DAGs

Tools for reasoning about causes

Einar Holsbø, Department of Computer Science, UiT — October, 2023

The Marko fallacy

The fundamental problem of observational research:

seeing X ≠ doing X

Baby’s first DAG: Marko’s causal model

(D)irected (A)cyclic (G)raph

(D)irected (A)cyclic **(G)raph**

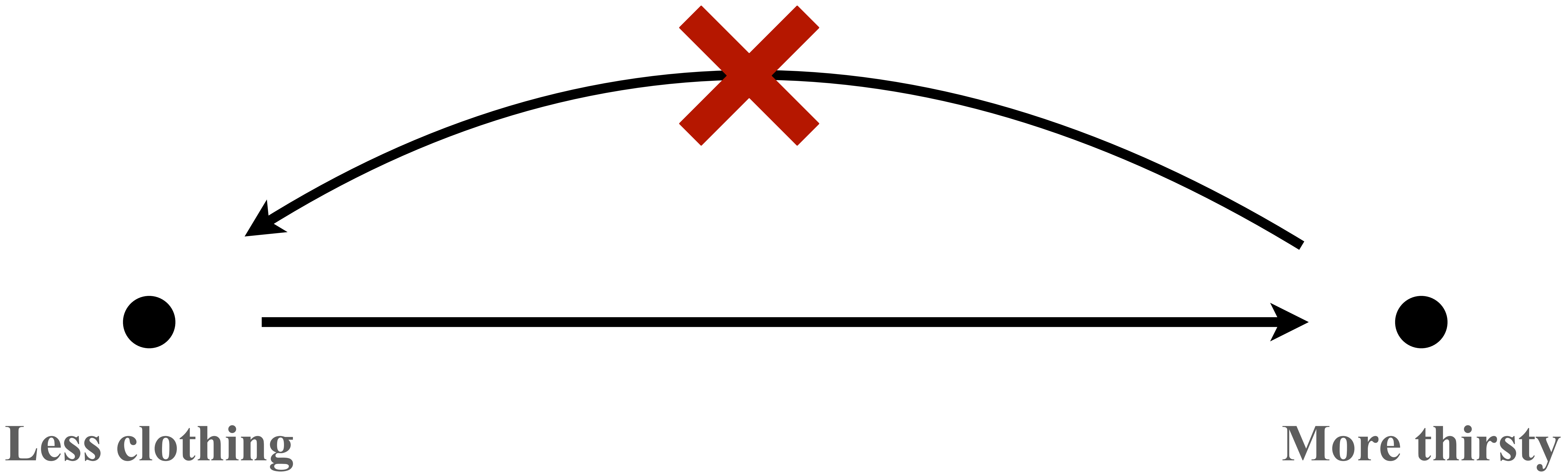
Technical term for some dots joined up with lines

