# Intro to Causal Inference

SDS 390 Structural Equation Modeling

Monday Feb 25, 2019

# Emma K. T. Benn

## Increasing Inclusion, Promotion and Evidence:

Uniquely Merging my Intersectionality
With My Profession as a
Biostatistician

**Tuesday, March 5 • 5 p.m.**

Seelye Hall, Room 106

Sponsored by the Statistical and Data Sciences Program and The Smith College Lect

# Agenda for today

- MORE Path analysis
  - Mediation versus moderation in SEM
- Intro to Causal Inference
- Project brainstorming time
  - Data search

# Moderation and Mediation

Where do interactions live in SEM?

# Moderation (interactions)

- In SEM moderation effects can be represented in
  1. Having different paths (not including equality constraints)
  2. Or by adding product terms to the model (Note: you would need to first create the product in your dataset)

# Recall Lab 3 Example

```{r}
reg_mod <- lm(satisfaction ~ tension + other_pos + gender, data = acitelli)

summary(reg_mod)
```

```
Call:
lm(formula = satisfaction ~ tension + other_pos + gender, data = acitelli)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.72552 -0.19603  0.02919  0.26636  0.88862 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.25764    0.24629  13.227  < 2e-16 ***
tension     -0.35943    0.03458 -10.394  < 2e-16 ***
other_pos    0.28633    0.04730   6.053 4.36e-09 ***
gender      -0.02370    0.02207  -1.073    0.284    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3765 on 292 degrees of freedom
Multiple R-squared:  0.4307,	Adjusted R-squared:  0.4248 
F-statistic: 73.63 on 3 and 292 DF,  p-value: < 2.2e-16
```

```
acitelli_dyad <- acitelli %>%
  gather(self_pos:simhob,
         key = "the_old_vars",
         value = "score") %>%
  select(-gender) %>%
  unite(the_new_vars, the_old_vars, gender_string,
        sep = "_", remove = TRUE) %>%
  spread(the_new_vars, score)

head(acitelli_dyad)
```

| cuplid <dbl> | Yearsmar <dbl> | other_pos_man <dbl> | other_pos_woman <dbl> |
|---|---|---|---|
| 3 | 8.202667 | 4.0 | 4.6 |
| 10 | 10.452667 | 4.0 | 3.8 |
| 11 | -8.297333 | 4.8 | 4.4 |
| 17 | -6.380667 | 4.4 | 3.6 |
| 21 | 10.202667 | 4.8 | 3.8 |
| 22 | 15.036000 | 4.6 | 5.0 |

# Recall Lab 3 Example

```
```{r}
model_ex10 <- 'satisfaction_woman  ~ 1 + b1*other_pos_woman + b2*tension_woman
               satisfaction_man  ~ 1 + b1*other_pos_man + b2*tension_man
               satisfaction_woman ~~ v*satisfaction_woman
               satisfaction_man ~~ v*satisfaction_man'

fit_ex10 <- sem(model_ex10, data = acitelli_dyad)

