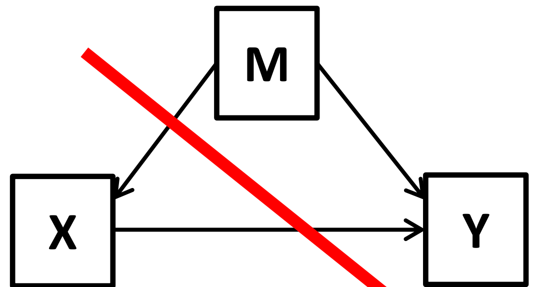
Y affects *M* through *X*

A Brief Caution on Causality

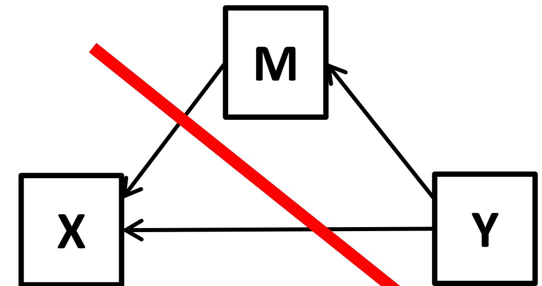
What you get by manipulating *X*.



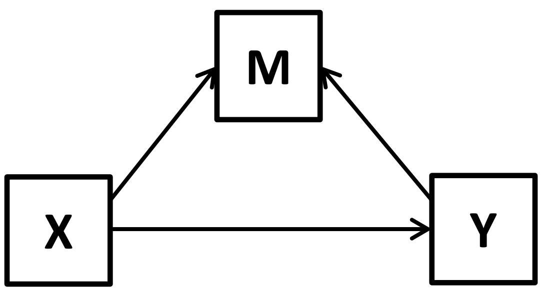
X affects *Y* through *M*



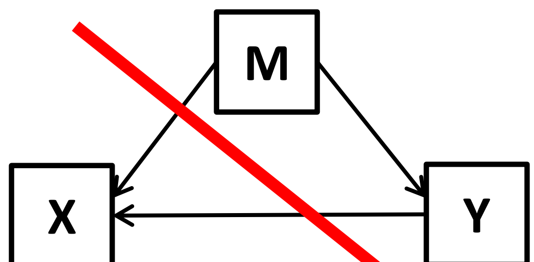
M affects *Y* through *X*



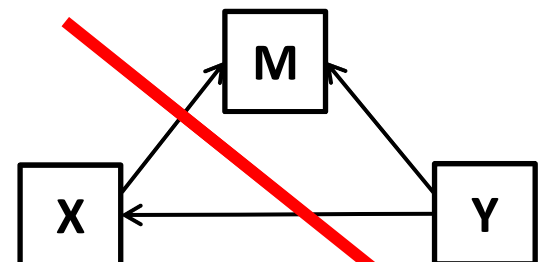
Y affects *X* through *M*



X affects *M* through *Y*



M affects *X* through *Y*



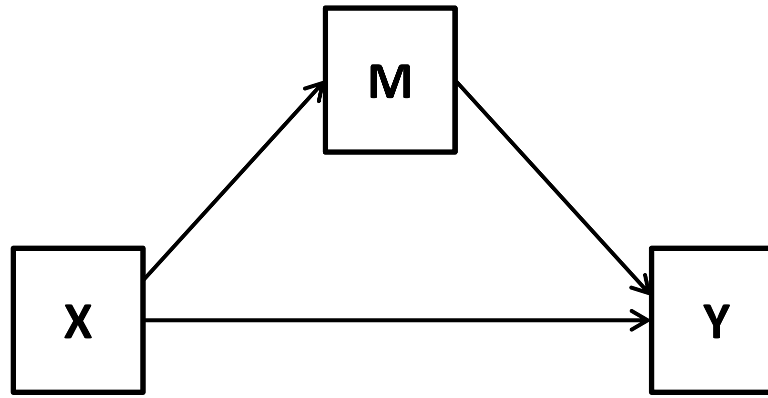
Y affects *M* through *X*

Even when *X* is manipulated, we can not provide evidence for the causal order between *M* and *Y*. This can only be supported using other experiments or previous research. *A statistically significant indirect effect does not lend credence to one model over another.*

Questions on Mediation?

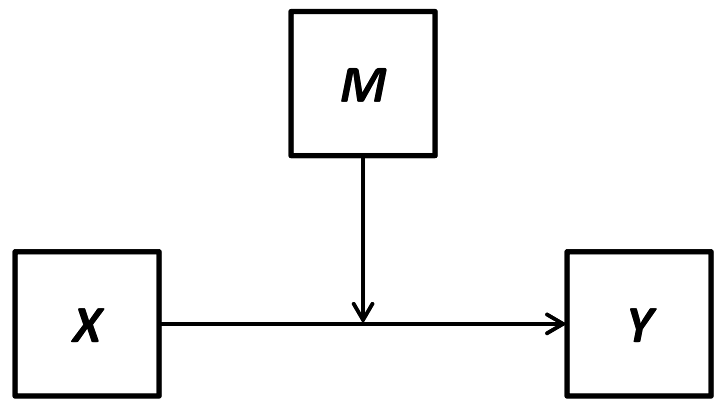
Conditional Process Analysis

Conditional process analysis combines the ideas of **mediation** and **moderation**. All conditional process models have a mediation model where one or more paths in this model is moderated.



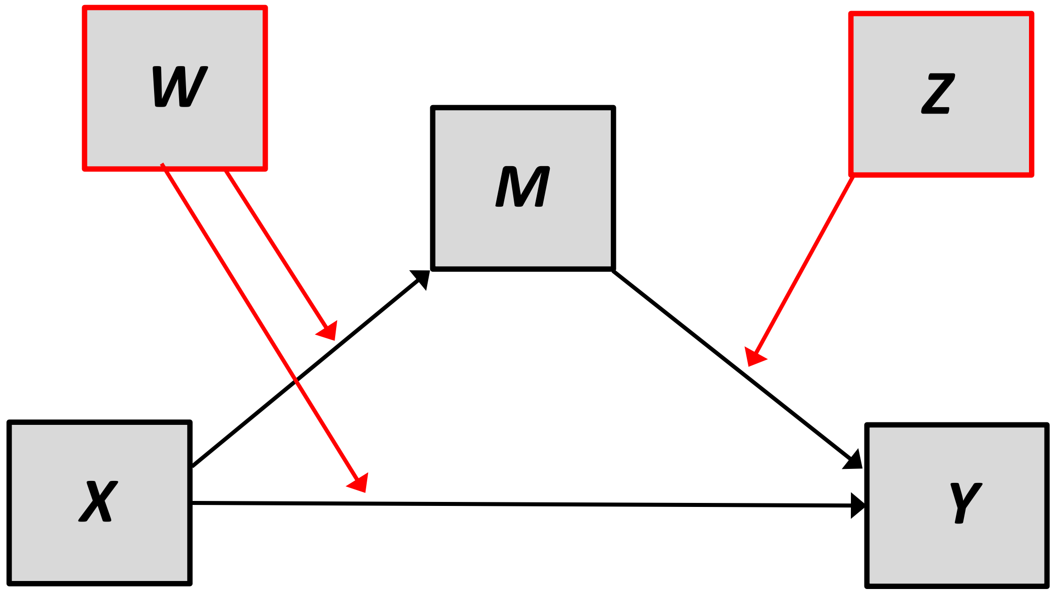
Conditional process analysis is useful for identifying the "when" or "for whom" certain mechanisms operate between *X* and *Y*.