(D)irected (A)cyclic (G)raph

The lines have direction

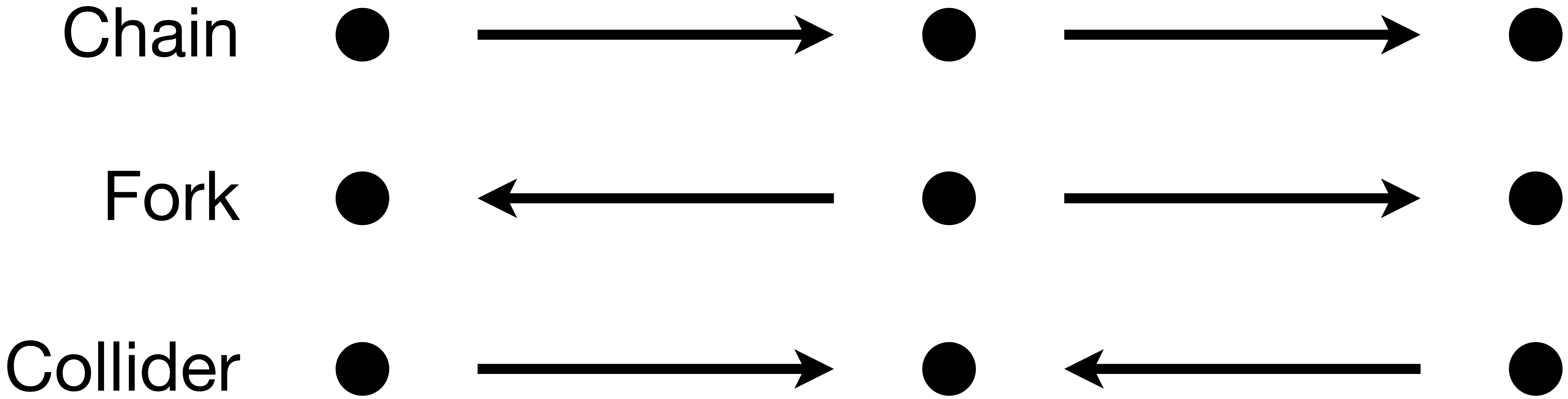
(D)irected **(A)cyclic** (G)raph

No cycles: you can't go backward, a thing cannot be its own cause

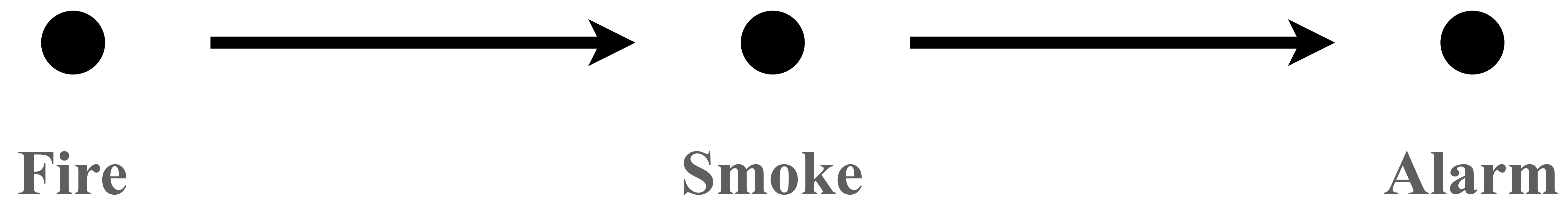


An arrow in the graph is a statement about a thought experiment

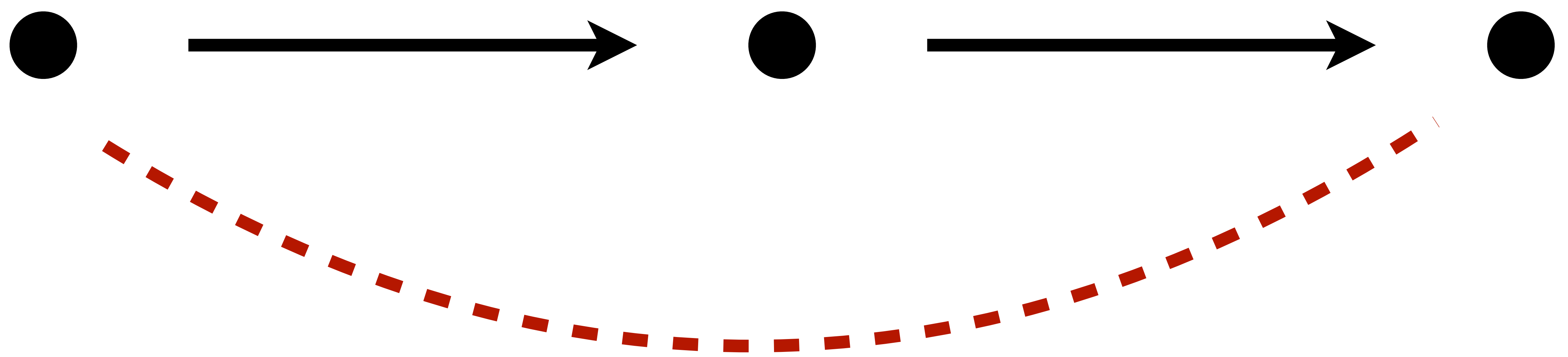
DAG anatomy
Three basic junctions



