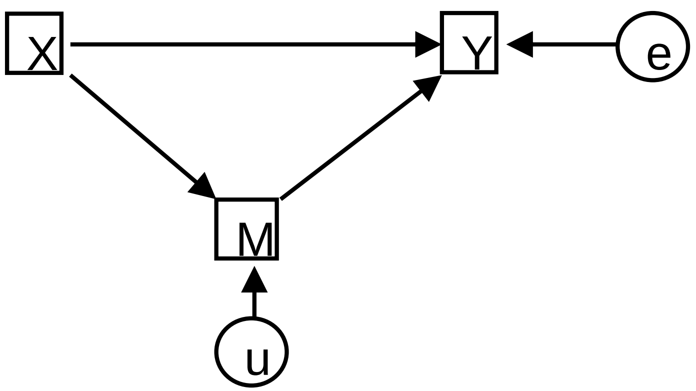
TCSM

Mediation



OUTLINE

- Relationships among 3 variables
- investigating mediation
- testing the indirect (or mediated) effect

Akademie für Soziologie - Academy of Sociology
@akadsoz

Sharing childcare matters, new @ESR_news paper by @renske_keizer , @cjvanlissa & colleagues.
Ungated paper!

ESR @ESR_news · 1h

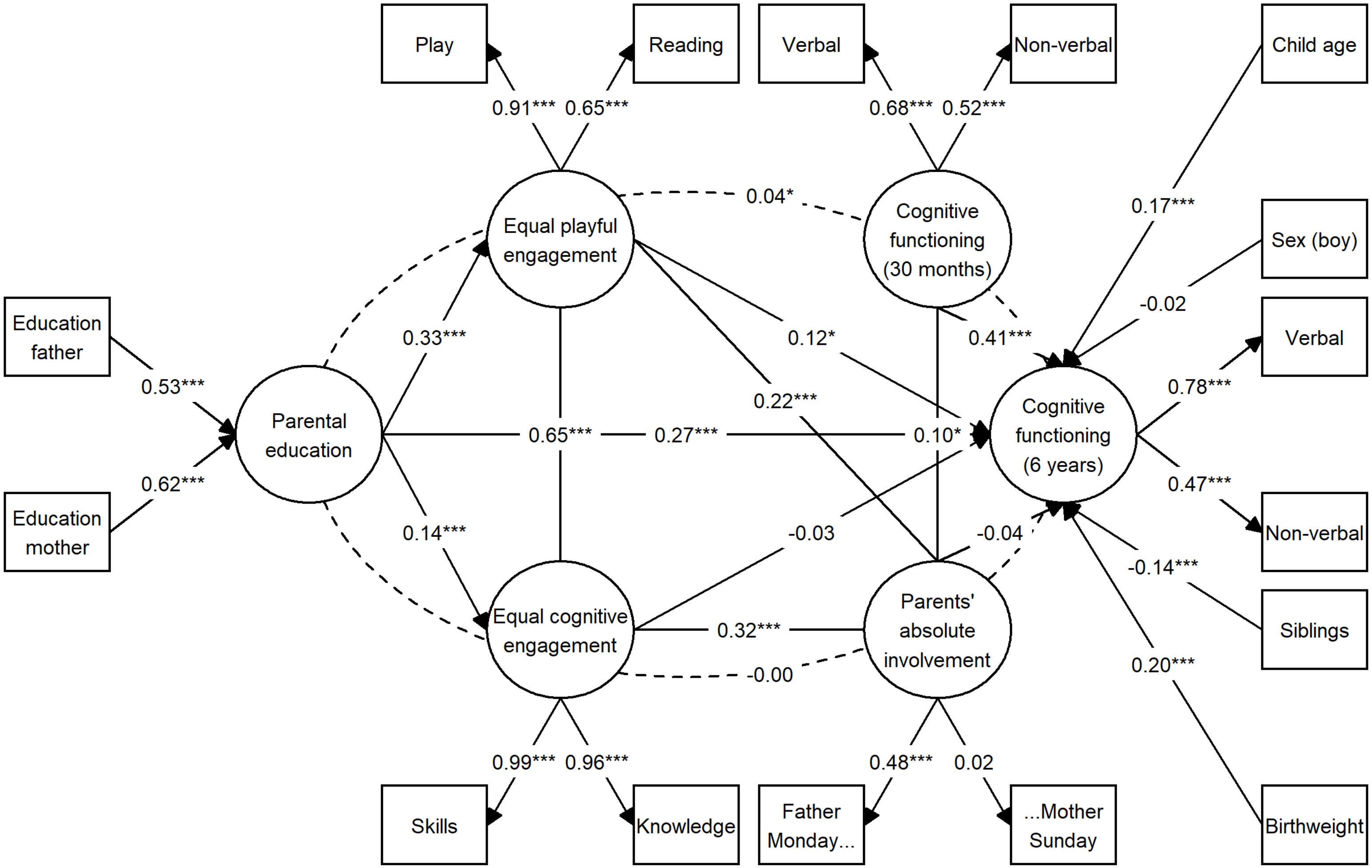
New paper by Keizer and colleagues shows that parents' equally sharing childcare responsibilities functions as an underlying mechanism for social class disparities in children's cognitive development doi.org/10.1093/esr/jc...

European Sociological Review, 2019, 1–15
doi: 10.1093/esr/jcz046
Original Article

OXFORD

The Influence of Fathers and Mothers Equally Sharing Childcare Responsibilities on Children's Cognitive Development from Early Childhood to School Age: An Overlooked Mechanism in the Intergenerational Transmission of (Dis)Advantages?

Renske Keizer[1,*], Caspar J. van Lissa[2], Henning Tiemeier[3,4] and Nicole Lucassen[5]



Play
Reading
Verbal
Non-verbal
Child age
0.91***
0.65***
0.68***
0.52***
Equal playful engagement
0.04*
Cognitive functioning (30 months)
0.17***
Sex (boy)
-0.02
Education father
0.33***
0.12*
0.41***
Verbal
0.53***
Parental education
0.22***
0.78***
0.65***
0.27***
0.10*
Cognitive functioning (6 years)
0.62***
0.47***
Education mother
0.14***
-0.03
-0.04
Non-verbal
-0.14***
Equal cognitive engagement
0.32***
Parents' absolute involvement
Siblings
-0.00
0.20***
0.99***
0.96***
0.48***
0.02
Skills
Knowledge
Father Monday...
...Mother Sunday
Birthweight

Mediation

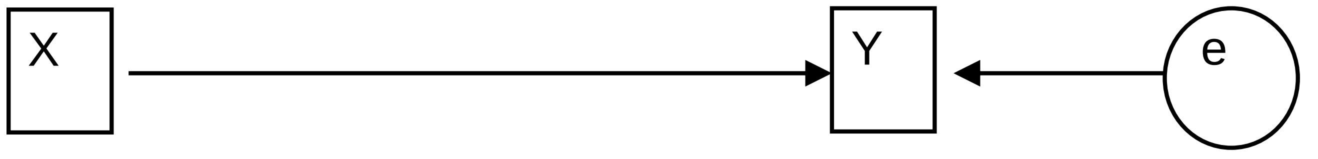
Snooping on your teenager leads to more parent-child conflict

WHY?

Because snooping interferes with teenagers' **autonomy needs**

Mediation

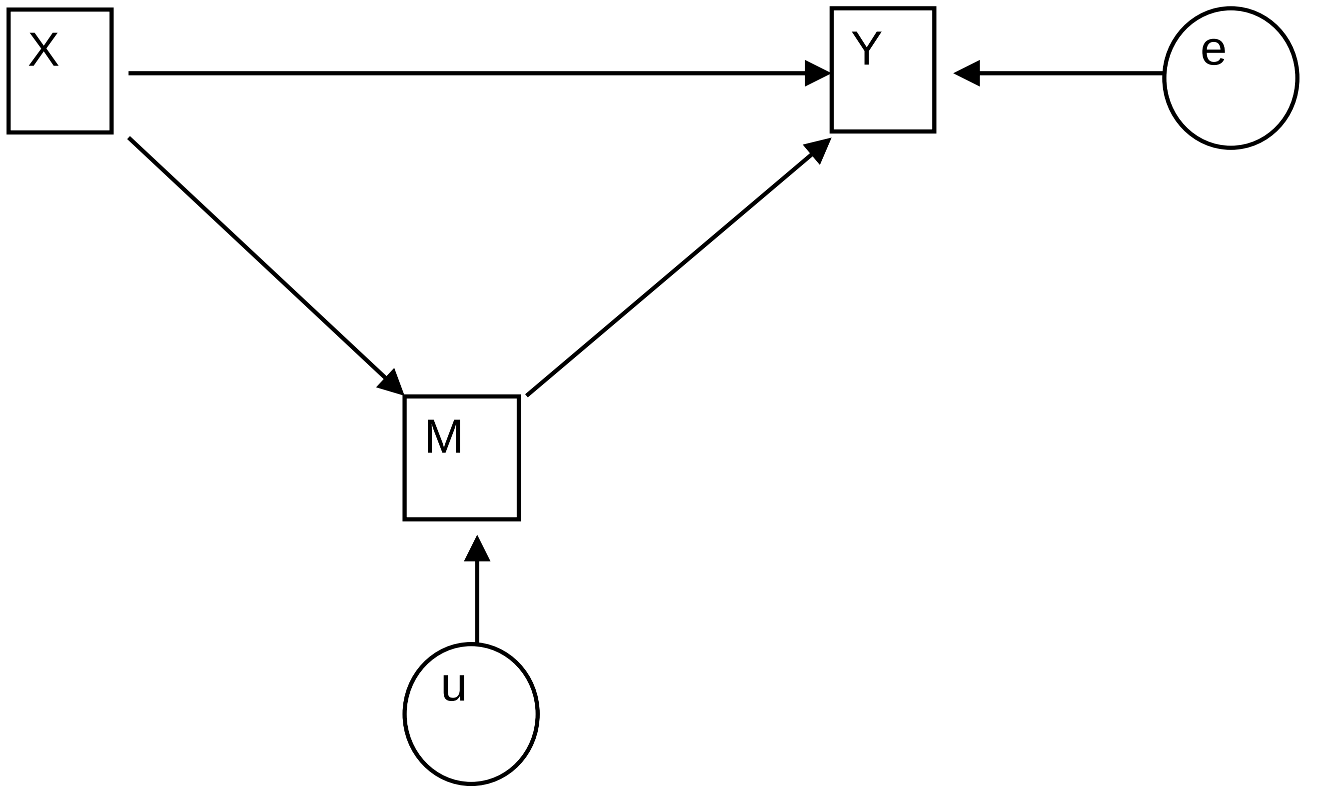
Snooping leads to more parent-child conflict