When one or more paths of the indirect effect is moderated, the indirect effect can be defined as a *function of the moderator*, allowing for tests of moderation and probing.



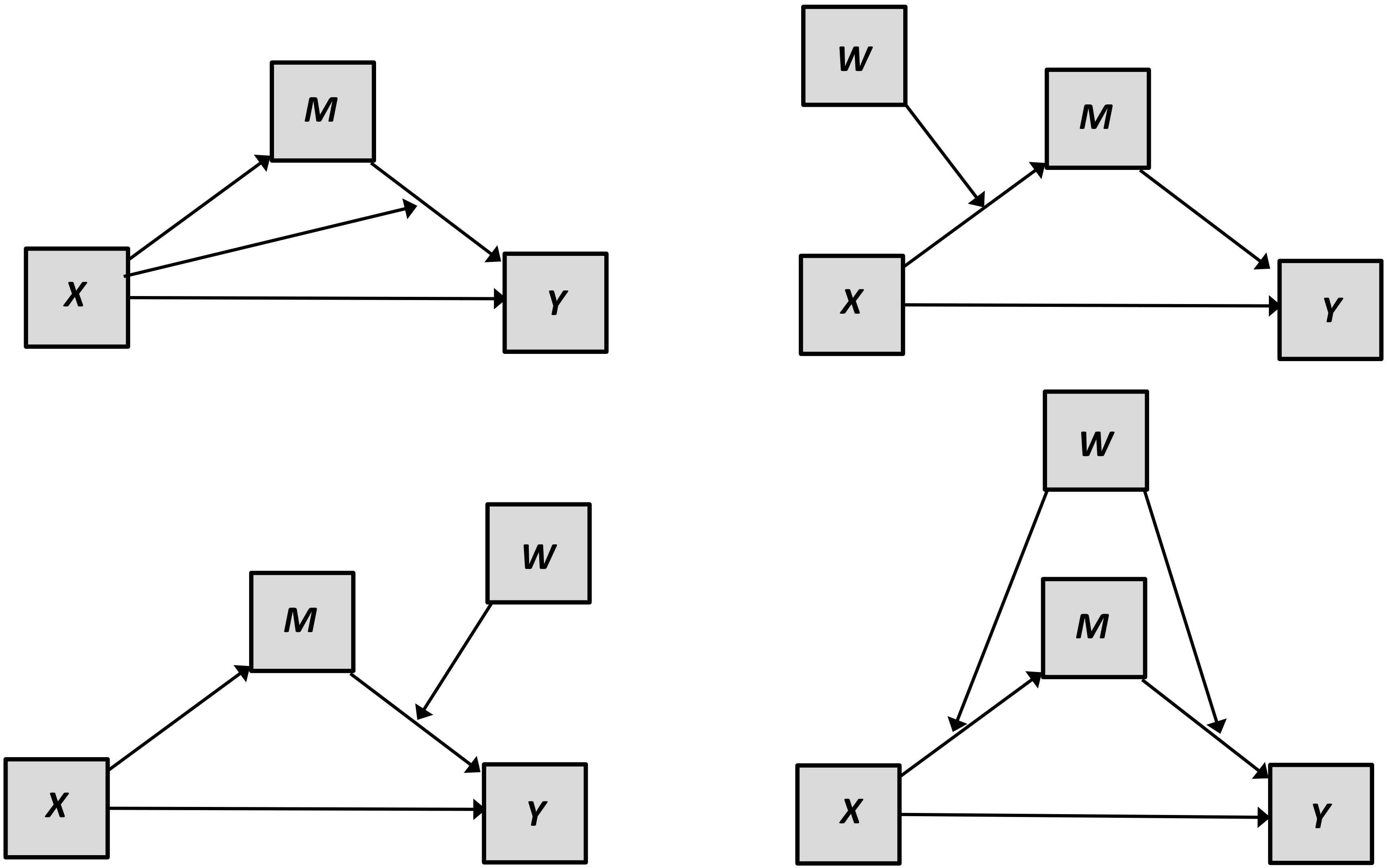
Conditional process analysis is not a new concept, really just a new name. You may know it by *moderated mediation* or *mediated moderation*.

Conditional Process Modeling

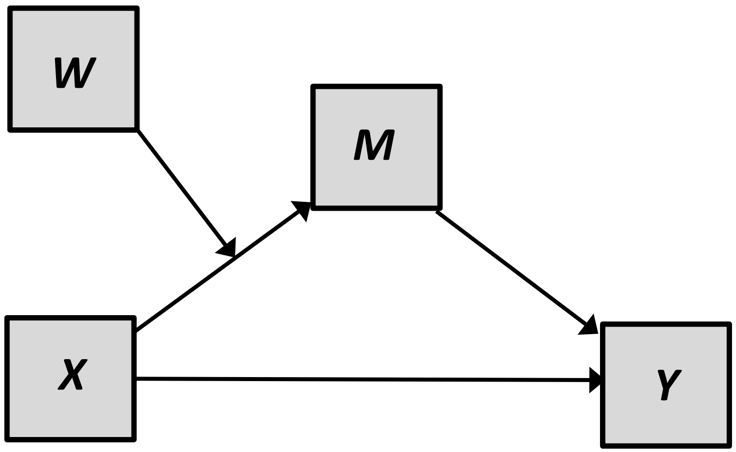


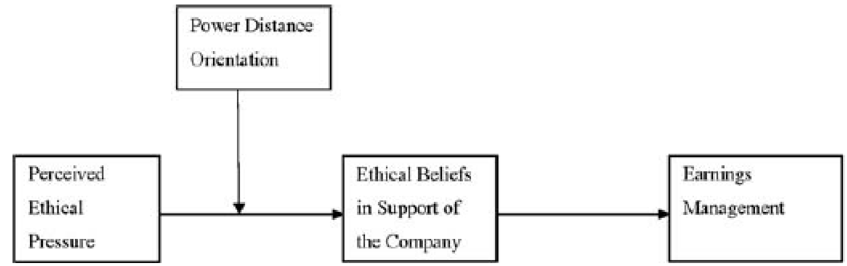
- The indirect effect of *X* on *Y* through *M* is estimated as the product of two paths
- But what if size of *the X* → *M* path or *M* → *Y path* (or both) depends on another variable (i.e., is moderated)?
- If so, then the magnitude of the indirect effect therefore depends on a third variable— "moderated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