Chain: A causes B causes C

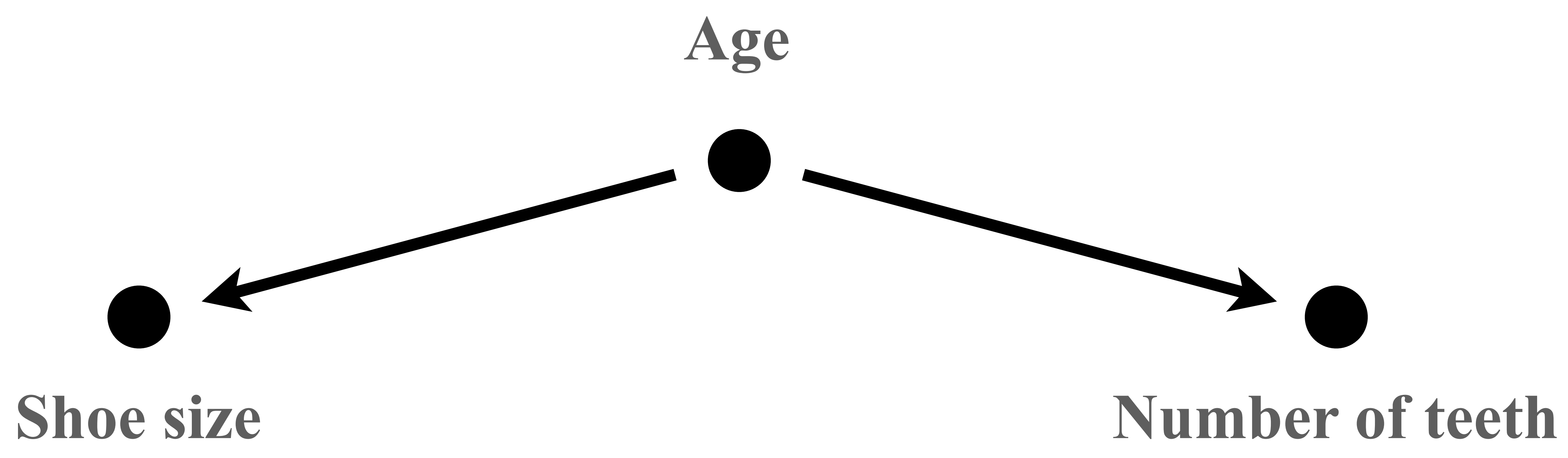


Chain: A causes B causes C

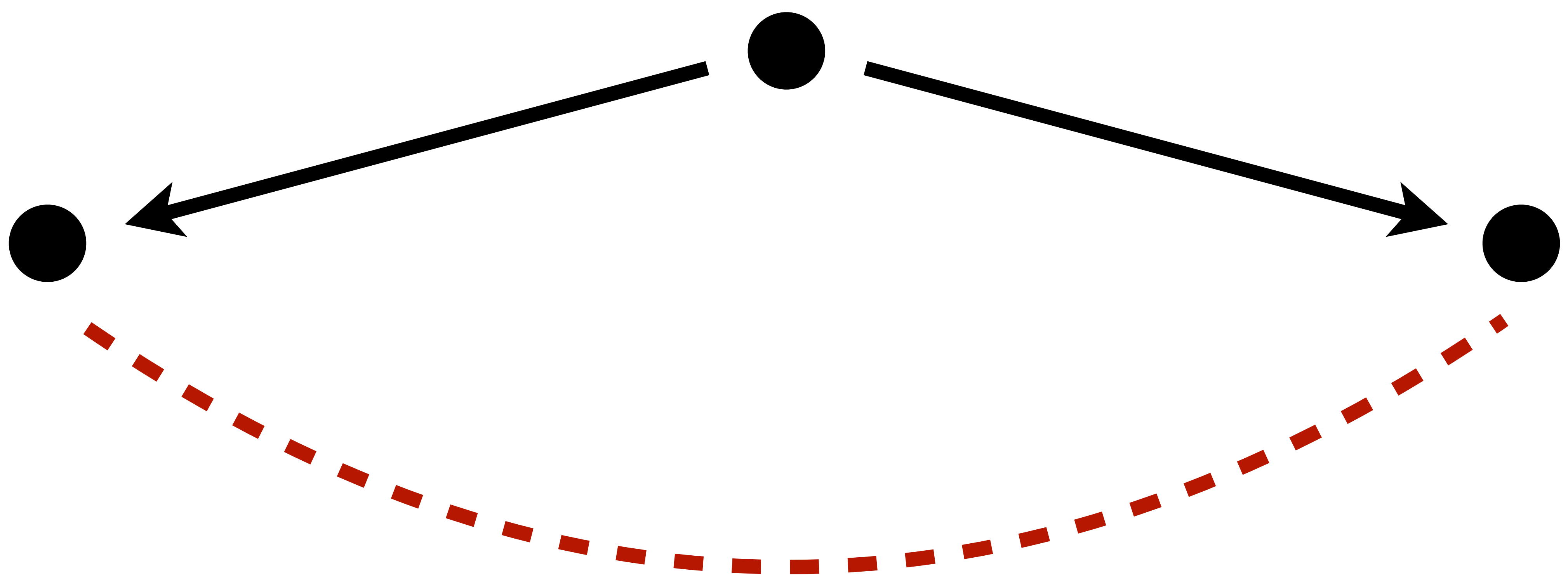


Transmits information:
an alarm indicates fire, fire indicates alarm

Fork: your typical confounder

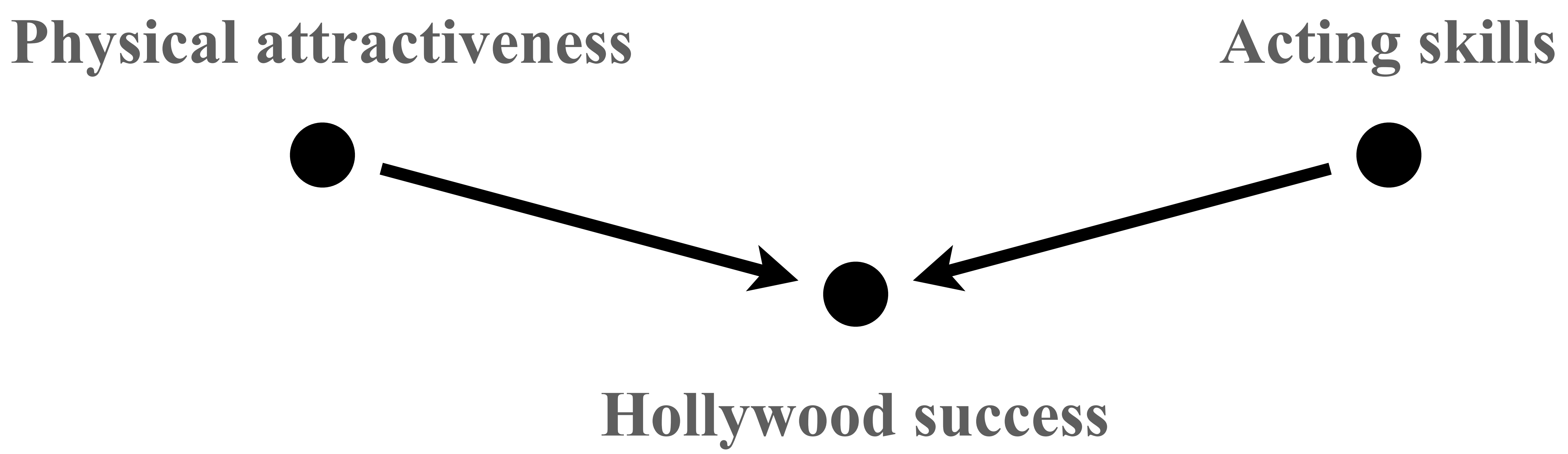


Fork: your typical confounder

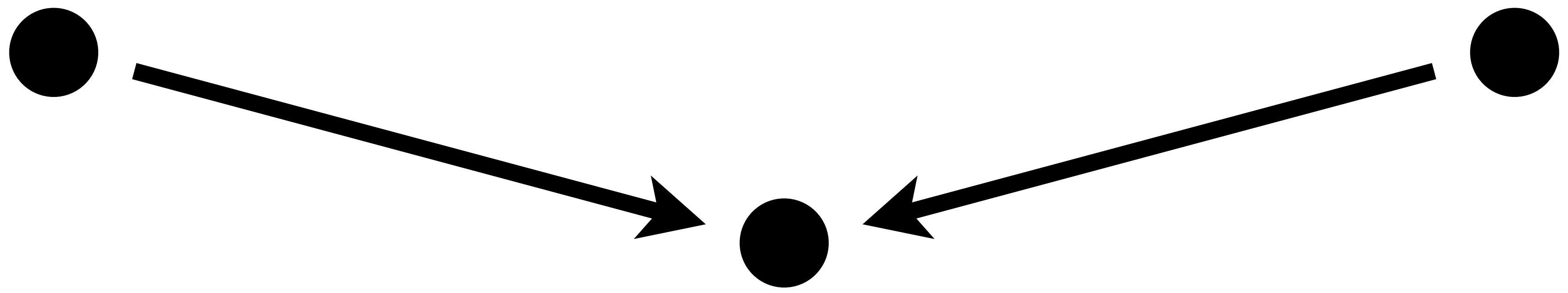


Transmits information:
Few teeth indicates small shoes,
small shoes indicates few teeth

Collider: a very interesting source of bias



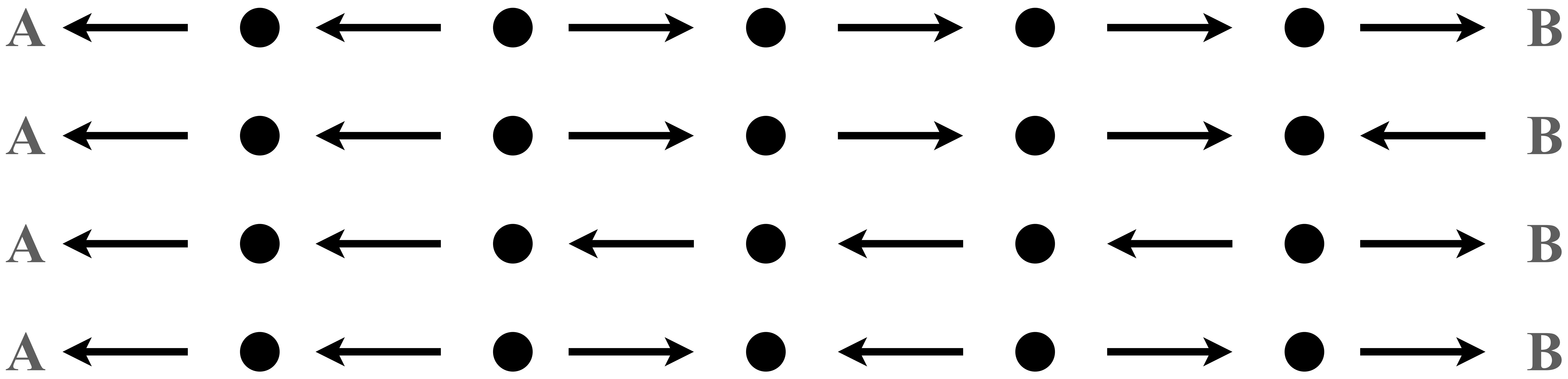
Collider: a very interesting source of bias



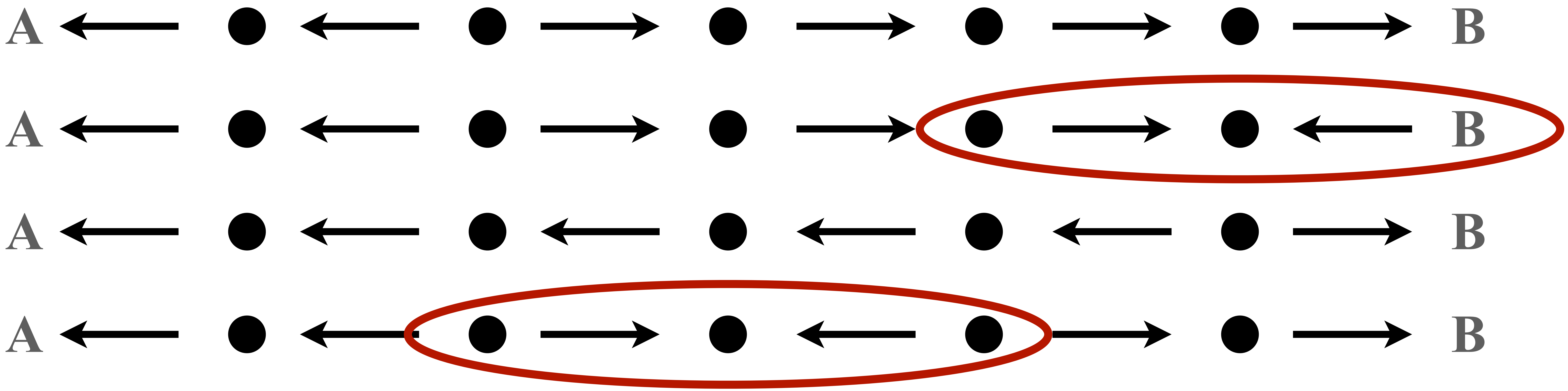
Does <u>not</u> transmit information:

acting skills doesn’t suggest attractiveness or vice versa

Puzzles: does information flow between A and B?