u

Mediation

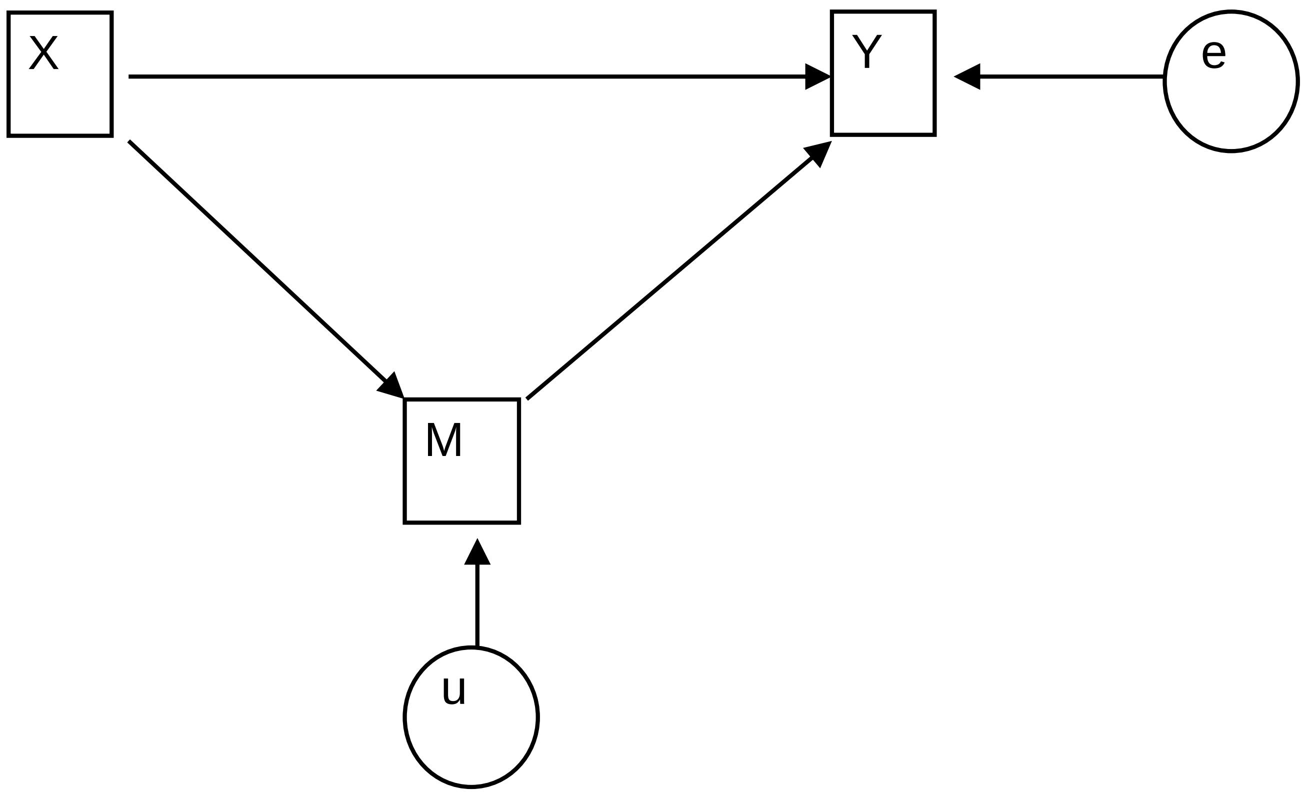
…because snooping frustrates autonomy needs!



Mediation

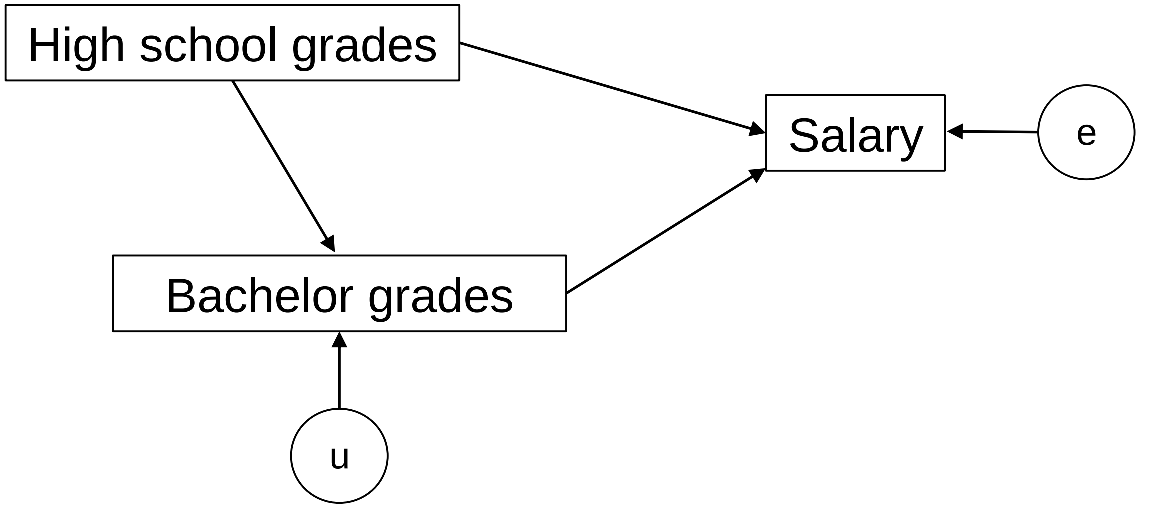
The effect of X on Y is (partially) **mediated** by M

Mediated: Explained by



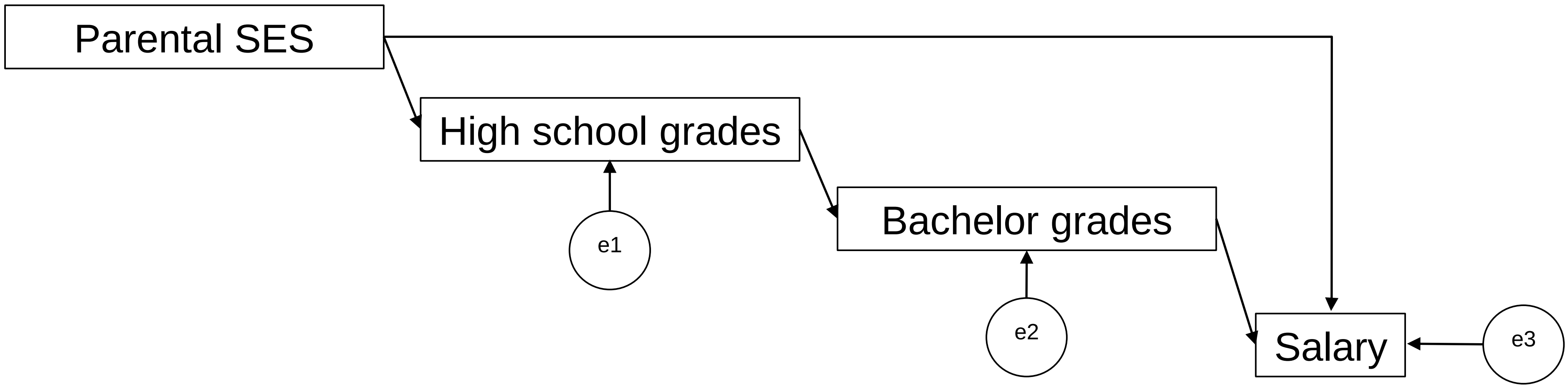
Mediation

X influences *Y* through a third variable: Mediator *M*.



Mediation

X influences *Y* through a third variable: Mediator *M*.



Effects in a mediation model

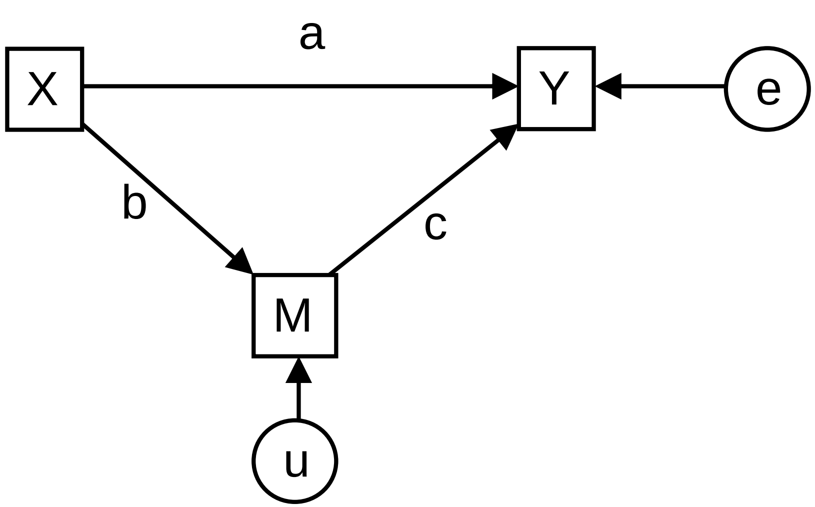
Direct and indirect effects

The influence of X on Y is (partially) mediated by M, *We also say: X has an indirect effect on Y*

Direct effect of X on Y: a

Indirect effect of X on Y: b*c

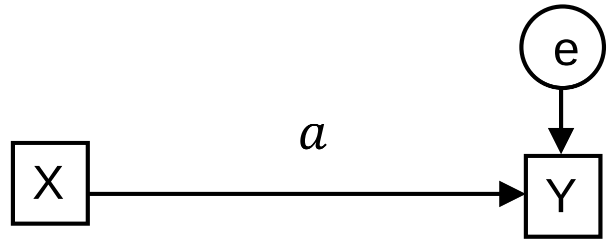
Total effect of X on Y: a + (b*c)



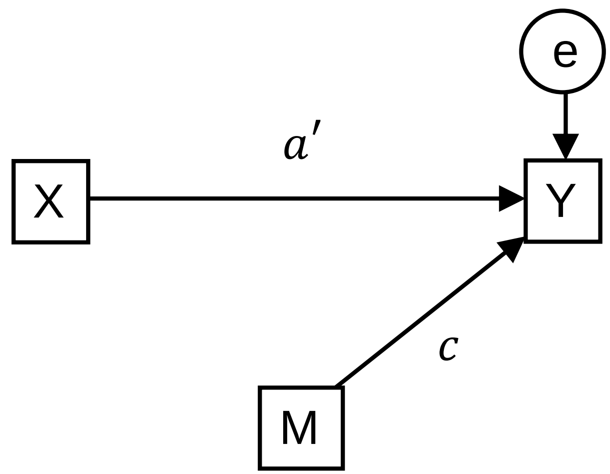
Determine indirect effect using regression:

Approach 1:

1) Simple regression



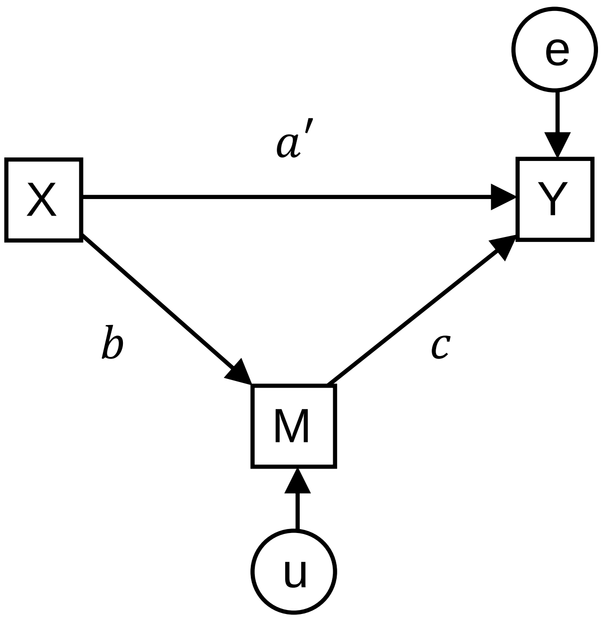
2) Multiple regression



Indirect effect: $a \quad - \ a'$

Determine indirect effect in SEM:

Approach 2:



Indirect effect: $b * c$

OUTLINE

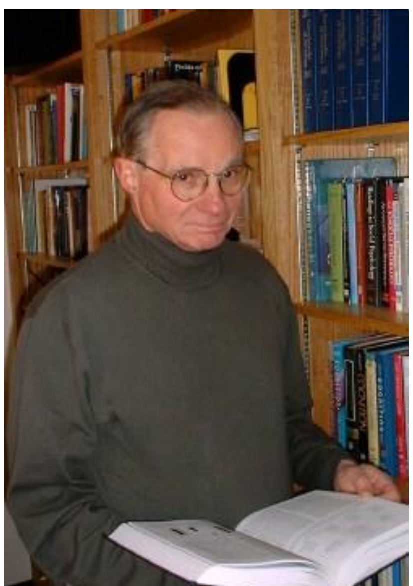
- relationships between 3 variables
- investigating mediation
- testing the indirect (or mediated) effect

Investigating mediation: old school

Baron and Kenny steps (cited 14.000+ times!!!).

Journal of Personality and Social Psychology
1986, Vol. 51, No. 6, 1173–1182

The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations



Reuben M. Baron and David A. Kenny
University of Connecticut

In this article, we attempt to distinguish between the properties of moderator and mediator variables at a number of levels. First, we seek to make theorists and researchers aware of the importance of not using the terms *moderator* and *mediator* interchangeably by carefully elaborating, both conceptually and strategically, the many ways in which moderators and mediators differ. We then go beyond this largely pedagogical function and delineate the conceptual and strategic implications of making use of such distinctions with regard to a wide range of phenomena, including control and stress, attitudes, and personality traits. We also provide a specific compendium of analytic procedures appropriate for making the most effective use of the moderator and mediator distinction, both separately and in terms of a broader causal system that includes both moderators and mediators.

Three steps of B&K

Several regressions, puzzle together the path model:

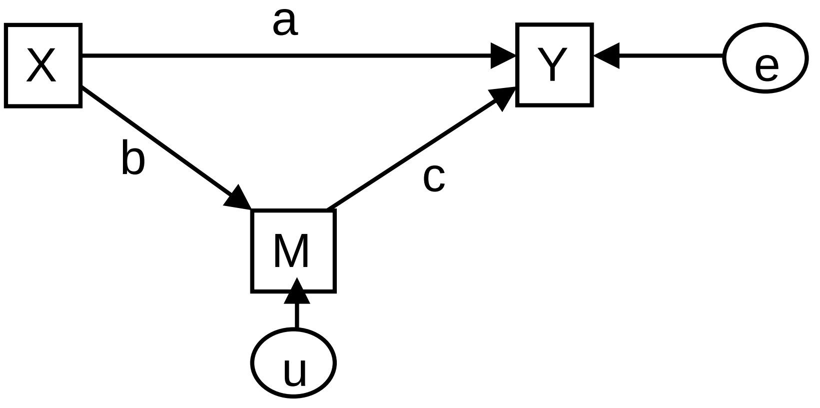
Step 1: Is X a significant predictor of Y? (a)

Step 2: Is X a significant predictor of M? (b)

Step 3: In model with both M and X as predictors, is M a significant predictor of Y? (c).

Did *a* decrease?

(or increase?)



Some problems with B&K

If there is a **suppression effect**

*(the direct and indirect effects **cancel out**), then*

<u>Step 1 would not show a significant effect</u>

Also: Low power when using B&K three steps

Investigating mediation: new school

Use **Structural Equation Modeling**

Advantages:

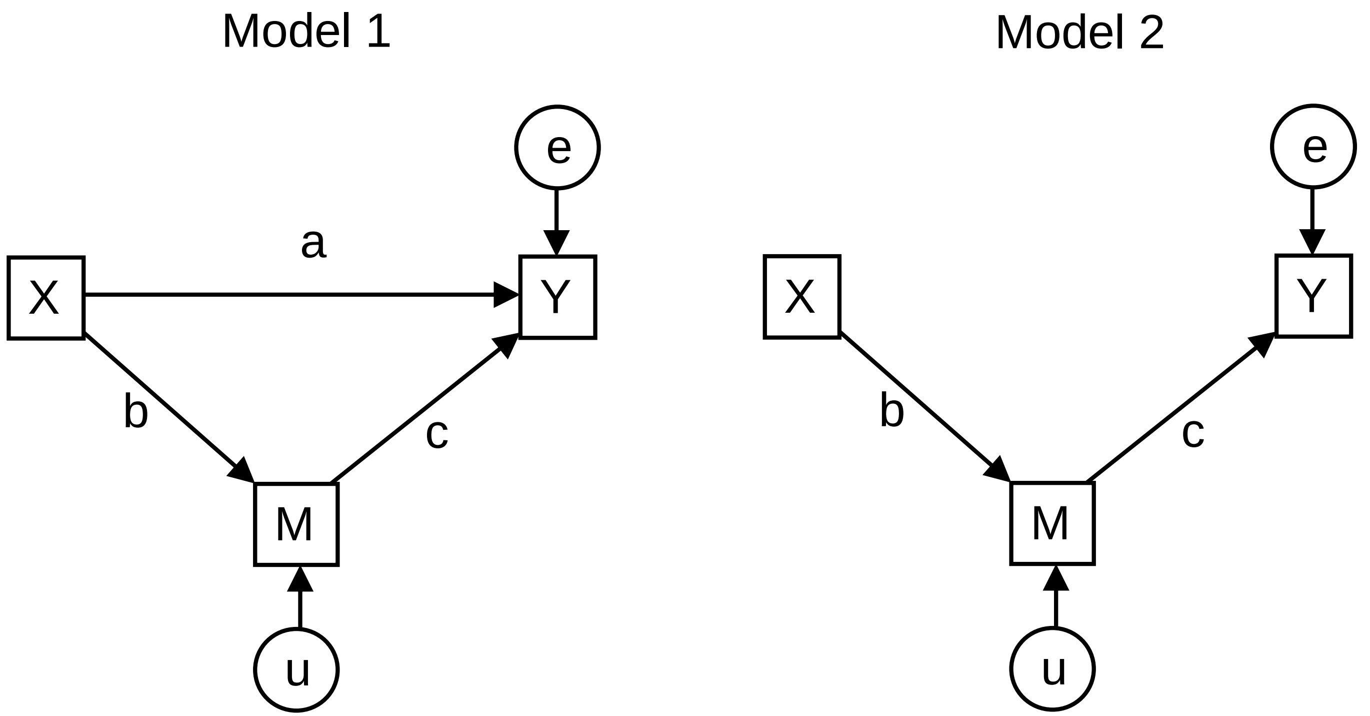
- easier (run 1 model, or 2 nested models)
- Automatically get direct, indirect and total effects; also standardized, and with SEs
- Easily investigate more complicated mediation, e.g.:
  - multiple mediators of one predictor,
  - multiple predictors with one mediator,
  - multiple outcome variables,
  - latent variables

Mediation in SEM

To investigate whether the effect is **fully mediated**, we can do two things:

1: Check the **significance** of coefficient a

2: Compare **nested models** (which is more complex?)



Nested models in SEM

Nested: By constraining (to be equal/to be zero) some parameters in model 1, you get model 2.

Compare nested models with a **chi-square diff test,**

$$\Delta\chi^2$$

Both models have a **model-implied vcov matrix,** $\widehat{\boldsymbol{\Sigma}}$

These are compared to the **observed vcov matrix, S**

$\Delta\chi^2$ is based on comparing the "distance" between **S** and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{1}}$ with the "distance" between **S** and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{2}}$

Nested models in SEM

- Remember Occam's razor:
  All else being equal, we should prefer simpler models
- Complex models have more "flexibility" to fit data
- Balance necessary complexity and elegant simplicity
- Model 2 has 1 parameter less; does this simplification make the fit significantly worse
- If difference is significant, model 2 isn't supported by the sample covariance matrix

Nested SEM models: removing path

Is the distance larger?