A Few of the Many Possibilities

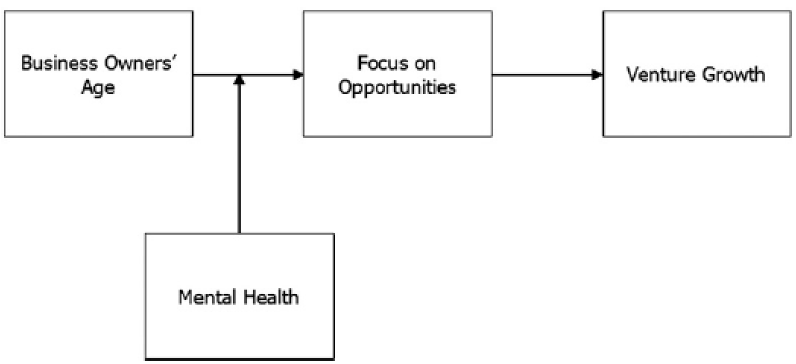


Examples: *X* to *M* Path Moderated by *W*

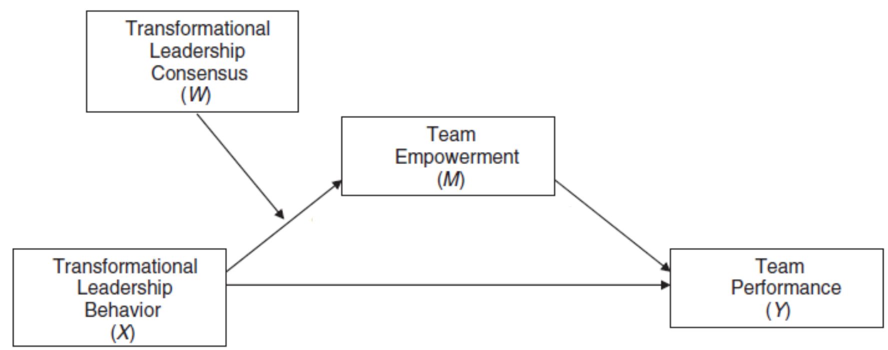




Tian, Q., & Peterson, D. K. (2016). The effects of ethical pressure and power distance orientation on unethical pro-organizational behavior: the case of earnings management. ***Business Ethics: A European Review,*** *25*(2), 159-171.

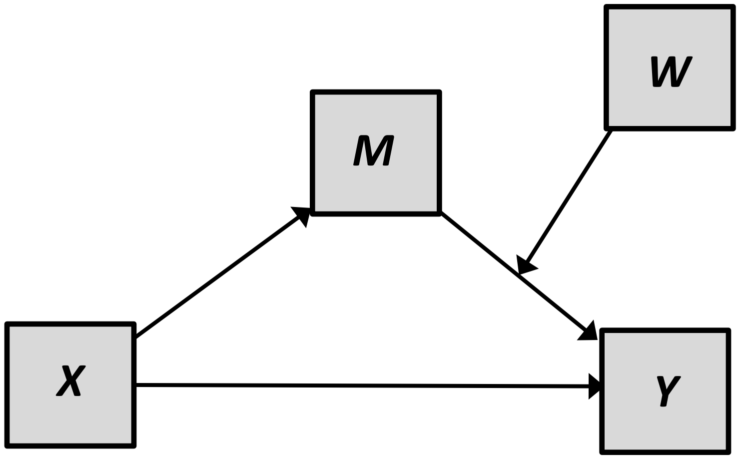


Gielnik, M. M., Zacher, H., & Frese, M. (2012). Focus on opportunities as a mediator of the relationship between business owners' age and venture growth. ***Journal of Business Venturing***, *27*, 127-142.

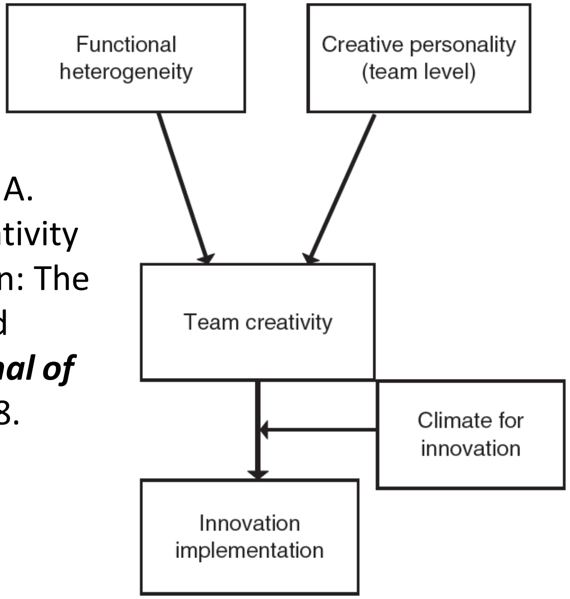


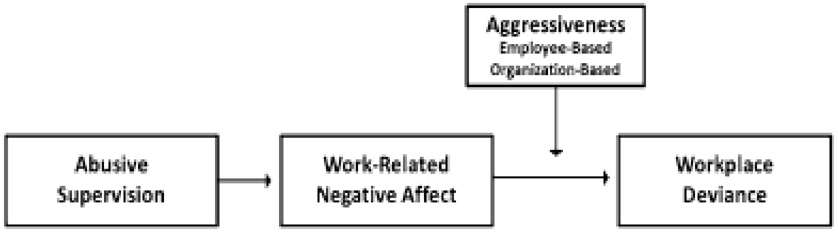
Cole, M. S., Bedeian, A. G., & Bruch, H. (2011). Linking leader behavior and leadership consensus to team performance: Integrating direct consensus and dispersion models to group composition. ***The Leadership Quarterly***, *22*, 383-398.

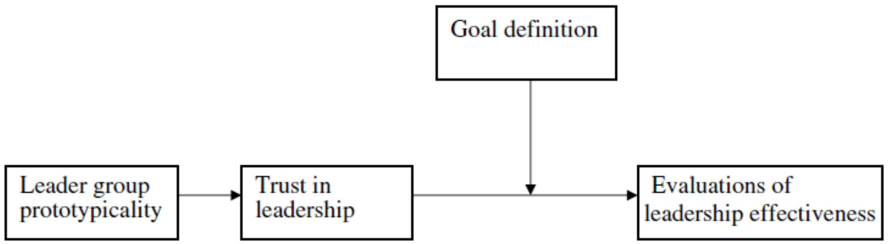
Examples: *M* to *Y* Path Moderated by *W*



Somech, A., & Drach-Zahavy, A. (2013). Translating team creativity to innovation implementation: The role of team composition and climate for innovation. ***Journal of Management***, 39(3), 684-708.