summary(fit_ex10)
```
```

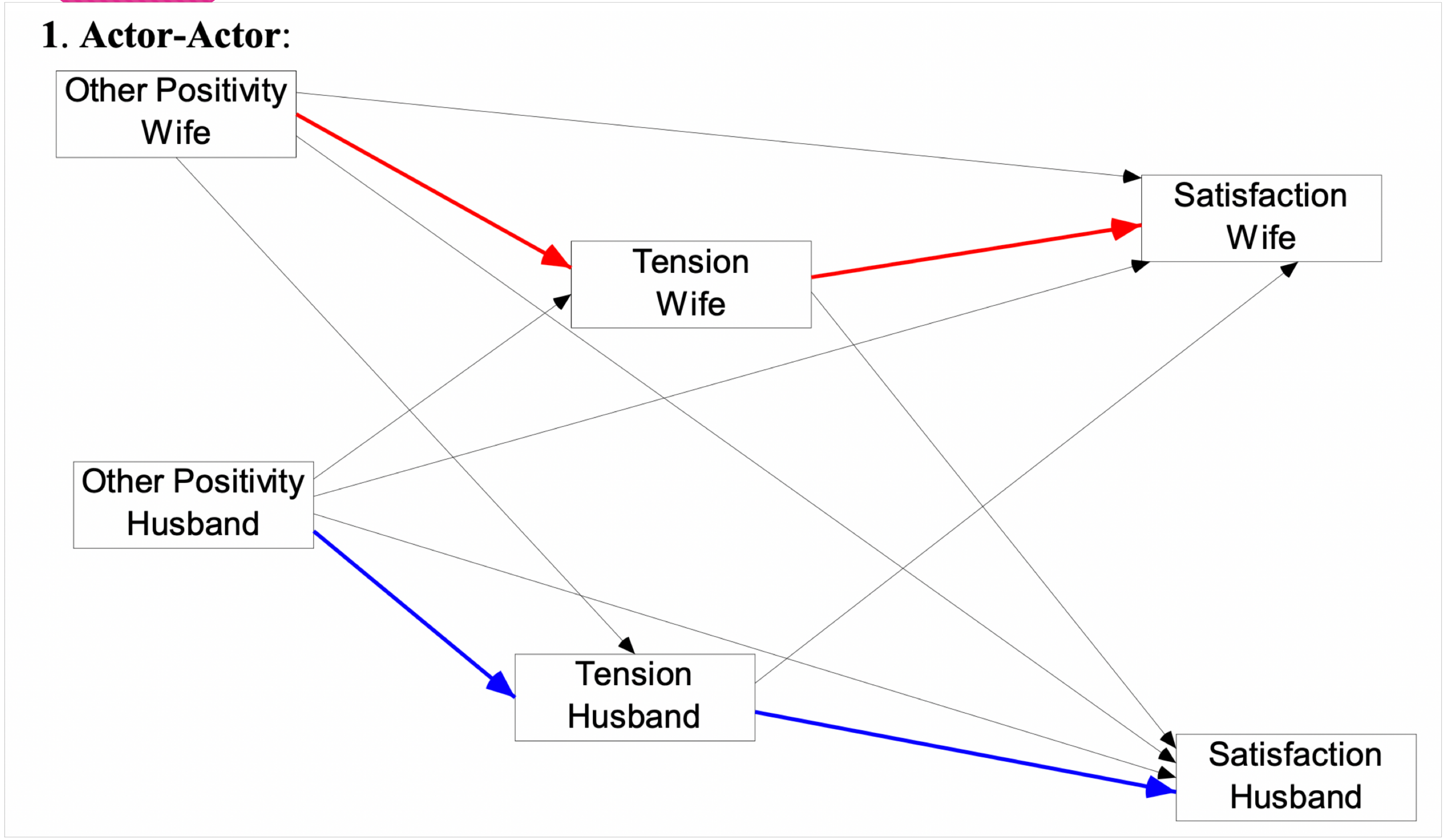
```
Regressions:
                         Estimate  Std.Err  z-value  P(>|z|)
  satisfaction_woman ~
    othr_ps_w (b1)          0.232    0.045    5.158    0.000
    tensn_wmn (b2)         -0.321    0.034   -9.565    0.000
  satisfaction_man ~
    othr_ps_m (b1)          0.232    0.045    5.158    0.000
    tensin_mn (b2)         -0.321    0.034   -9.565    0.000

Covariances:
                         Estimate  Std.Err  z-value  P(>|z|)