larger Chi-square

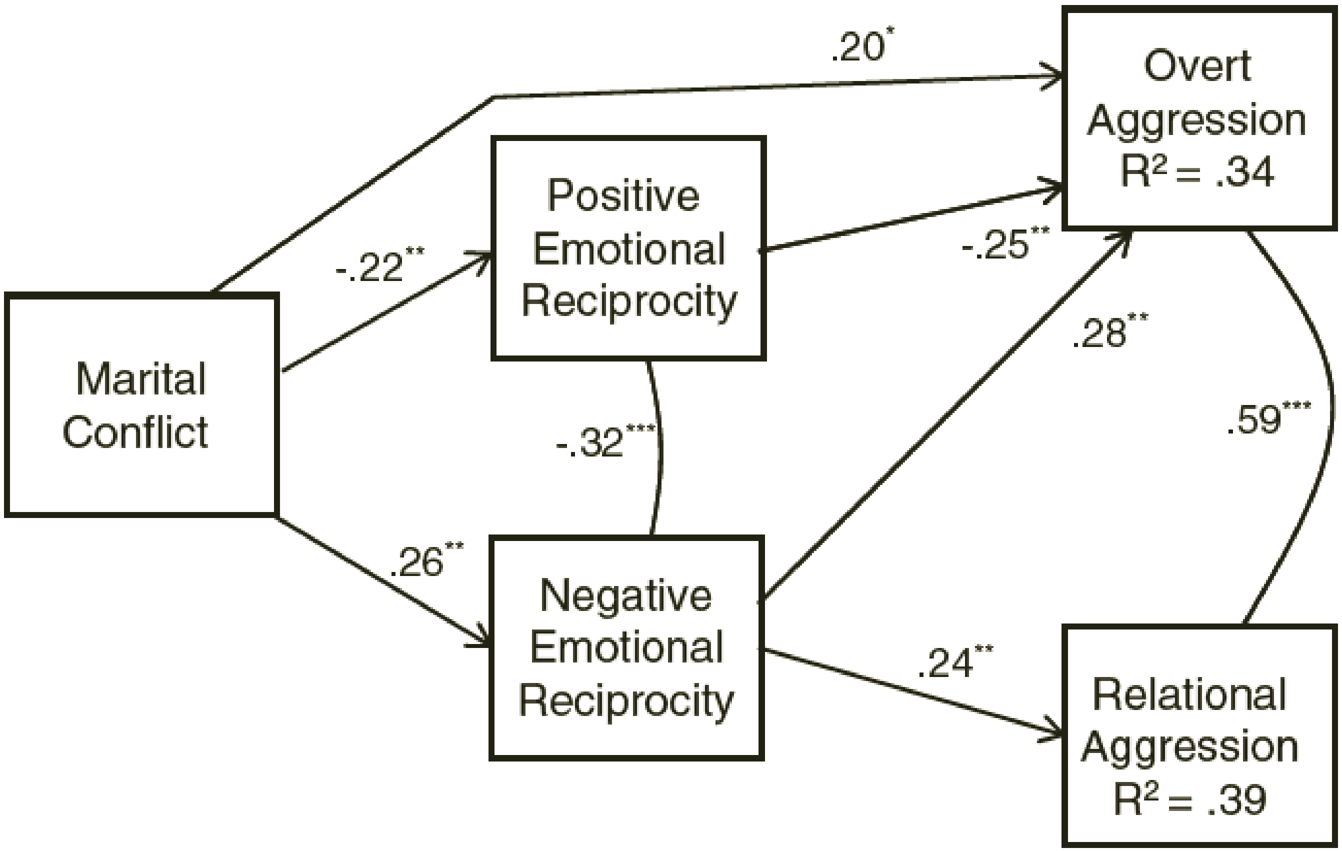
significant result

Model 2 is not an improvement, Model 1 is better

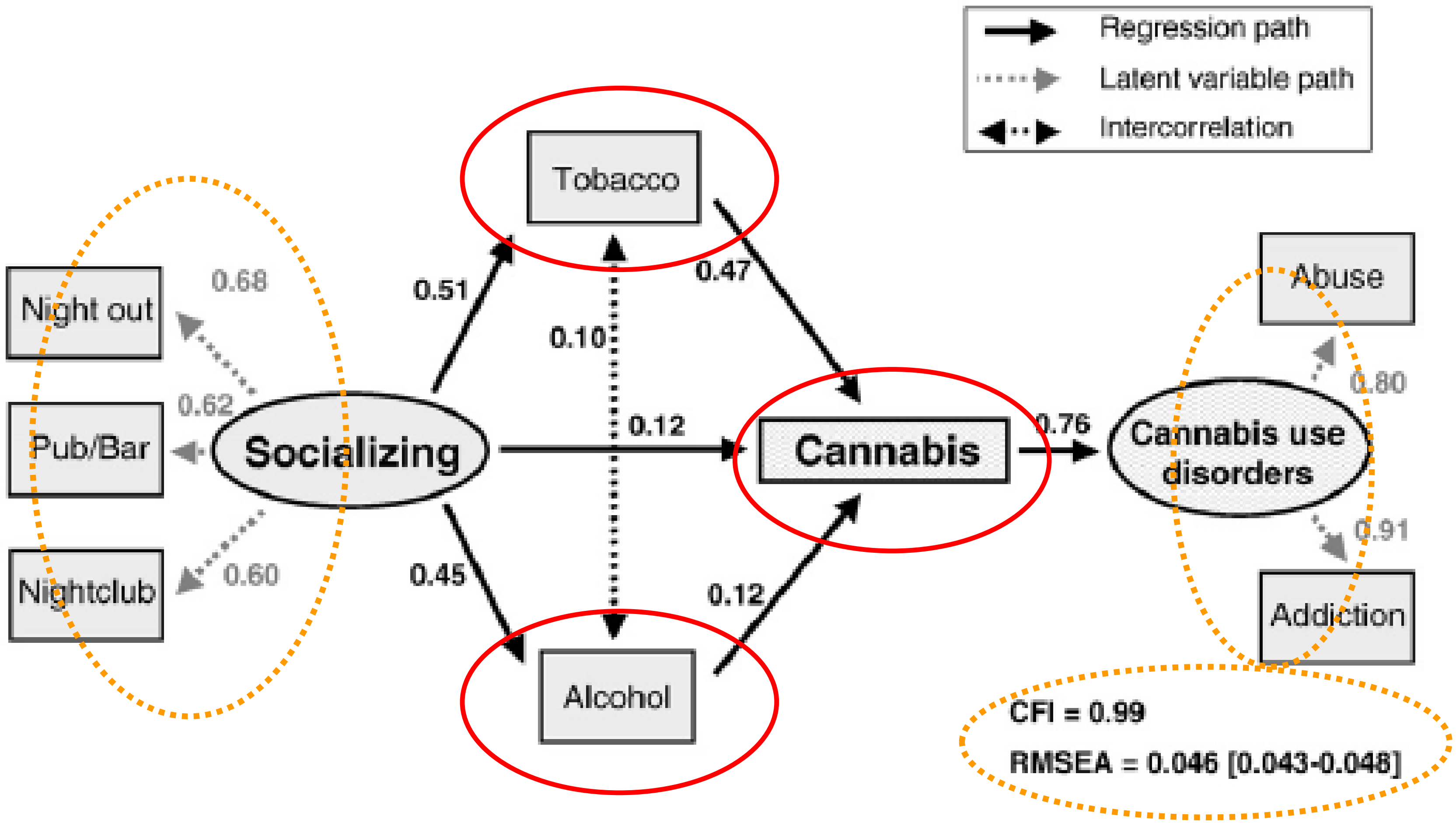
In contrast, a non-significant result
means that Model 2 fits equally well, but is simpler:
choose model 2

Applied mediation examples

Aggression in adolescents



Cannabis use disorders



Emo. intelligence and life satisfaction

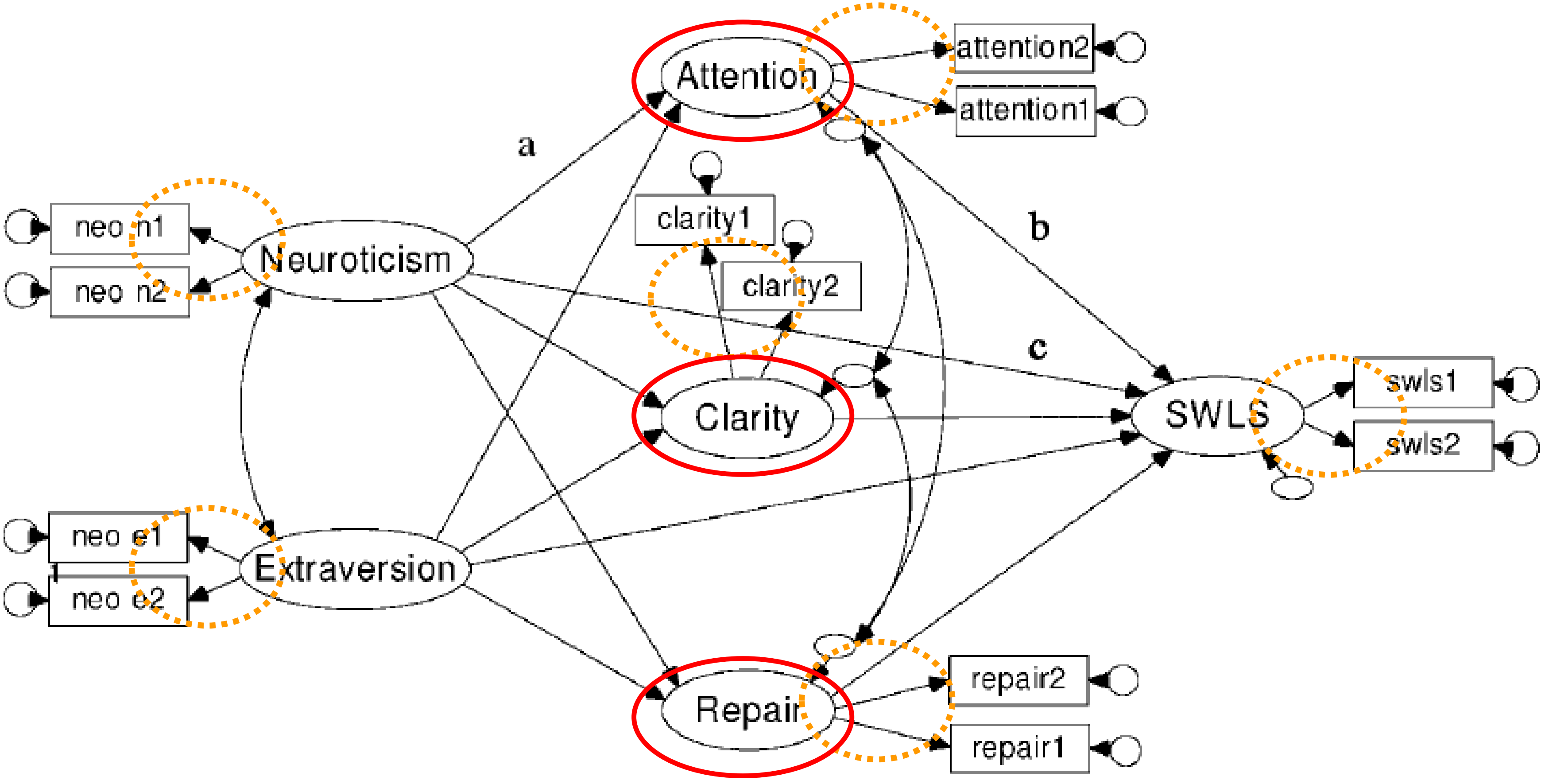


Table 1
SWLS Items and Factor Loadings

Item	Factor Loadings	Item-Total Correlations
1. In most ways my life is close to my ideal.	.84	.75
2. The conditions of my life are excellent.	.77	.69
3. I am satisfied with my life	.83	.75
4. So far I have gotten the important things I want in life.	.72	.67
5. If I could live my life over, I would change almost nothing.	.61	.57

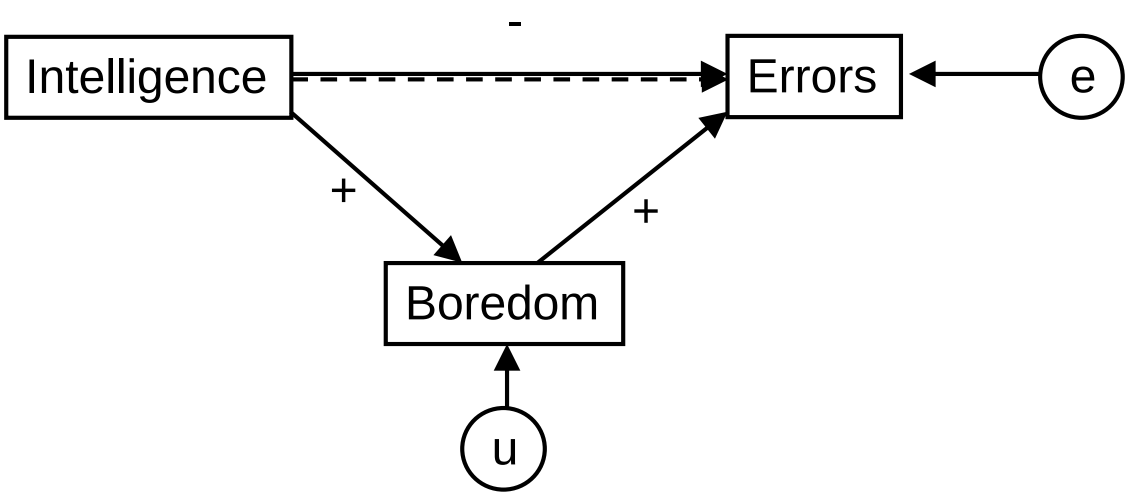
Note: $n = 176$. SWLS = Satisfaction With Life Scale.