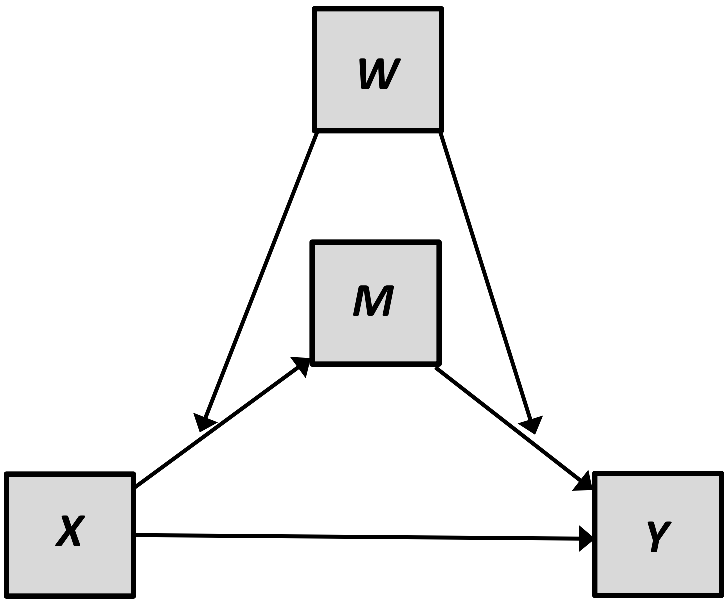


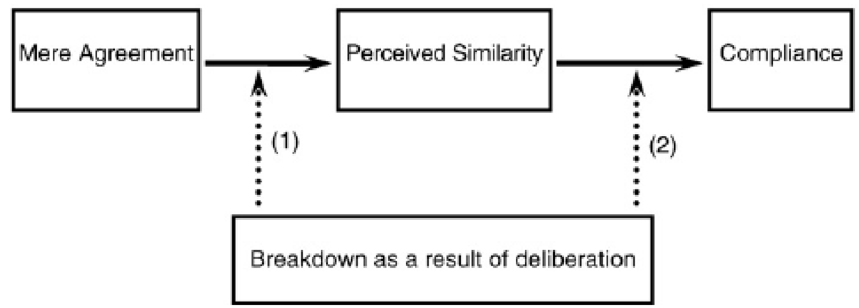


Michel, J. S., Newness, K., & Duniewicz, K. (2015). How abusive supervision affects workplace deviance: A moderated-mediation examination of aggresiveness and work-related negative affect. ***Journal of Business Psychology***, 31(1), 1-22.

Giessner, S., van Knippenberg, D., & Sleebos, E. (2008). "License to fail": Goal definition, leader group prototypicality, and perceptions of leadership effectiveness after leader failure. ***Organizational Behavior and Human Decision Processes***, *105*, 14-35

Examples: *X* to *M* and *M* to *Y* Path Moderated by *W*

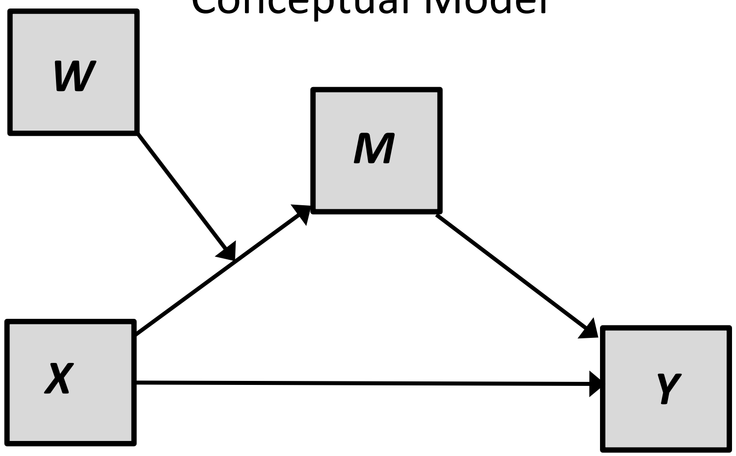




Pandelaere, M., Briers, B., Dewitte, S., & Warlop, L. (2010). Better think before agreeing twice. Mere agreement: A similarity-based persuasion mechanism. ***International Journal of Research in Marketing***, *27,* 133-141.

X to *M* Path Moderated by *W*

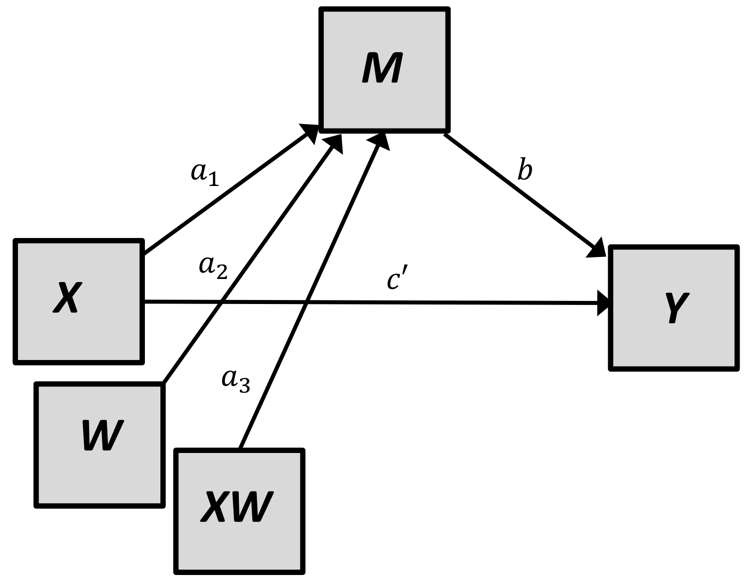
Conceptual Model



X is proposed to influence Y indirectly through M, but the effect of X on M is moderated by W. Therefore the degree to which M operates as a mediator of the effect of X on Y depends on W.

In order to describe this relationship it will be worthwhile to construct an estimate of the indirect effect at different values of W.

Statistical Model



$$M_i = i_M + a_1 X_i + a_2 W_i + a_3 X_i W_i + \epsilon_{Mi}$$

equivalently

$$M_i = i_M + (a_1 + a_3 W_i) X_i + a_2 W_i + \epsilon_{Mi}$$

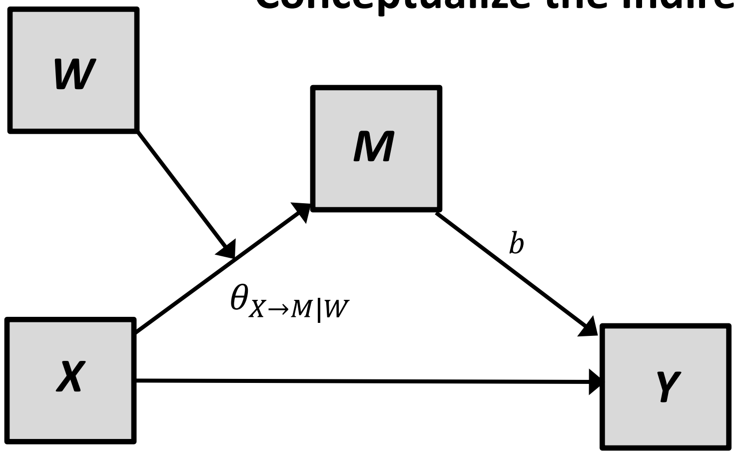
$$M_i = i_M + \theta_{X \to M|W} X_i + a_2 W_i + \epsilon_{Mi}$$