 .satisfaction_woman ~~
   .satisfactin_mn          0.051    0.012    4.098    0.000

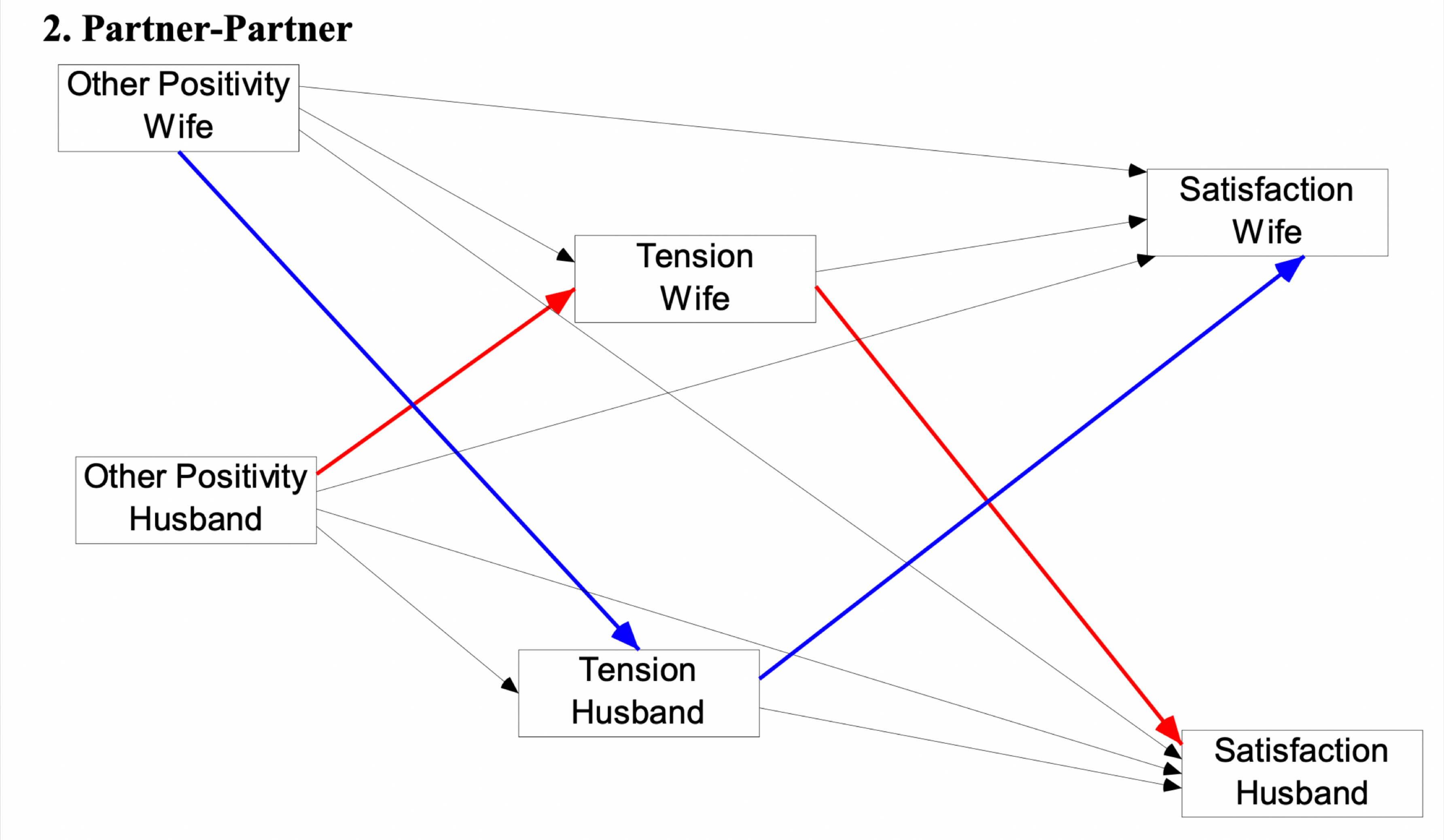
Intercepts:
                    Estimate  Std.Err  z-value  P(>|z|)
   .satisfactn_wmn     3.416    0.233   14.674    0.000
   .satisfactin_mn     3.377    0.231   14.642    0.000

Variances:
                    Estimate  Std.Err  z-value  P(>|z|)
   .stsfctn_wm (v)     0.142    0.012   11.454    0.000
   .stsfctn_mn (v)     0.142    0.012   11.454    0.000
```