Puzzles: does information flow between A and B?



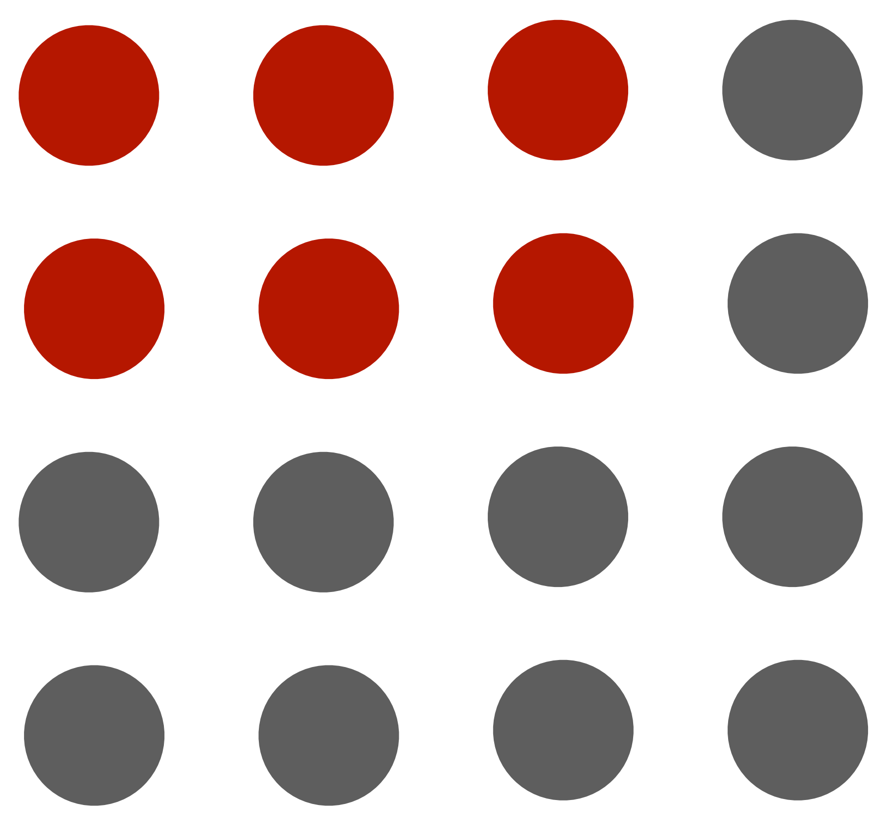
Probability intermission

Probability as counting of outcomes

$$\frac{\#\ \text{interesting outcomes}}{\#\ \text{possible outcomes}}$$

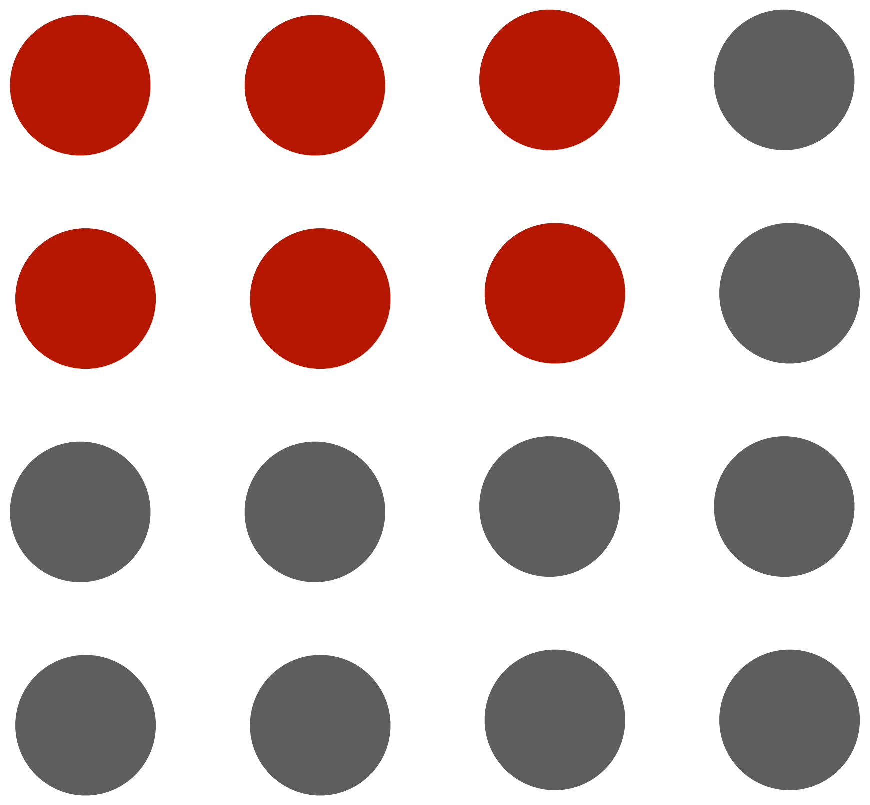
Probability as counting of outcomes

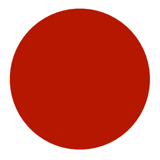
$$\frac{\#\ \text{interesting outcomes}}{\#\ \text{possible outcomes}}$$