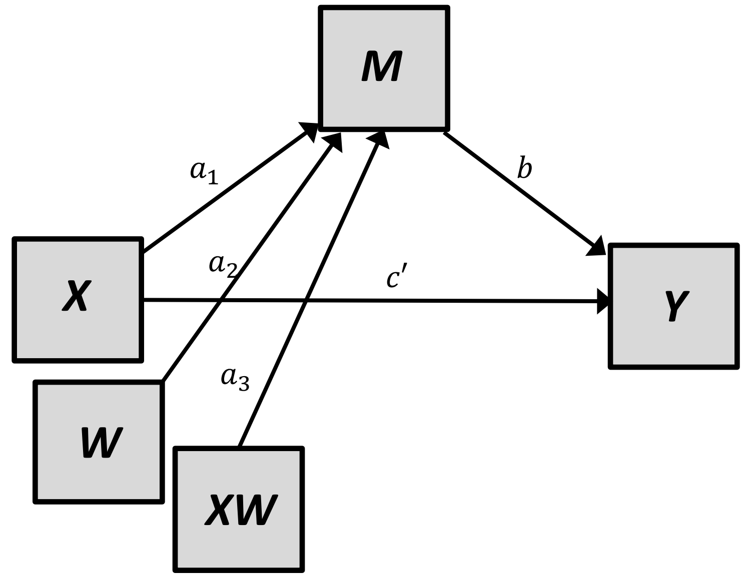
Where $\theta_{X \to M|W} = a_1 + a_3 W_i$

$$Y_i = i_Y + c' X_i + b M_i + \epsilon_{Yi}$$

X to *M* Path Moderated by *W*

Conceptualize the Indirect Effect as the Product of Two Paths





$\theta_{X\to M|W}$ is a quantification of the effect of *X* on *M* conditional on *W*. The estimate of the effect of *M* on *Y* is *b*.

$$M_i = i_M + \theta_{X\to M|W} X_i + a_2 W_i + \epsilon_{Mi}$$

Where $\theta_{X\to M|W} = a_1 + a_3 W_i$

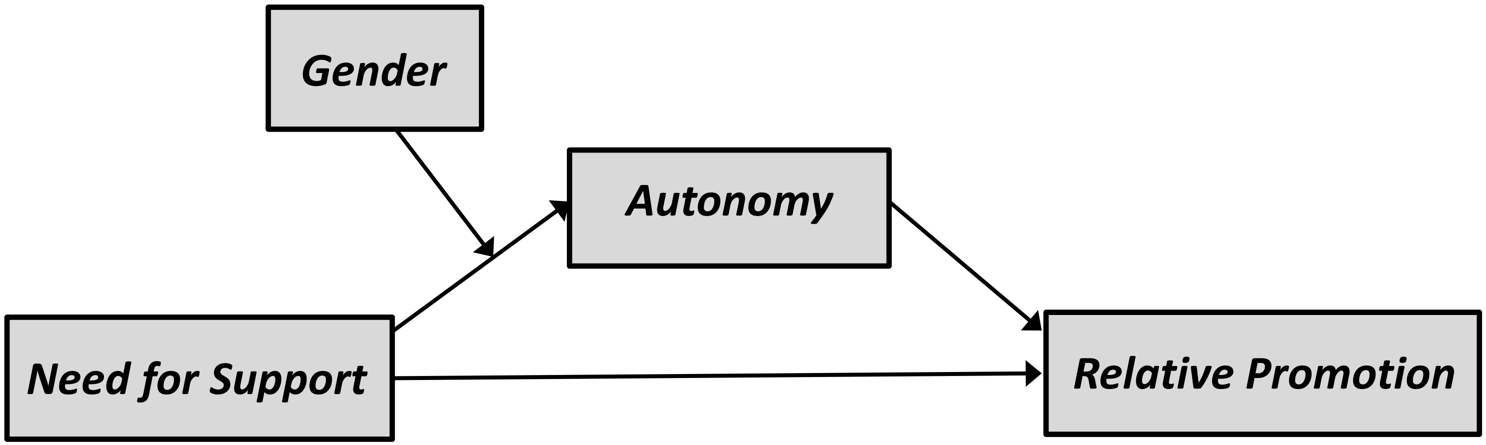
An estimate of the conditional indirect effect of *X* on *Y* through *M* conditional on *W*

$$\theta_{X\to M|W} b = (a_1 + a_3 W) b$$

The indirect effect is now a function of W

Now we have an estimate of the indirect effect at any value of *W.*

Conceptual Model with Protest Data



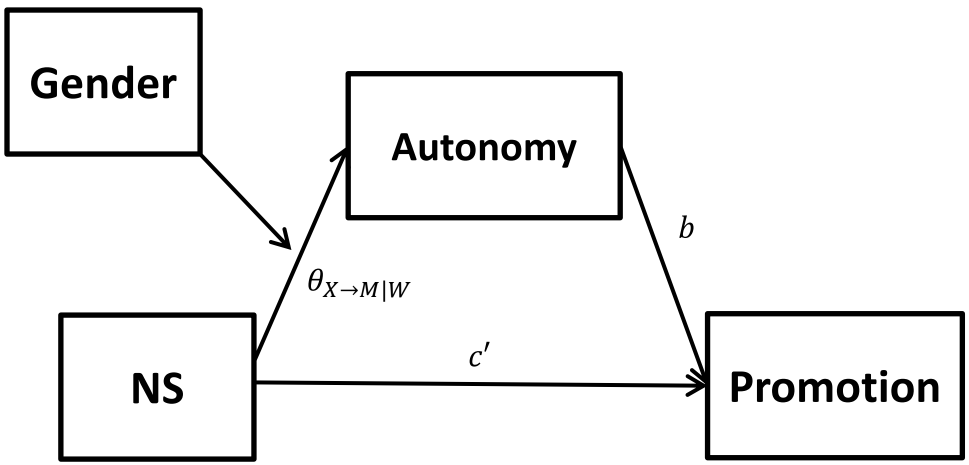
We have set up the mediation model that we examined before: Remember we found that the relationship between need for support and relative promotion depended on gender. So could it be that gender moderates the path from need for support to autonomy?