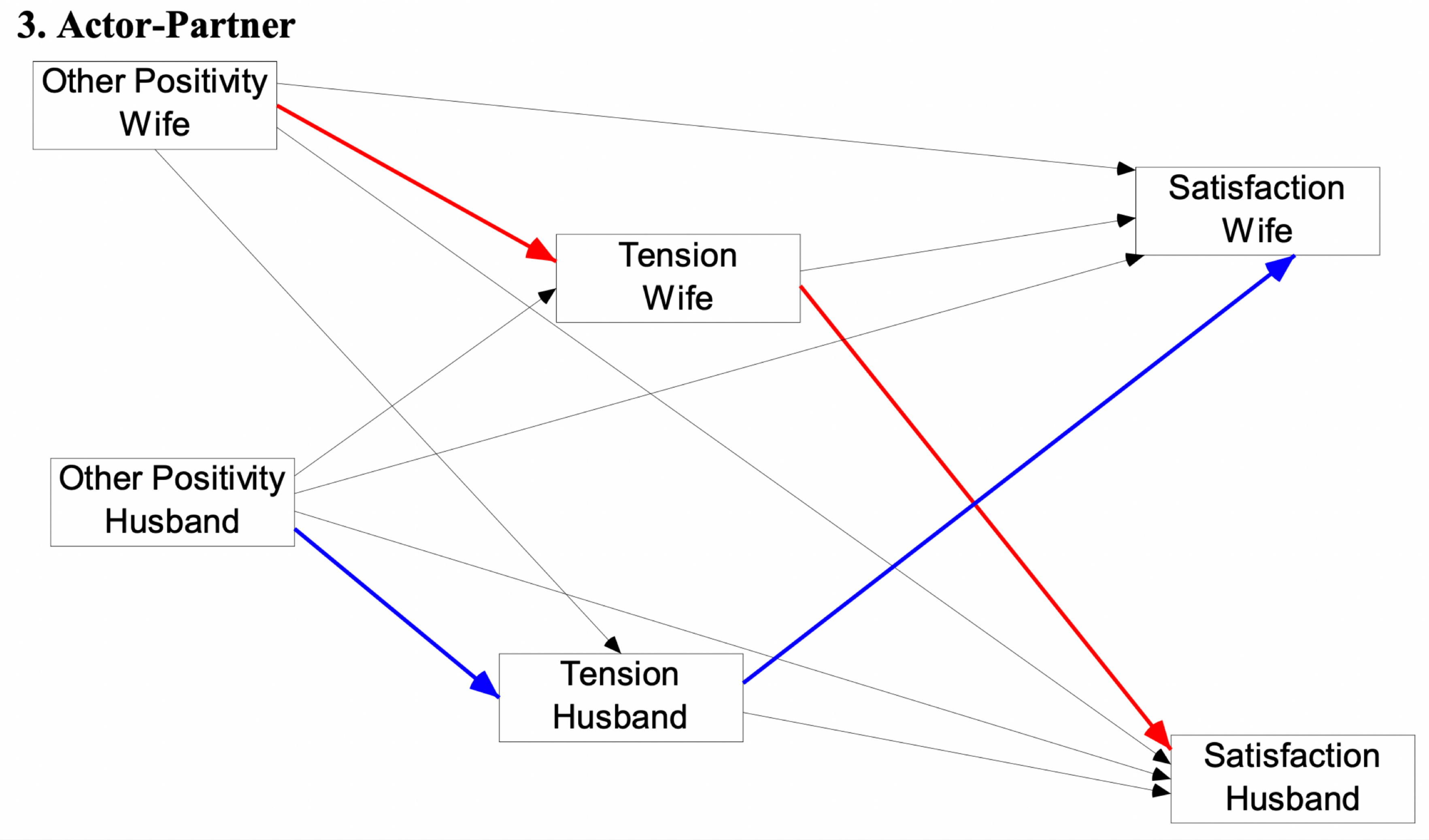
# Mediation in the Acitelli data?