Special cases

Suppressing mediation

Suppression:

The direct and the indirect are of **opposite signs**, and (partly) cancel each other out.



Note: In case of suppression, including the mediator will **increase** the predictive ability of X on Y.

OUTLINE

- relationships between 3 variables
- investigating mediation
- testing the indirect (or mediated) effect
  - Testing a parameter estimate (in general)
  - Testing the indirect effect (Ind := b*c)
    - Classic: Sobel test - WARNING!
    - Better: Bootstrapping procedure in SEM

Testing indirect effects

Introducing bootstrapping

Testing a parameter estimate

Central limit theorem:

The **sampling distribution** for many parameters is (approximately) **normal**.

The sampling distribution is the distribution we would get if we would:

- Take many samples from the same population of the same size
- Estimate the parameter of interest (i.e., θ) each time

Testing a parameter estimate

1) Estimate the parameter in the sample, e.g. T

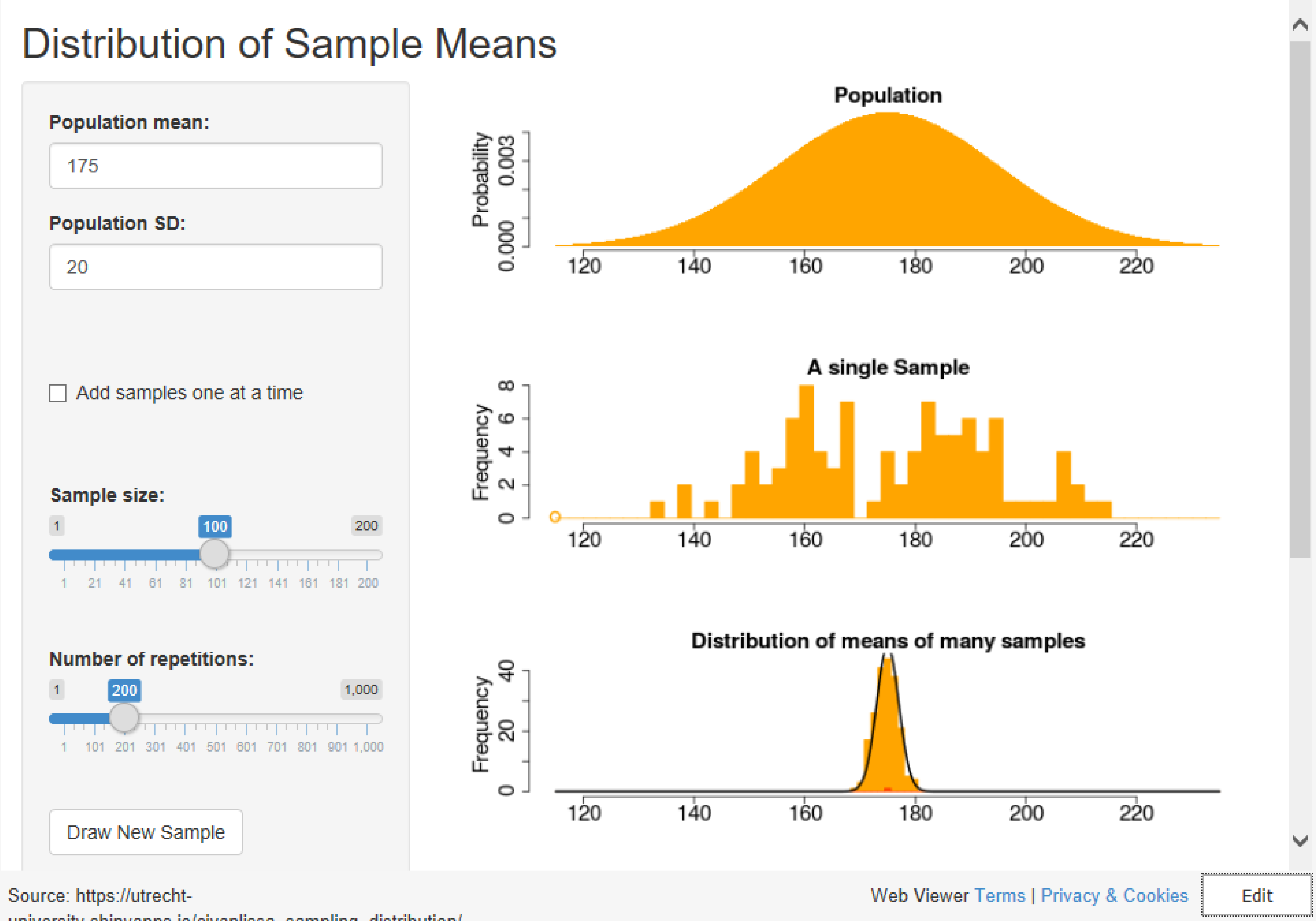
2) Estimate the SE of the parameter, SE_T

3) Derive the sampling distribution **under the null hypothesis** (i.e., Θ_0=0, SE = SE_T)

In other words: We draw a normal distribution with mean = 0 and sd = SE_T

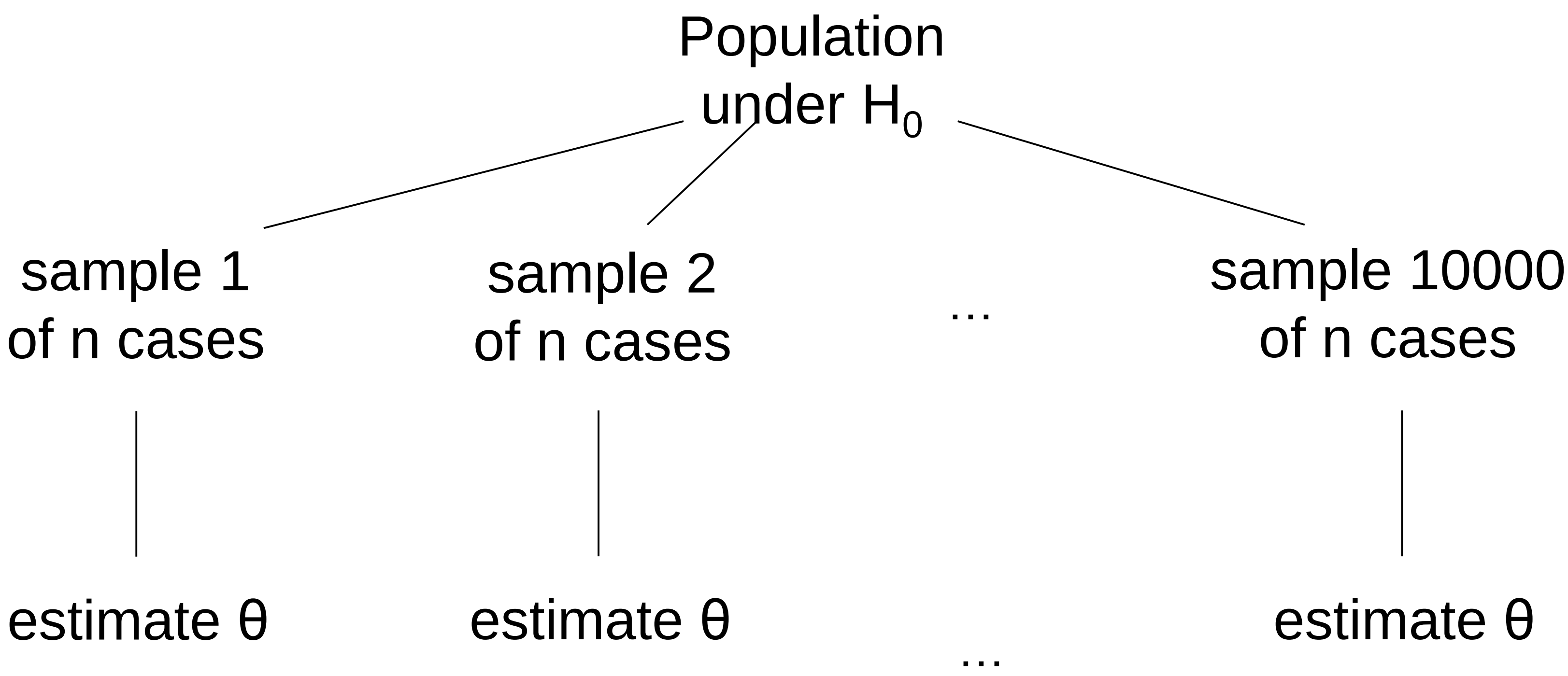
We then test:

How likely is it to get a value for T at least as extreme as we observed in our data, IF the null hypothesis were true?