Now we can ask **Does the degree to which autonomy acts as a mediator between need for support and relative promotion depend on gender?**

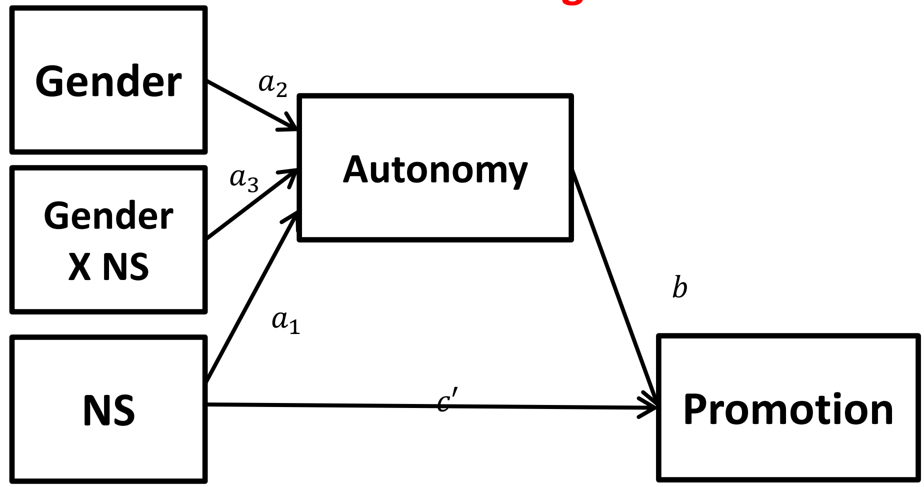
Note: To avoid confusion of "*M*" vs. "*M*" we're going to call gender "*W*" now.

Conceptual and Statistical Diagram

Conceptual Diagram



Statistical Diagram

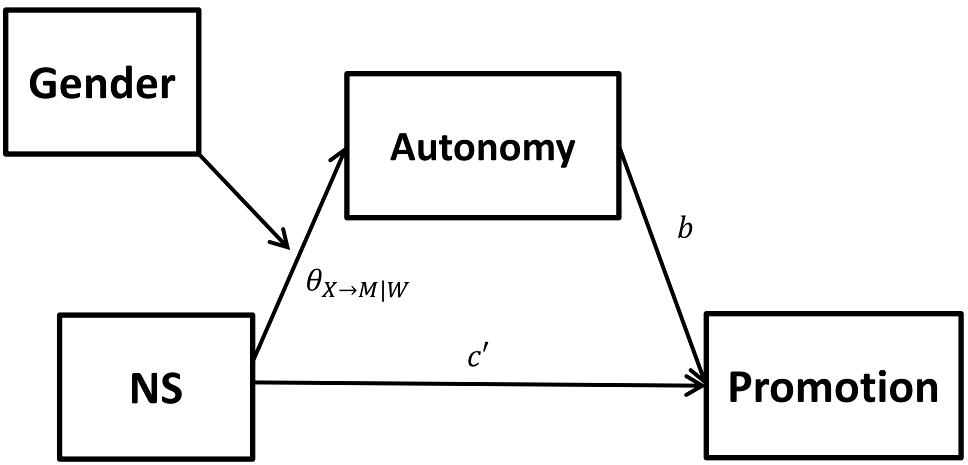


$$M_i = i_M + a_1 X_i + a_2 W_i + a_3 X_i W_i + \epsilon_{Mi}$$

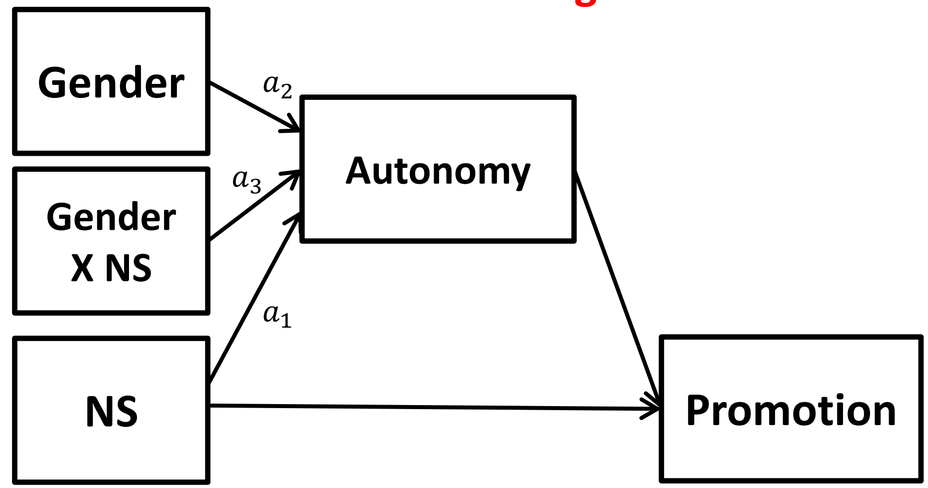
$$Y_i = i_Y + c' X_i + b M_i + \epsilon_{Yi}$$

The Model of *M*

Conceptual Diagram



Statistical Diagram



The effect of protesting on response appropriateness can be modeled as a function of perceived prevalence of sexism

$$\boxed{M_i = i_M + a_1 X_i + a_2 W_i + a_3 X_i W_i + \epsilon_{Mi}}$$

$$M_i = i_M + (a_1 + a_3 W_i) X_i + a_2 W_i + \epsilon_{Mi}$$

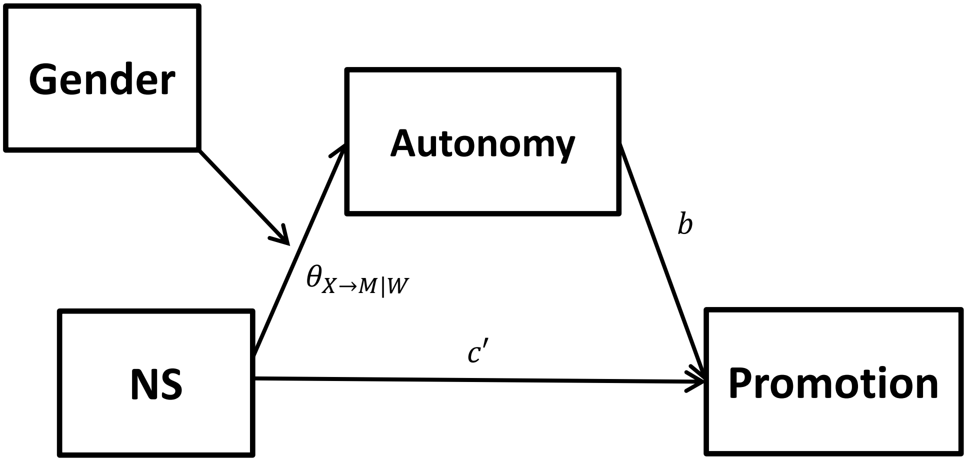
$$\theta_{X \to M|W} = a_1 + a_3 W_i$$

$$Y_i = i_Y + c' X_i + b M_i + \epsilon_{Yi}$$

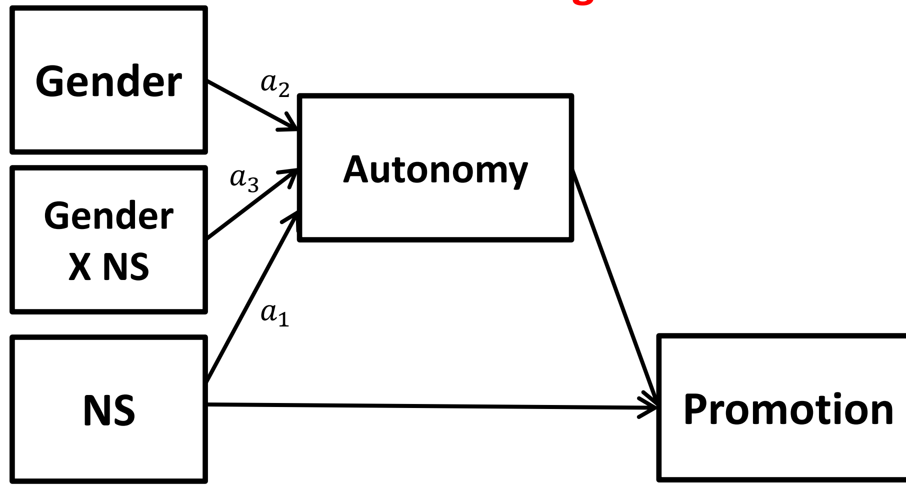
A quantification of the effect of *X* on *M*: $\theta_{X \to M|W} = (a_1 + a_3 W)$