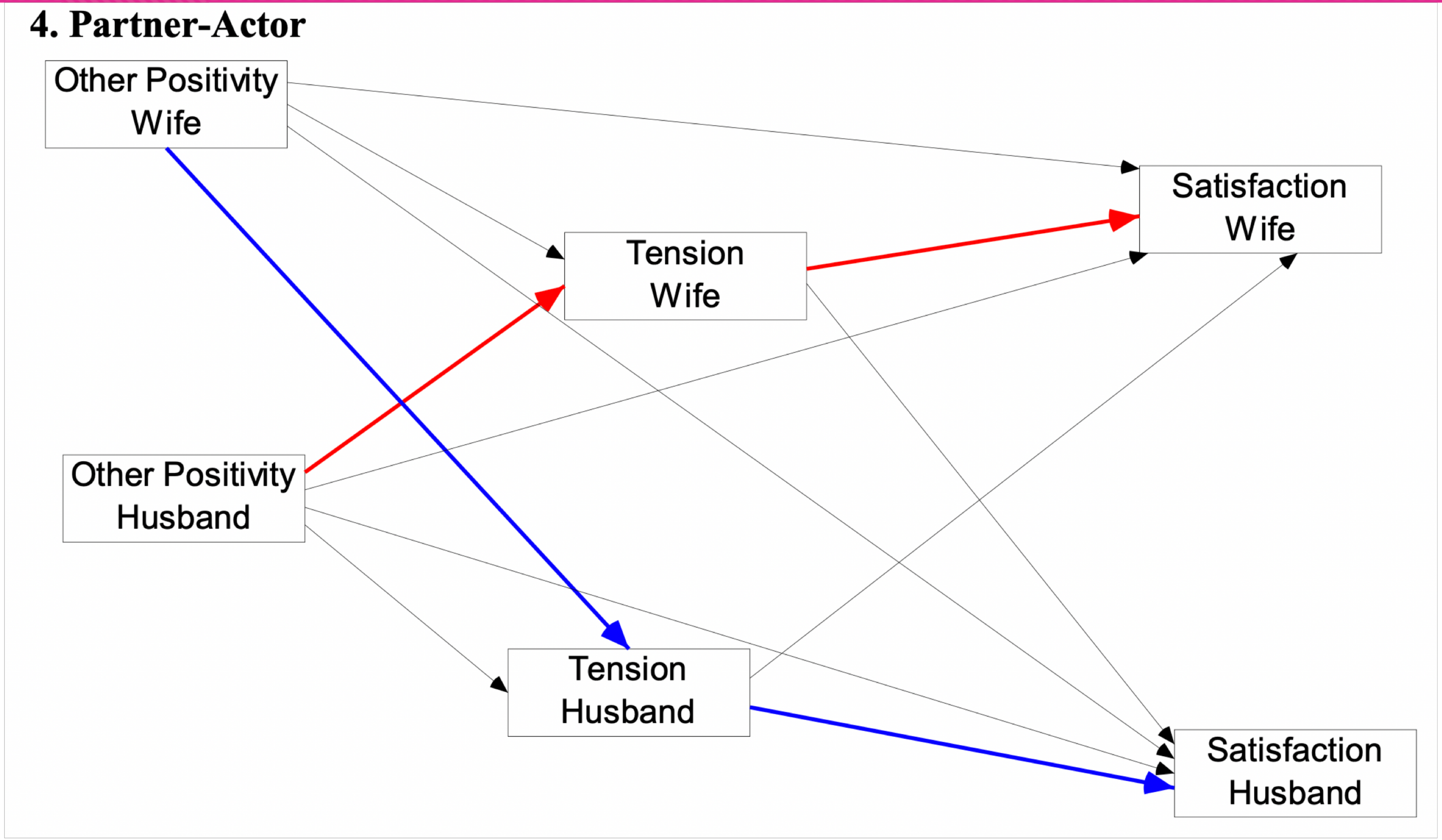
# Mediation in the Acitelli data?