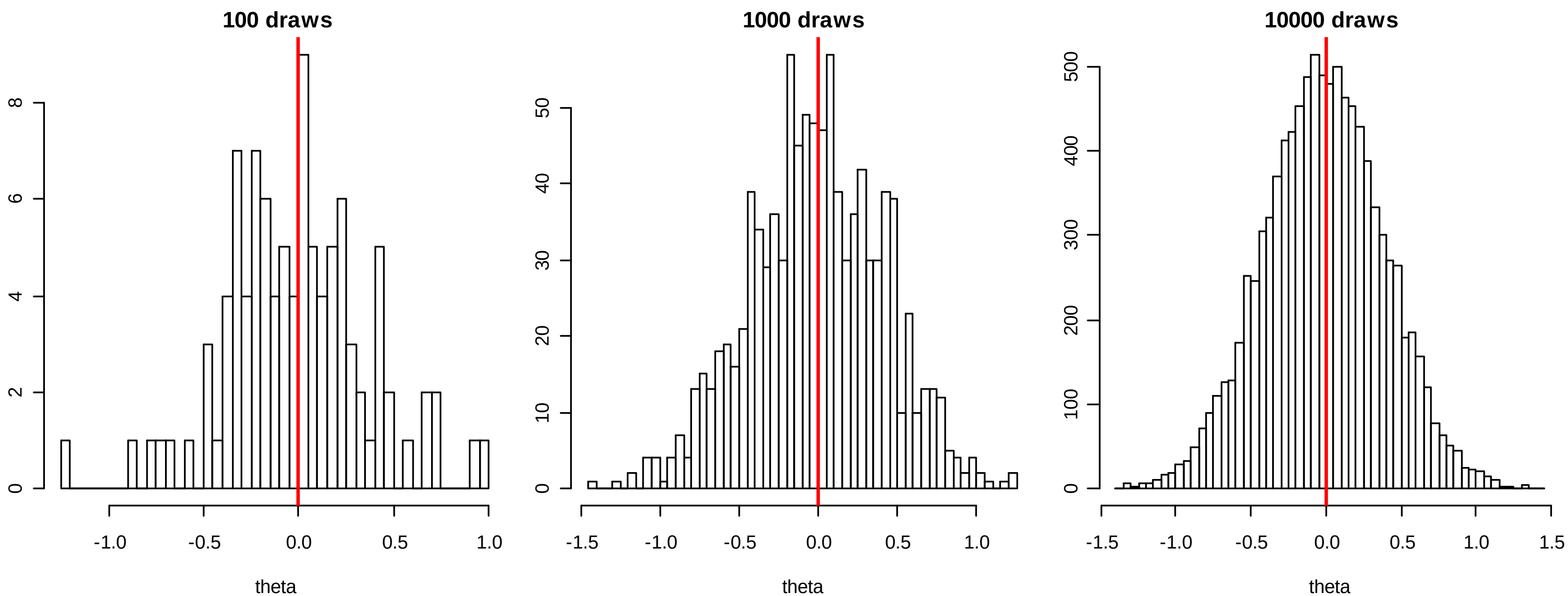


https://utrecht-university.shinyapps.io/cjvanlissa_sampling_distribution/

Estimate normally distributed?



Yes, normally distributed!



Under H_0: $$\hat{\theta} \sim N\left(\theta = 0, \sigma^2_{\hat{\theta}(n)}\right)$$

Standard error

The **standard error** is an estimate of the **standard deviation** of the sampling distribution.

Hence, it can be used to compute a z-statistic and matching p-value (under H_0: $\theta=0$):

$$z = \frac{\hat{\theta}}{SE_{\theta}}$$

Alternatively, one can compute a 95%-confidence interval around the parameter estimate:

$$\hat{\theta} \pm 1.96 * SE_{\theta}$$

Test indirect effect: Sobel test

The **sampling distribution** for many parameters is (approximately) **normal**.

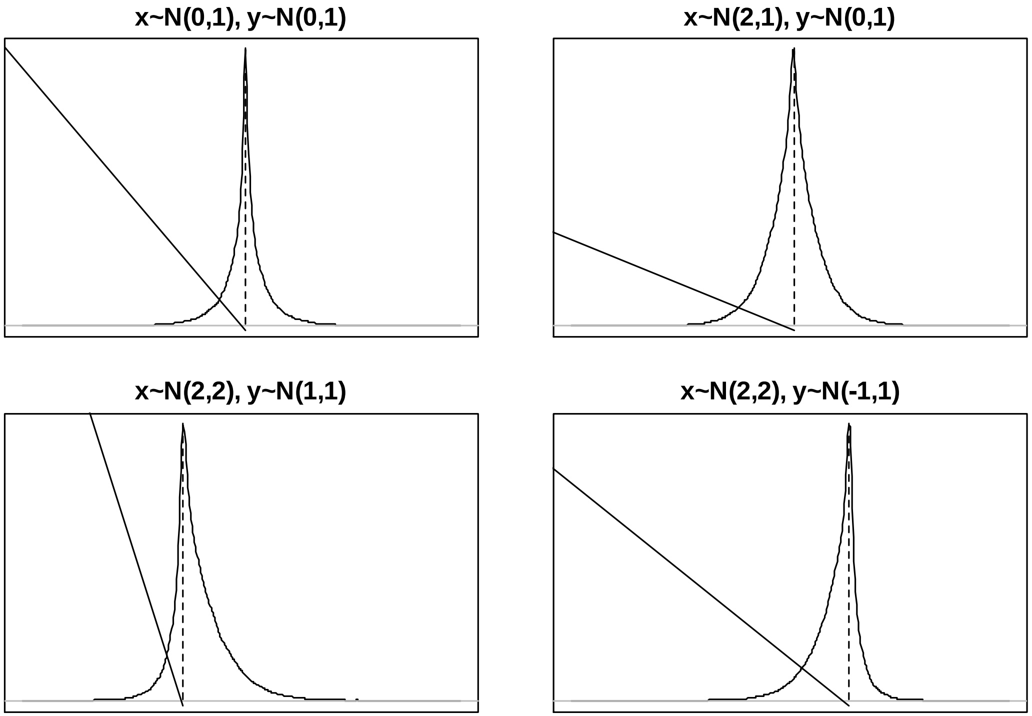
Hence a **z-test** is appropriate.

Sobel test for an indirect effect is based on the **assumption** that the sampling distribution of the product of coefficients (b*c) is normal.

Normal product distribution

The **indirect effect** (=b*c) is the product of two normally distributed variables.

This does **not** result in a normally distributed quantity!



As a result, the p-value is incorrect (may be too small or too large).

Solution: bootstrapping

We can **bootstrap** our confidence intervals:

1) Re-sample your data (1000x).

2) Estimate same model on each bootstrap sample

3) Treat the distribution of parameters across bootstrap samples as a sampling distribution

We've "empirically derived" the sampling distribution

Basics of bootstrapping

Observed sample
of n cases

Resample 1:
draw n cases
(with replacement)

Resample 2:
draw n cases
(with replacement)

…

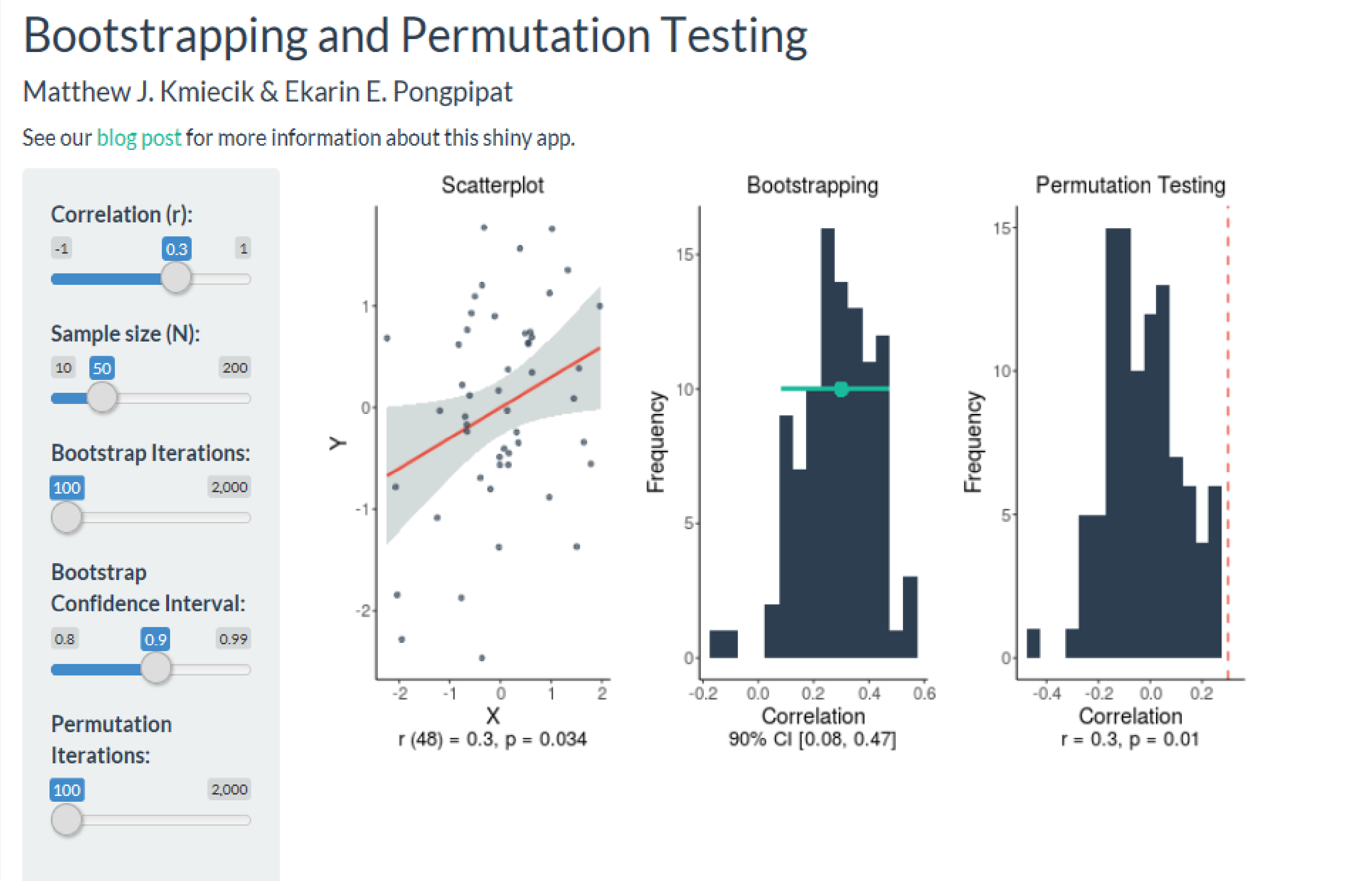
Resample 1000:
draw n cases
(with replacement)

Estimate the
indirect effect:

Estimate the
indirect effect:

…

Estimate the
indirect effect



Source: https://mattkmiecik.shinyapps.io/boot-perm-app/

https://mattkmiecik.shinyapps.io/boot-perm-app/

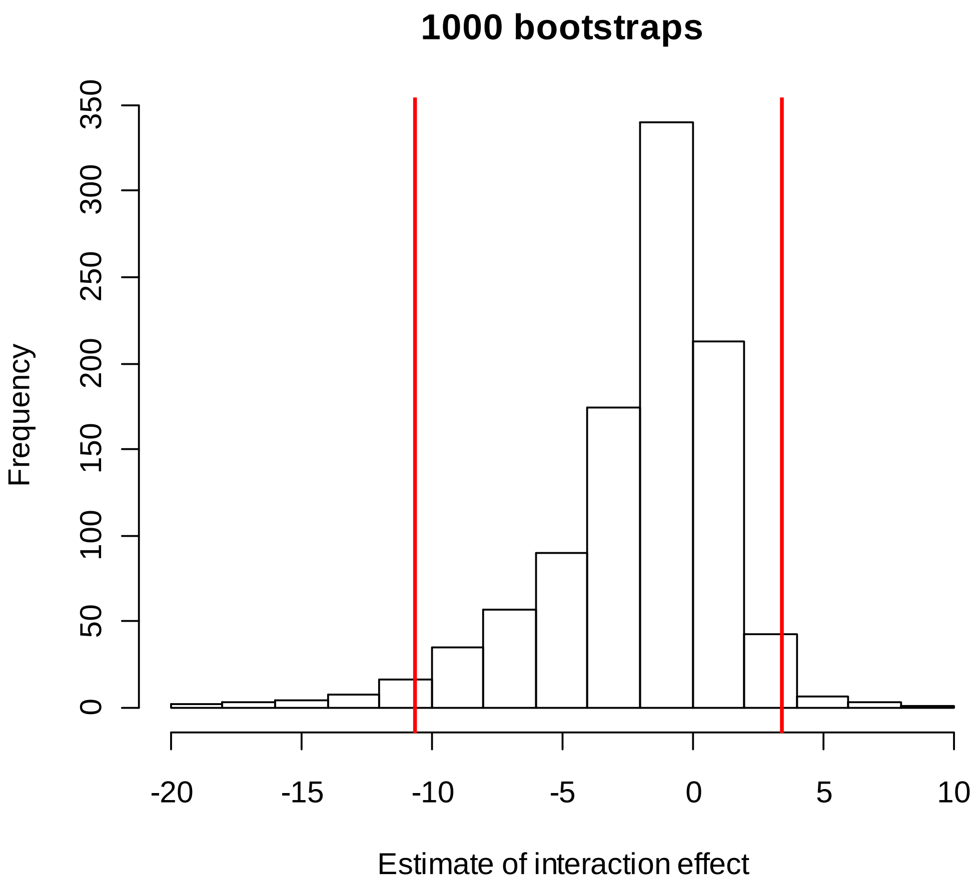
Solution in lavaan: bootstrapping

1000 bootstraps gives us:

- 1000 estimates of every parameter, including indirect effect
- The mean of these 1000 estimates = the parameter estimate
- The SD of these 1000 estimates = the SE of the parameter
- **The .025 and .975 quantiles of these 1000 estimates = the (non-parametric) 95% confidence interval**

Bootstrap confidence interval

Bootstrap samples **approximate the sampling distribution**.

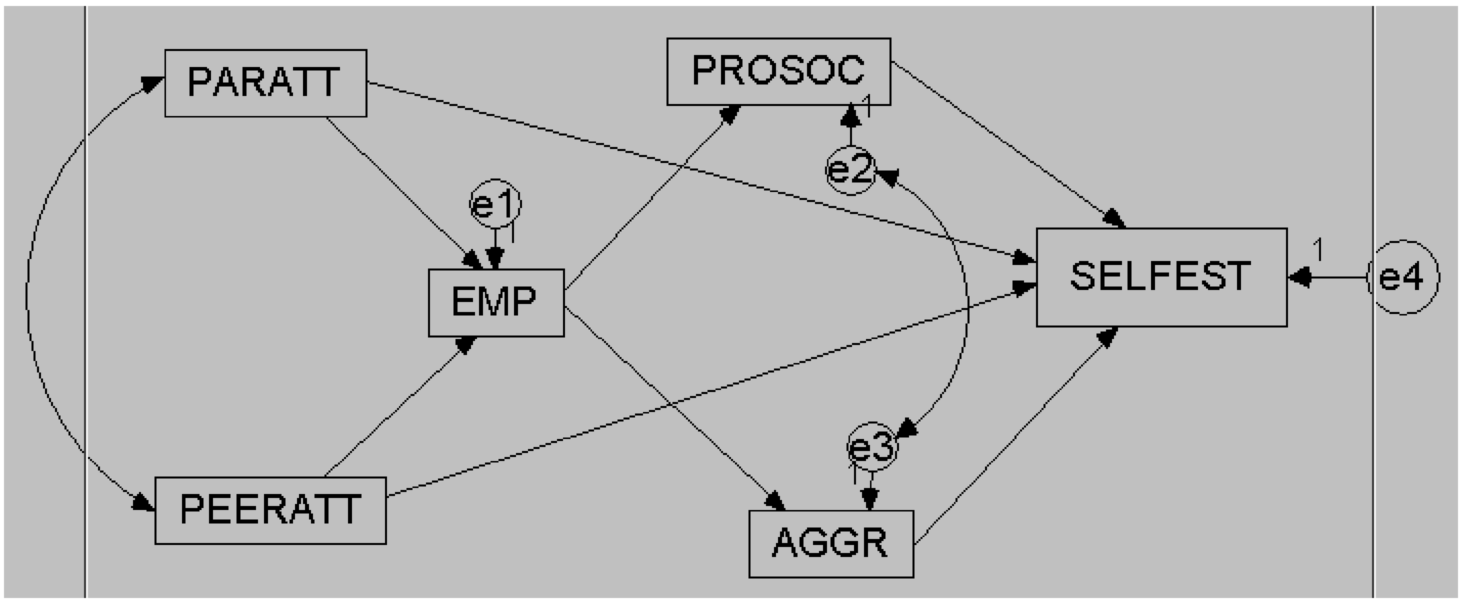


We obtain a lower and upper bound of the 95% confidence interval.

If zero lies inside this interval, we conclude the parameter estimate does not differ significantly from zero.

Hence, there is no indirect effect.

Model with indirect effects



Lavaan syntax for indirect effects

```
model <- ' # direct effect
             Y ~ c*X
           # mediator
             M ~ a*X
             Y ~ b*M
           # indirect effect (a*b)
             ab := a*b
           # total effect
             total := c + (a*b)
         '
```

Lavaan output for indirect effects

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  Y ~
    X         (c)     0.036    0.104    0.348    0.728
  M ~
    X         (a)     0.474    0.103    4.613    0.000
  Y ~
    M         (b)     0.788    0.092    8.539    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .Y                 0.898    0.127    7.071    0.000
   .M                 1.054    0.149    7.071    0.000

Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)
    ab                0.374    0.092    4.059    0.000
    total             0.410    0.125    3.287    0.001
```

Lavaan syntax for bootstrapped SE

```
fit <- sem(mediation_model,
           data = data,
           se = "bootstrap",
           bootstrap = 1000)
```

To obtain the confidence intervals, use the following syntax:

```
parameterestimates(fit, boot.ci.type = "bca.simple")
```

```
lhs op      rhs label   est    se      z pvalue ci.lower ci.upper
Y   ~       X      c 0.036 0.116 0.312  0.755
M   ~       X      a 0.474 0.098 4.837  0.00
Y   ~       M      b 0.788 0.094 8.361  0.00
Y  ~~       Y        0.898 0.149 6.044  0.00
M  ~~       M        1.054 0.178 5.917  0.00
X  ~~       X        0.999 0.000    NA     N