Probability as counting of outcomes

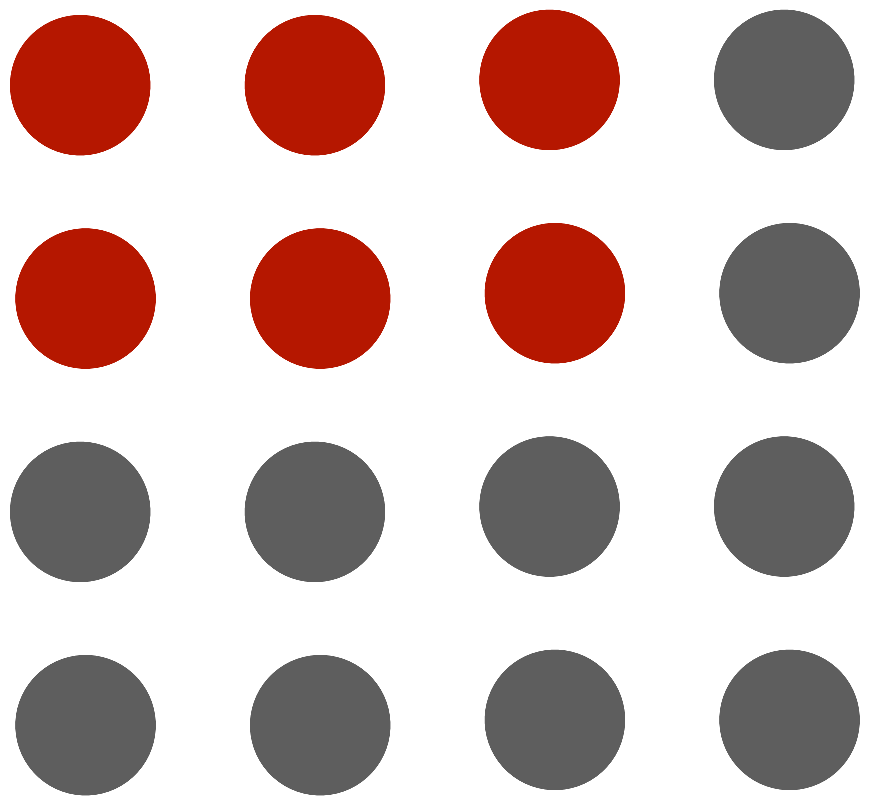
$$\frac{\#\ \text{interesting outcomes}}{\#\ \text{possible outcomes}}$$