# Mediation in the Acitelli data?



# Mediation in the Acitelli data?



# Causal Inference

# Causal Models

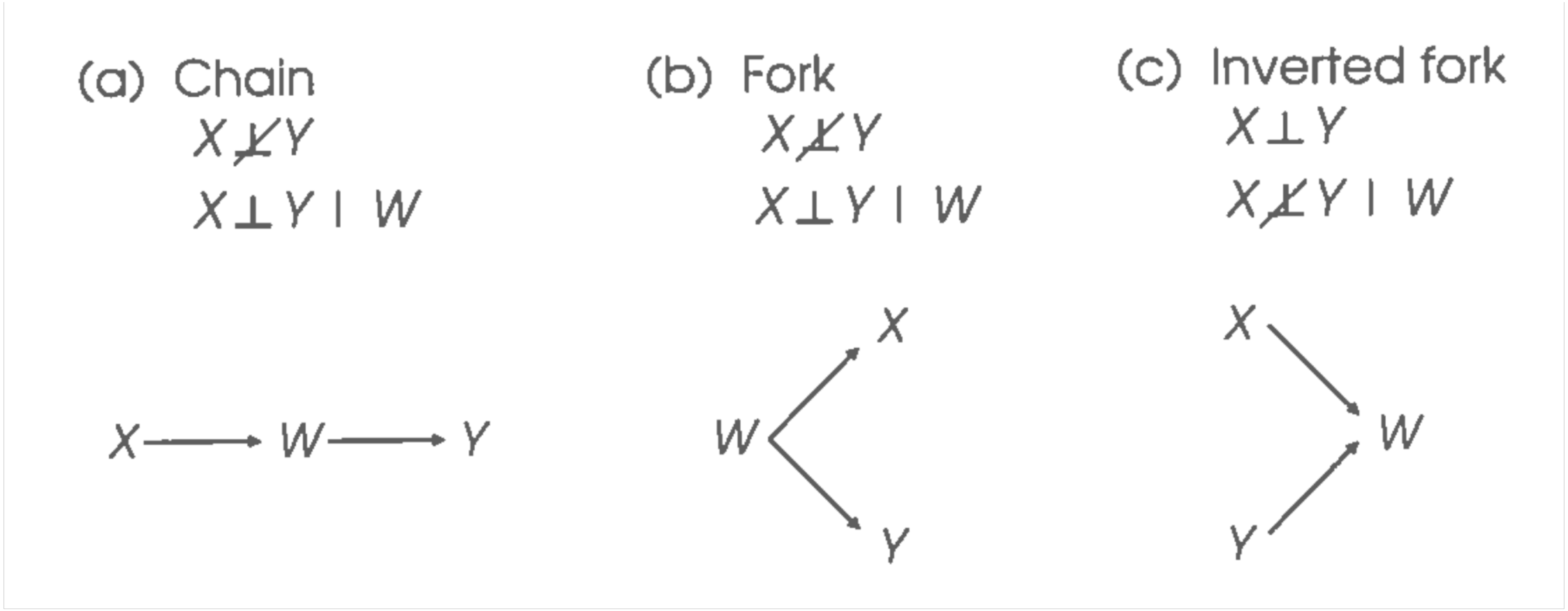
- Causal Modeling aka Structural Causal Models aka Causal inference
  - A way of depicting the causal relations among variables such that you can see which variables are implied to be conditionally independent, and thus, testable.
- Directed Acyclic Graphs (DAG)
  - No causal loops
- Directed Cyclic Graphs (DCG)
  - Causal loops

# Casual Models (DAGs and DCGs)

Just as in Structural Equation Models, Causal Models have their own names for things in the visualization.