The Model of *Y*

Conceptual Diagram



Statistical Diagram



$$M_i = i_M + a_1 X_i + a_2 W_i + a_3 X_i W_i + \epsilon_{Mi}$$

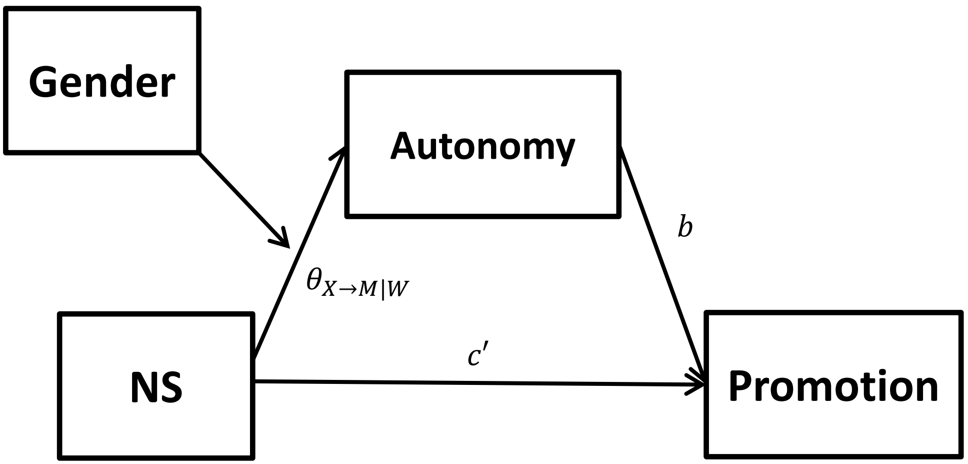
The effect of response appropriateness on liking is constant. This is a modeling choice!

$$\boxed{Y_i = i_Y + c' X_i + b M_i + \epsilon_{Yi}}$$

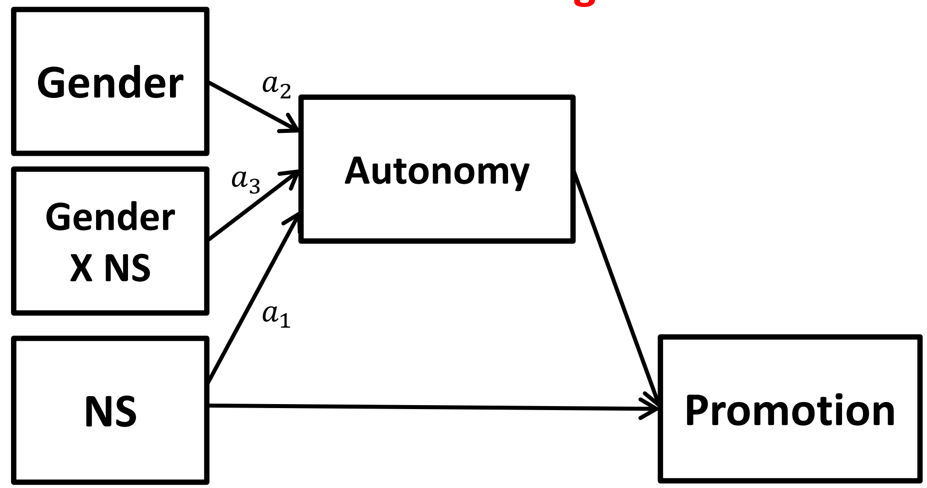
A quantification of the effect of *M* on *Y*: b

The Conditional Indirect Effect

Conceptual Diagram



Statistical Diagram



$$M_i = i_M + a_1 X_i + a_2 W_i + a_3 X_i W_i + \epsilon_{Mi}$$

$$Y_i = i_Y + c' X_i + b M_i + \epsilon_{Yi}$$

The indirect effect of *X* on *Y* through *M* is **not** $a_1 b$. It is $\theta_{X\rightarrow M|W} b$ where $\theta_{X\rightarrow M|W} = a_1 + a_3 W$, so the indirect effect of *X* is $(a_1 + a_3 W)b$. It is conditional because it *depends on W*.

A quantification of the conditional indirect effect: $\theta_{X\rightarrow Y|W} b = (a_1 + a_3 W) b$

Estimation in R

```
vaughnlm.cpath <- lm(relativepromotion ~ nscond, data =
vaughndata)
coef(vaughnlm.cpath)
cpath <- coef(vaughnlm.cpath)[2]

vaughnlm.apath <- lm(autonomy ~ nscond*gender, data =
vaughndata)
coef(vaughnlm.apath)
a1path <- coef(vaughnlm.apath)[2]
a2path <- coef(vaughnlm.apath)[3]
a3path <- coef(vaughnlm.apath)[4]

vaughnlm.bpath <- lm(relativepromotion ~ nscond+autonomy, data
= vaughndata)
coef(vaughnlm.bpath)
cprimepath <- coef(vaughnlm.bpath)[2]
bpath <- coef(vaughnlm.bpath)[3]

cond.ind.W1 <- (a1path + a3path*1)bpath #Indirect for Males
cond.ind.W2 <- (a1path + a3path*2)bpath #Indirect for Females
```

A Formal Test of Moderated Mediation

Are the two indirect effects (one for men and one for women) different from each other?

What is the difference between the two?

$$\theta_{X \to M|W} b = a_1 b + \boxed{a_3 b} W$$

If $a_3 b$ is different from zero, this is evidence that the indirect effect depends on W. We call this term the **Index of Moderated Mediation**. A formal test on $a_3 b$ can provide a formal test for if the mediation is moderated.

We can do this formal test using bootstrapping.