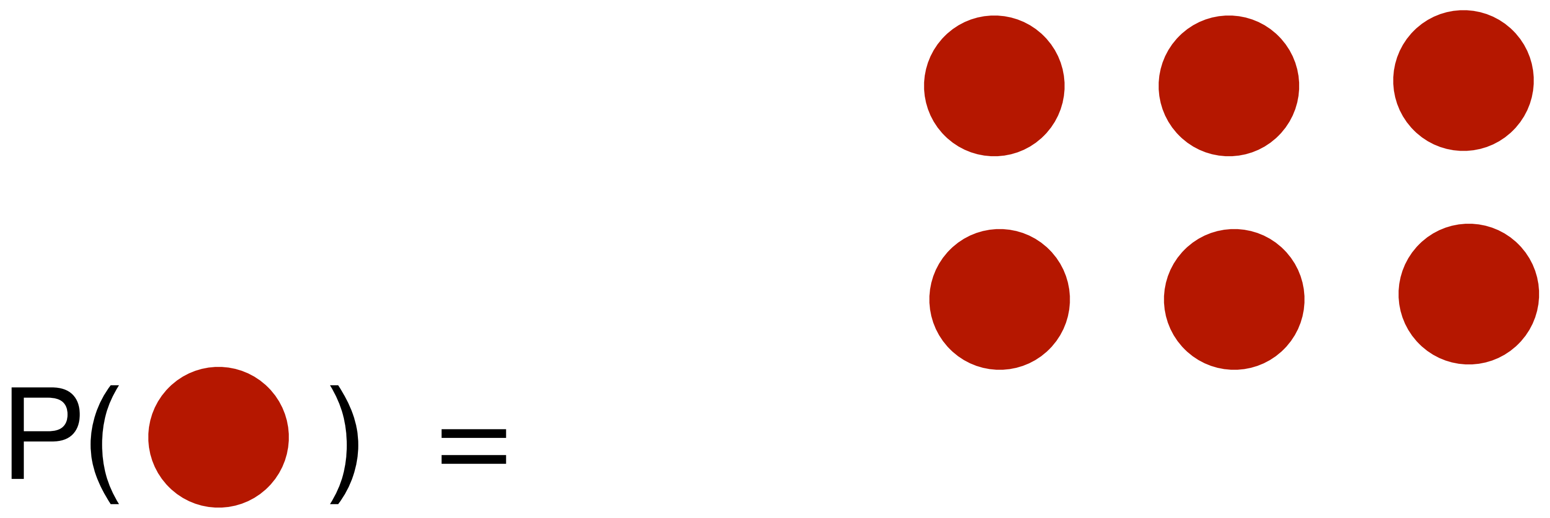


P() =

Probability as counting of outcomes

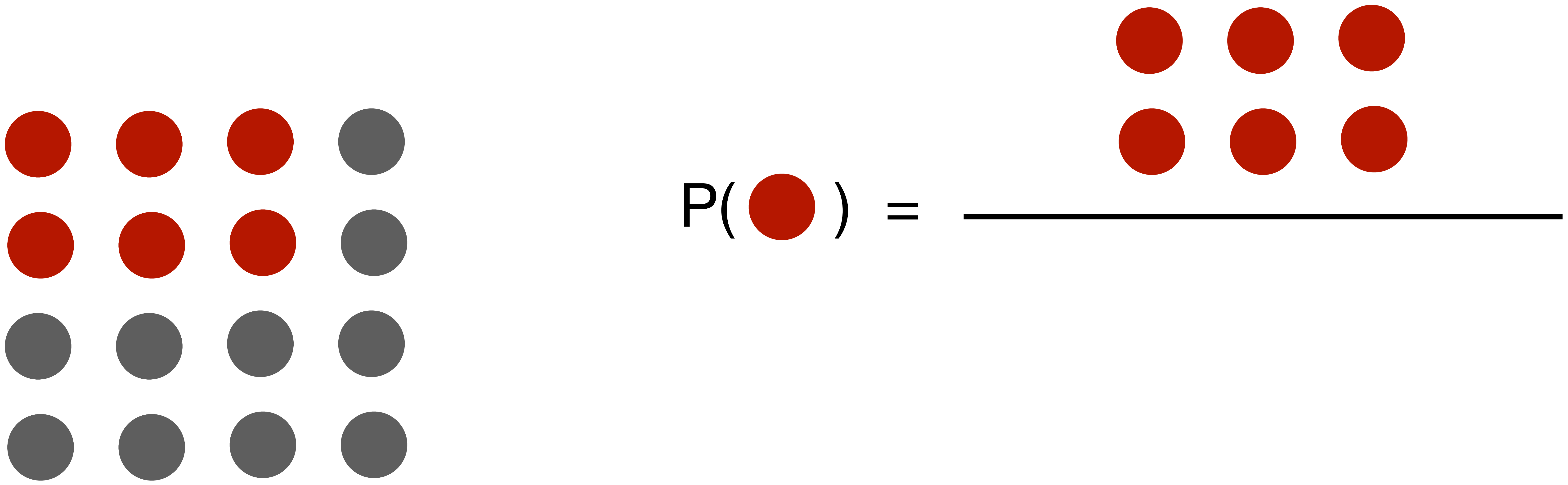
$$\frac{\#\ \text{interesting outcomes}}{\#\ \text{possible outcomes}}$$