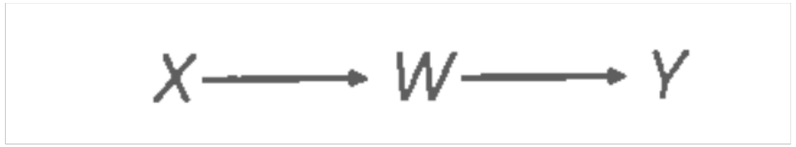
- Nodes – the variables
- Edge – link between nodes
- Path – a sequence of adjacent edges
- Parents – the direct causes of a variable
  - Ancestors – the direct and indirect causes of a variable
- Children – the variables directly caused by a variable
  - Descendants - all variables directly and indirectly caused by the variable

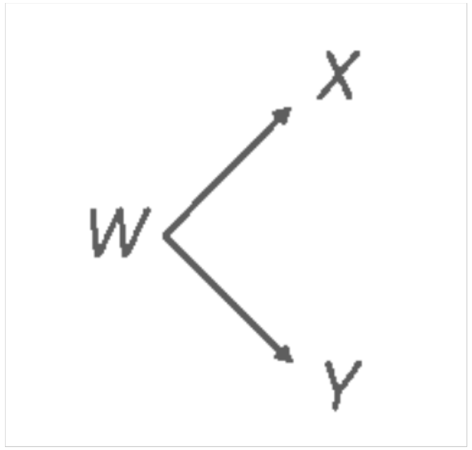
# Causal Models

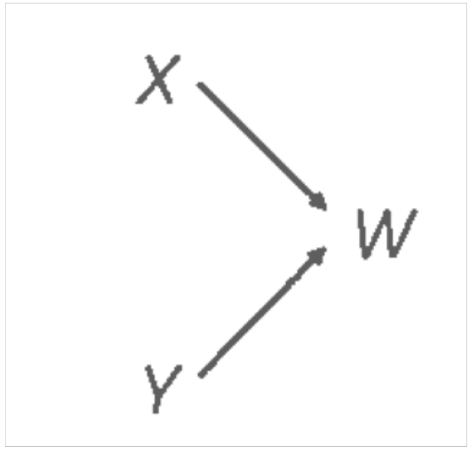


# Casual Models

- Indirect effects
- Common cause
  - Your garden variety confounder
  - Example: Ice cream sales and murder
- Collider
  - [Blog post on dangers of conditioning on a collider]



# d-Separation

- Finding all the pairs of variables that can be d-separated (conditional independence) in the model points to testable hypotheses.
- A pair of variables in a DAG is d-separated by a set of covariates, Z, if either

1. One of the noncolliders on the path is in Z; or
2. There is a collider on the path, but neither the collider nor neither the collider nor any of its descendants is in Z

$$X \rightarrow A \rightarrow B \rightarrow Y$$

$$X \perp Y \mid A \quad X \perp Y \mid B \quad \text{and} \quad X \perp Y \mid (A, B)$$