IMM as Test of Difference in Indirect Effects

We may wonder if the indirect effect of *X* on *Y* through *M* at some value of the moderator *W* (call this w_1) is different from the indirect effect of *X* on *Y* through *M* at some **other** value of the moderator *W* (call this w_2)

$H_0: \theta_{X\to M|W=w_1} b = \theta_{X\to M|W=w_2}\, b$ or equivalently $H_0: \theta_{X\to M|W=w_1} b - \theta_{X\to M|W=w_2}\, b = 0$

$H_A: \theta_{X\to M|W=w_1} b - \theta_{X\to M|W=w_2}\, b \neq 0$

We can test this!

$$\theta_{X\to M|W=w_1} b = a_1 b + a_3 b w_1$$

$$\theta_{X\to M|W=w_2} b = a_1 b + a_3 b w_2$$

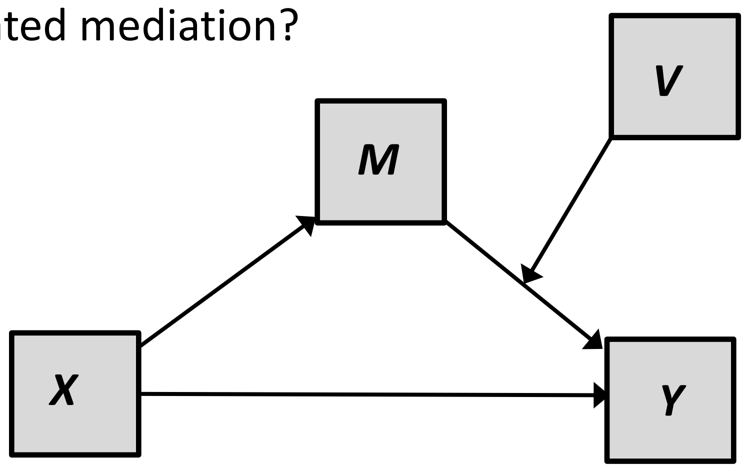
$$\begin{aligned}\theta_{X\to M|W=w_1} b - \theta_{X\to M|W=w_2}\, b &= (a_1 b + a_3 b w_1) - (a_1 b + a_3 b w_2)\\ &= a_3 b w_1 - a_3 b w_2\\ &= a_3 b (w_1 - w_2)\end{aligned}$$

So long as w_1 does not equal w_2, this means that the difference between the indirect effects is zero **only if a_3b is non zero**. Thus the *index of moderated mediation* is a test for if any two indirect effects differ from each other.

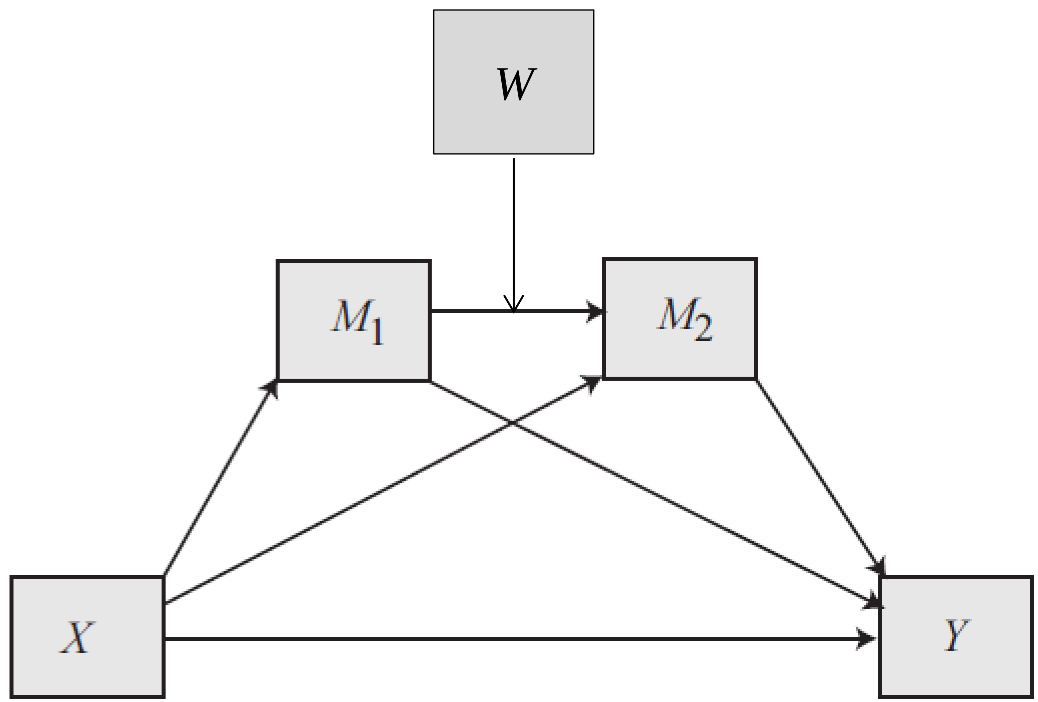
Practice: Index of Moderated Mediation

1. Build a function which estimates all the paths in a first stage moderated mediation model.
2. Build a function which calculates the conditional indirect effect, conditional on the coded values of W (1 and 2).
   a. Can you make this more general? What if W were coded as different values?
3. Add to your function the ability to calculate the index of moderated mediation.
4. Add to the function to calculate a 95% confidence interval using the percentiles of the bootstrapped index of moderated mediation.

5. Challenge: Can you make a new function which examines second stage moderated mediation?



The ultimate challenge: Moderated Serial Mediation



Resources

Books

Introduction to Mediation, Moderation, and Conditional Process Analysis by Andrew F. Hayes, Guilford Press

Interaction Effects in Multiple Regression by James Jaccard & Robert Turrisi, Sage

Introduction to Statistical Mediation Analysis, David MacKinnon, Taylor & Francis

Bootstrapping, Christopher Mooney & Robert Duval, Sage

A first course in Statistical Programming with R Braun and Murdoch

Ggplot2, Hadley Wickham

Monte Carlo simulation and resampling methods for social science, Carsey and Harden

Workshops

Mediation, Moderation, and Conditional Process Analysis, Andrew Hayes, June 6 – 7, Brisbane, Australia

Mediation, Moderation, and Conditional Process Analysis, Andrew Hayes, July 11 - 15, Chicago Illinois

Websites:

Afhayes.com

Quantpsy.org (Kris Preachers Webpage)

Akmontoya.com

Njrockwood.com

Reading List

Montoya A. K. & Hayes, A. F. (*in press*) Two condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods.* **Available at afhayes.com**

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.*

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of interential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, 24, 1918-1927.

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.

Spiller, S. A., Fitzsimons, G. J., Lynch Jr., J. G., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, *100*, 277-288.

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.

Preacher, K. J. & Selig, J. P. (2012) Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, *6*(2), 77-98

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.

Reading List

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 millennium. *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). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, and L. B. Snyder (Eds), *The Sage sourcebook of advanced data analysis methods for communication research (pp. 13-54). Thousand Oaks, CA: Sage* Publications.

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.*

Edwards, J. R. & Lambert, L. S. (2007) Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis, *Psychological Methods,* 12*(1), 1-22.*

Bauer, D. J., & Curran, P. J. (2005) Probing interactions in fixed and multilevel regression: Inferential and graphical techniques, *Multivariate Behavioral Researcher*, *40*(3), 373 - 400

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers, 36, 717-731.*

Thank you!

I am available for questions after the workshop and via email at montoya.29@osu.edu