ab     :=     a*b    ab 0.374 0.087 4.314  0.000     0.213     0.559
total  := c+(a*b) total 0.410 0.139 2.942  0.003     0.140     0.689
```

Zero is not in C.I.:
Indirect effect
significant

P-value based on bootstrapping

P-value is based on bootstrapped standard errors if you specify `se = “bootstrap”`

Other use of bootstrapping

Bootstrapping is also useful:

- if **sample size is small**, such that normal approximations are not appropriate
- if the data are (multivariate) **non-normally** distributed

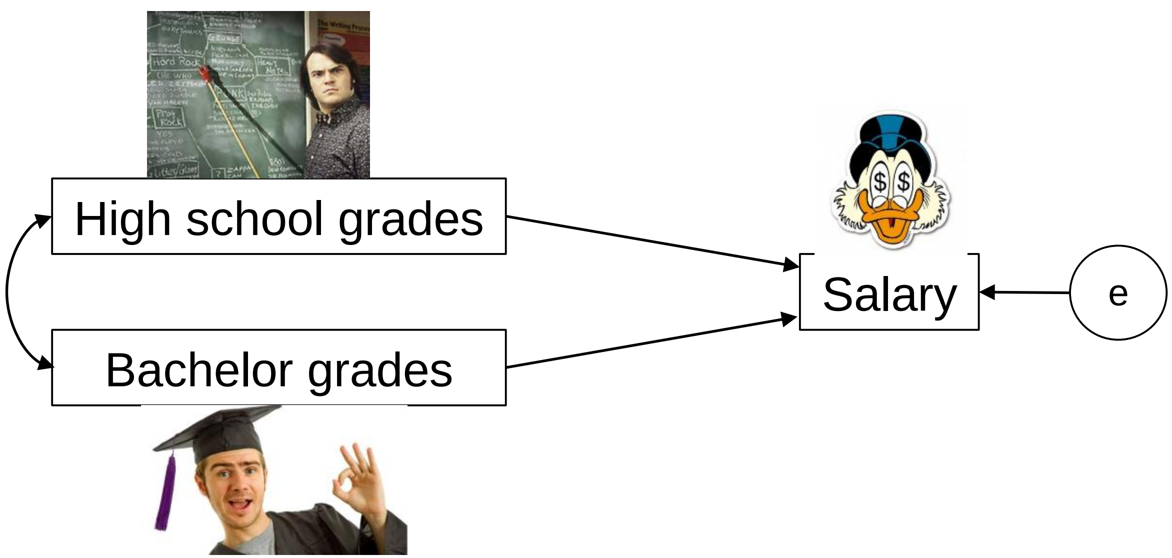
Some people say: Bootstrap everything, all the time.

This allows you to relax the assumption of normality

Different causal models

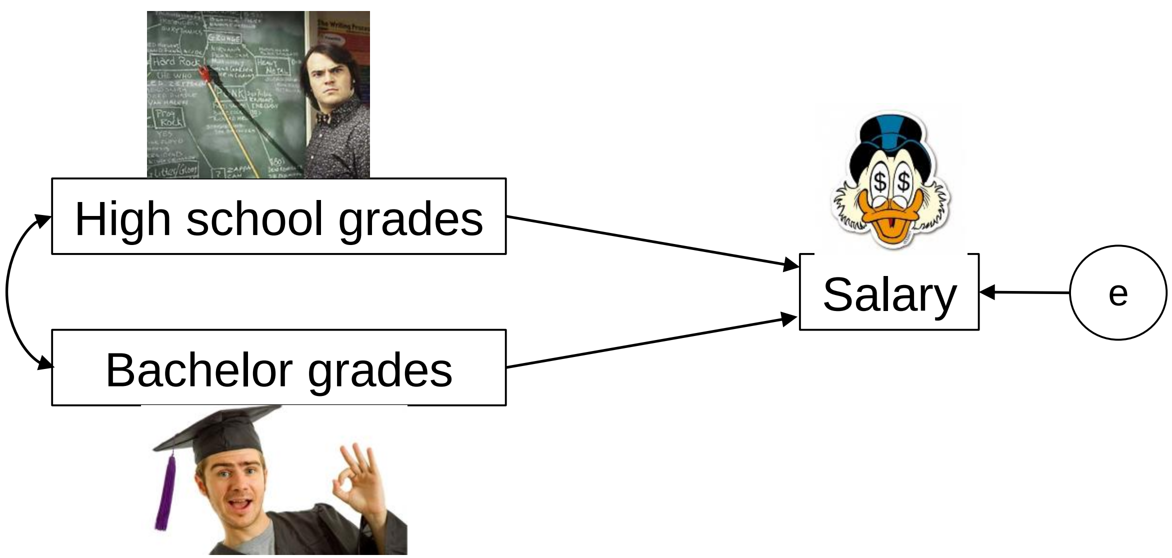
Causality

- You have several correlated variables
- You're imposing a "causal structure" on the variables
- E.g., these two predictors (IVs) are correlated
- You can ask **why** they are correlated.



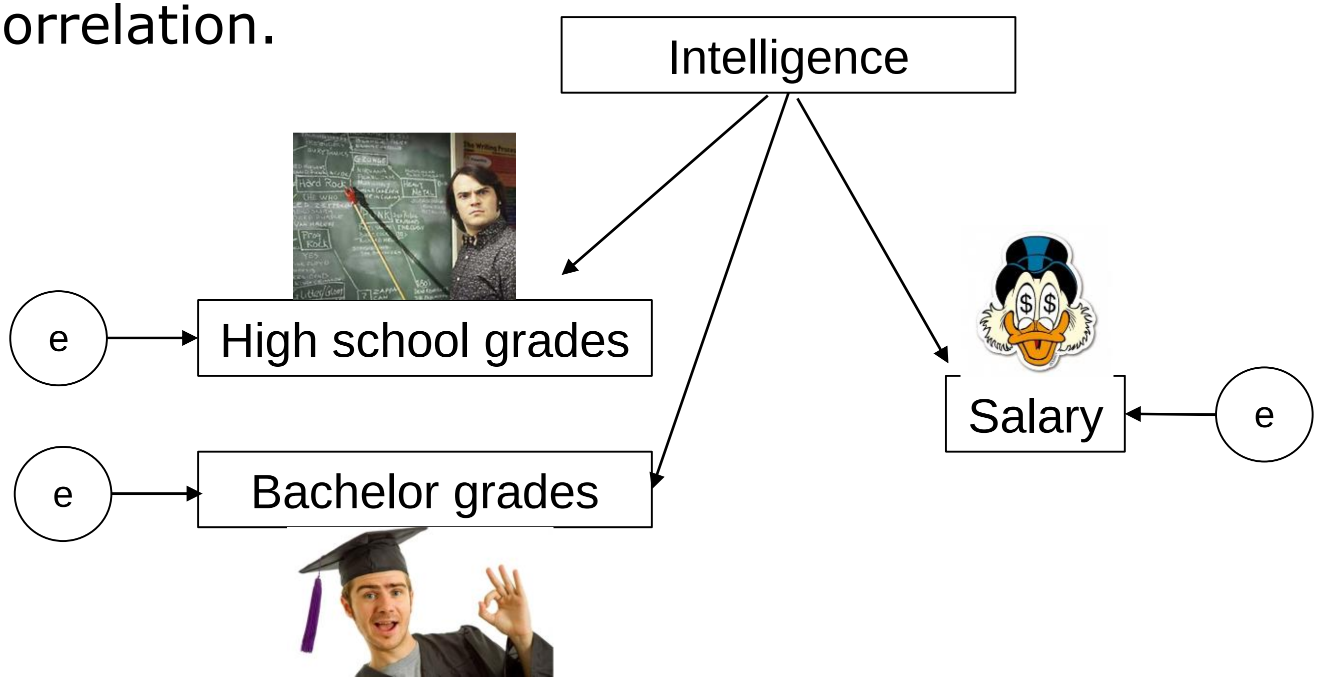
Causality

- These two predictors (IVs) are correlated
- You may have a theory about **why** they are correlated:
  - X_1 may influence X_2 directly (or “reverse causality”)
  - Alternatively: A third variable X_3 is responsible for the correlation.



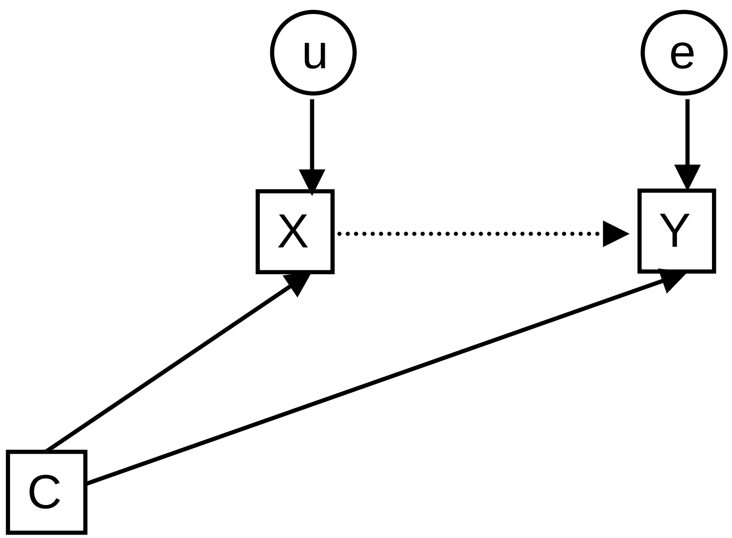
Causality

- These two predictors (IVs) are correlated
- You may have a theory about **why** they are correlated:
  - X_1 may influence X_2 directly (or “reverse causality”)
  - Alternatively: A third variable X_3 is responsible for the correlation.



Confounders

A confounder is a third variable that once it is included, changes the relationship between X and Y.

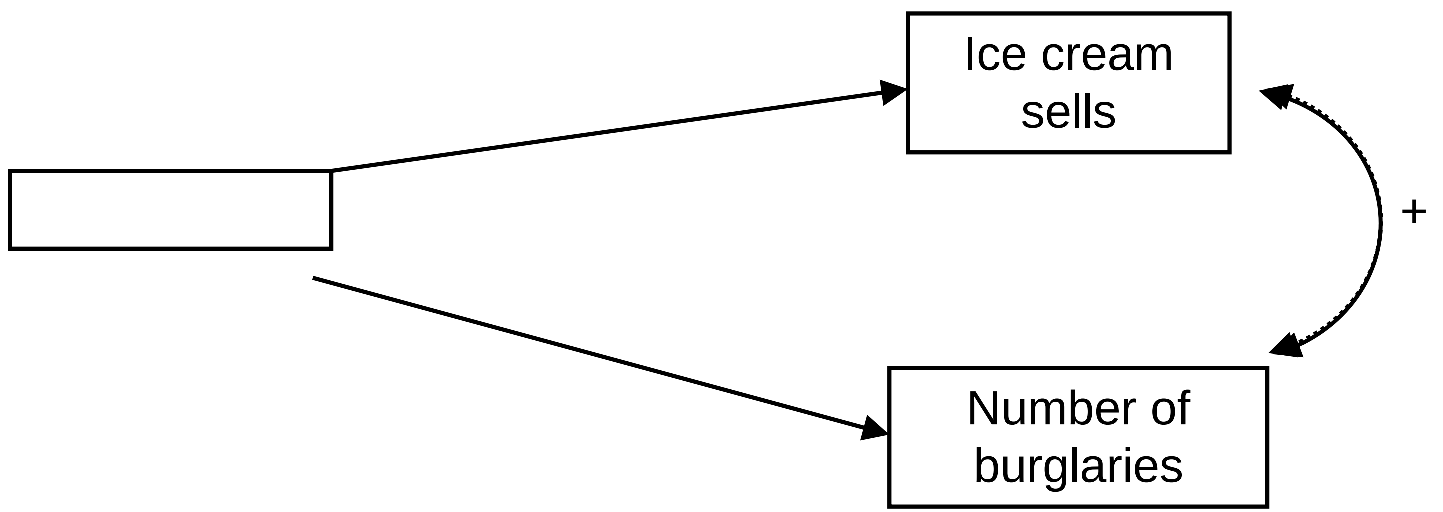


Spurious effects

X is associated with Y, because Z causes X and Y.

The relationship between X and Y is spurious.

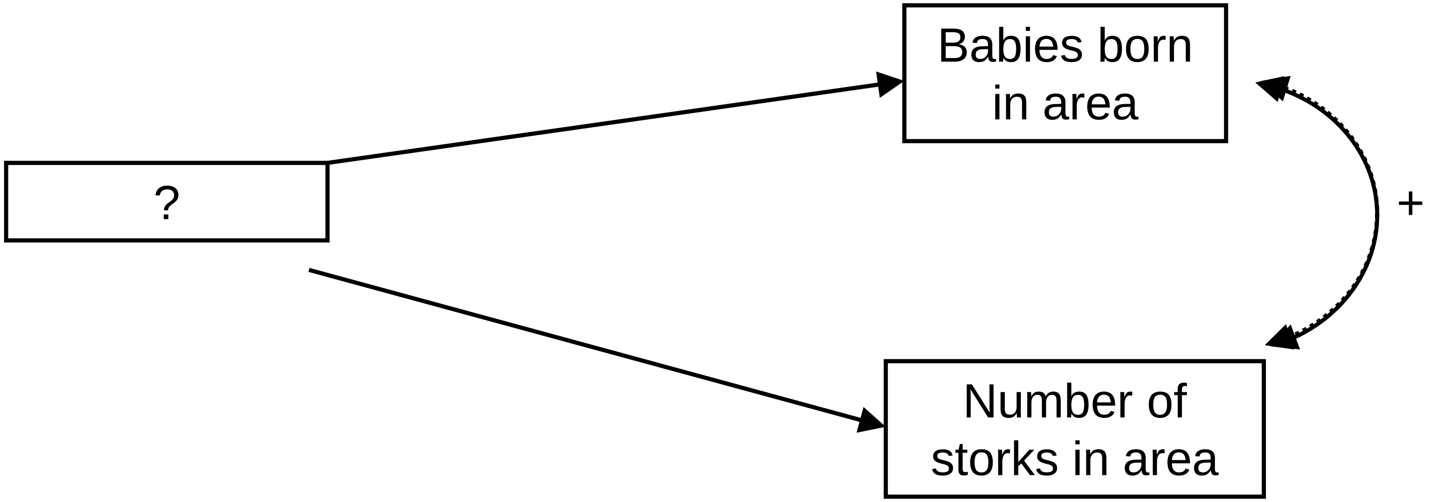
Warning: If you analyze these data with mediation model, you will probably find significant mediation. Why?



Spurious effects?

Number of storks in area → Babies born in area

Spurious effects?



Causality

- If I fit a mediation model
- And the model has good fit
- Can I conclude that the effect of X on Y is indeed explained/mediated by M?

Causality

- If I fit a mediation model
- And the model has good fit
- Can I conclude that the effect of X on Y is indeed explained/mediated by M?

- **NO**! Causality is always in the **METHODS** (or theory), not in the **STATISTICS**
- My model reflects my theory
- Different causal models will have identical fit (if you flip some of the paths around)