$$X \perp B \mid A \quad \text{and} \quad X \perp B \mid (A, Y)$$

$$A \perp Y \mid B \quad \text{and} \quad A \perp Y \mid (B, X)$$

# d-Separation

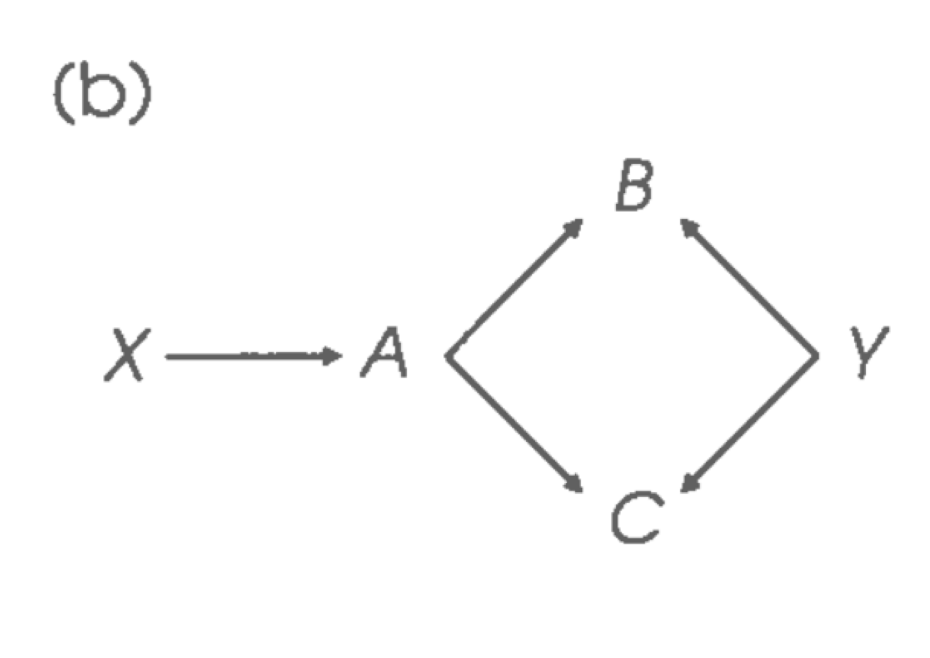



**TABLE 8.1. Conditional Independences Located by the d-Separation Criterion in Figure 8.2(b)**

| Nonadjacent pair | Conditional independences | |
|---|---|---|
| $X, Y$ | $X \perp Y$ | $X \perp Y \mid A$ |
| | $X \perp Y \mid (A, B)$ | $X \perp Y \mid (B, C)$ |
| | $X \perp Y \mid (A, B, C)$ | $X \perp Y \mid (A, C)$ |
| $X, B$ | $X \perp B \mid A$ | $X \perp B \mid (A, C)$ |
| | $X \perp B \mid (A, Y)$ | $X \perp B \mid (A, C, Y)$ |
| $X, C$ | $X \perp C \mid A$ | $X \perp C \mid (A, B)$ |
| | $X \perp C \mid (A, Y)$ | $X \perp C \mid (A, B, Y)$ |
| $B, C$ | $B \perp C \mid (A, Y)$ | $B \perp C \mid (A, X, Y)$ |
| $A, Y$ | $A \perp Y$ | $A \perp Y \mid X$ |

# d-Separation

- Basis set
  - Smallest number of conditional independencies (d-separations) that imply all others.
  - All other conditional independences are implied by the basis set, and thus are redundant and do not need to be tested.

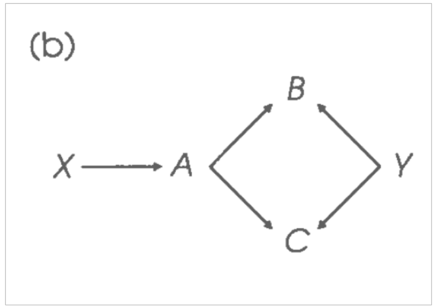



**TABLE 8.1. Conditional Independences Located by the d-Separation Criterion in Figure 8.2(b)**

| Nonadjacent pair | Conditional independences | |
|---|---|---|
| $X, Y$ | $X \perp Y$ | $X \perp Y \mid A$ |
| | $X \perp Y \mid (A, B)$ | $X \perp Y \mid (B, C)$ |
| | $X \perp Y \mid (A, B, C)$ | $X \perp Y \mid (A, C)$ |
| $X, B$ | $X \perp B \mid A$ | $X \perp B \mid (A, C)$ |
| | $X \perp B \mid (A, Y)$ | $X \perp B \mid (A, C, Y)$ |
| $X, C$ | $X \perp C \mid A$ | $X \perp C \mid (A, B)$ |
| | $X \perp C \mid (A, Y)$ | $X \perp C \mid (A, B, Y)$ |
| $B, C$ | $B \perp C \mid (A, Y)$ | $B \perp C \mid (A, X, Y)$ |
| $A, Y$ | $A \perp Y$ | $A \perp Y \mid X$ |

# Project time