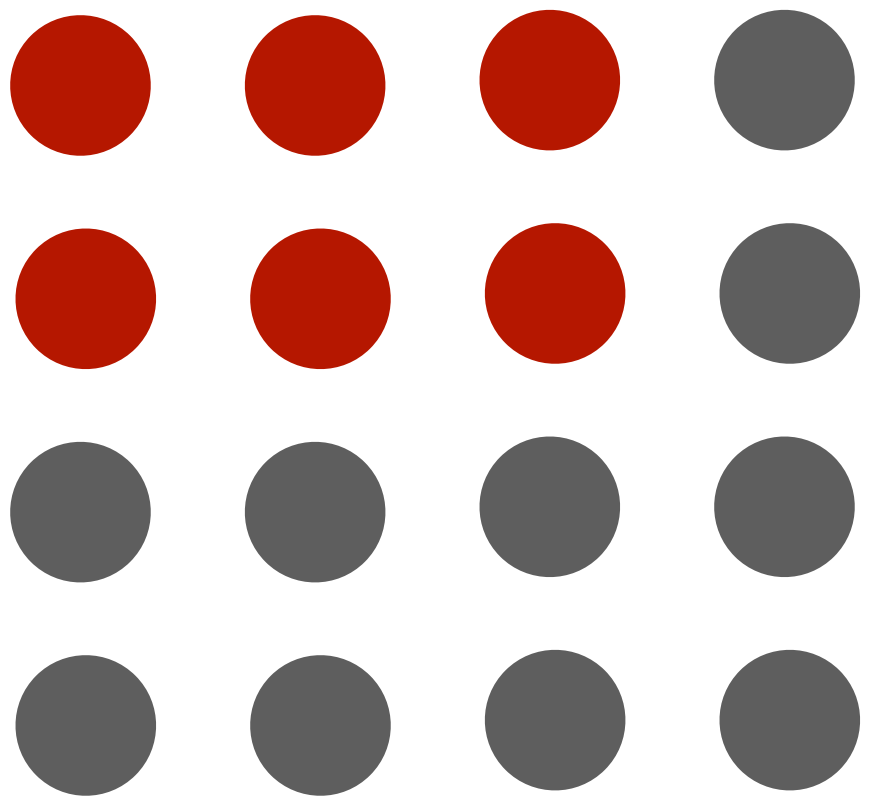
Probability as counting of outcomes

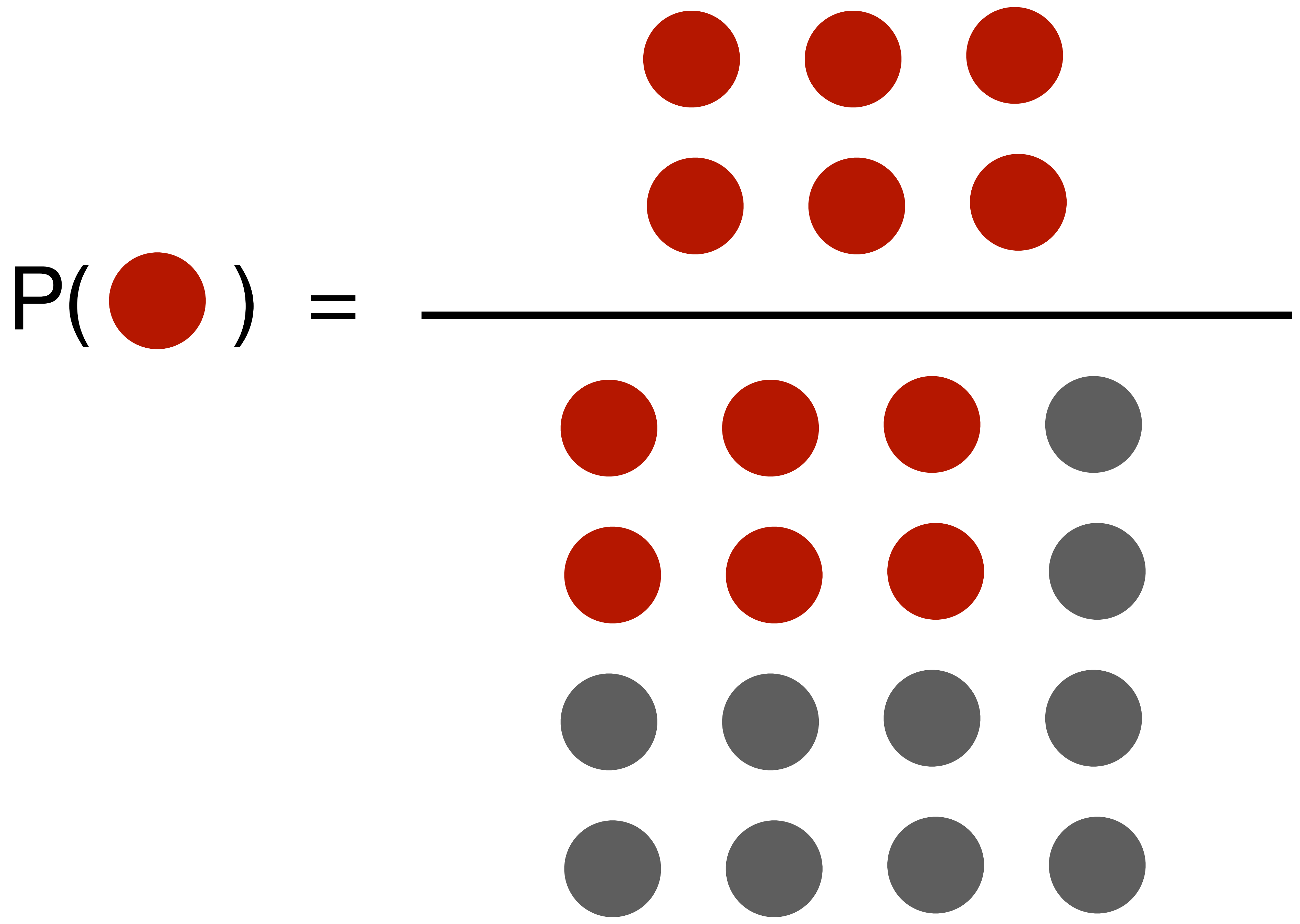
$$\frac{\#\ \text{interesting outcomes}}{\#\ \text{possible outcomes}}$$



Probability as counting of outcomes

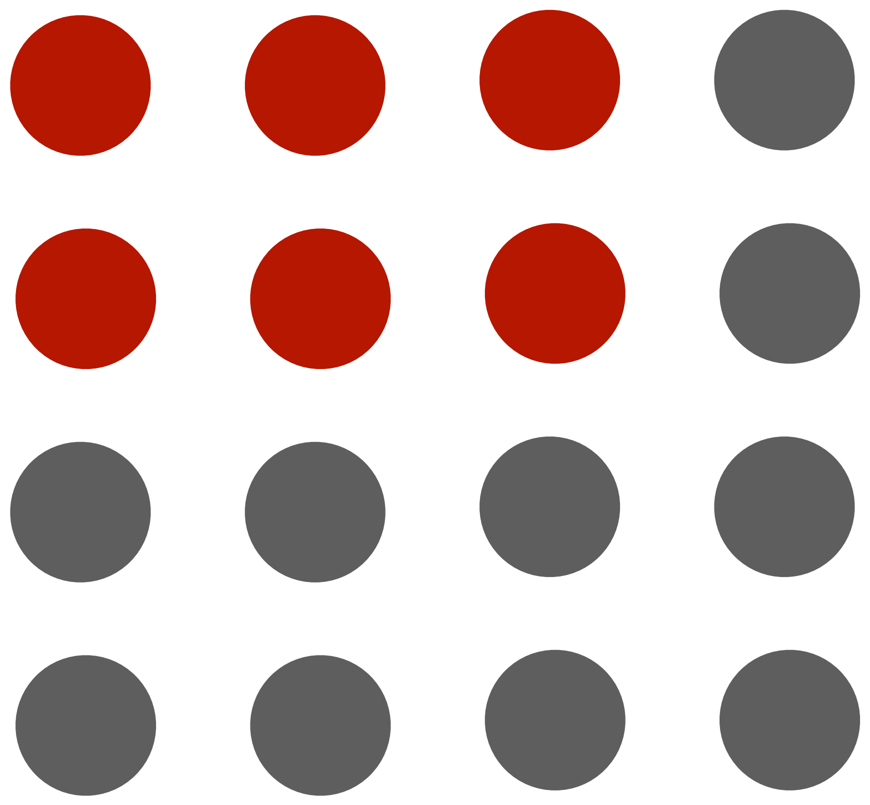
$$\frac{\#\ \text{interesting outcomes}}{\#\ \text{possible outcomes}}$$

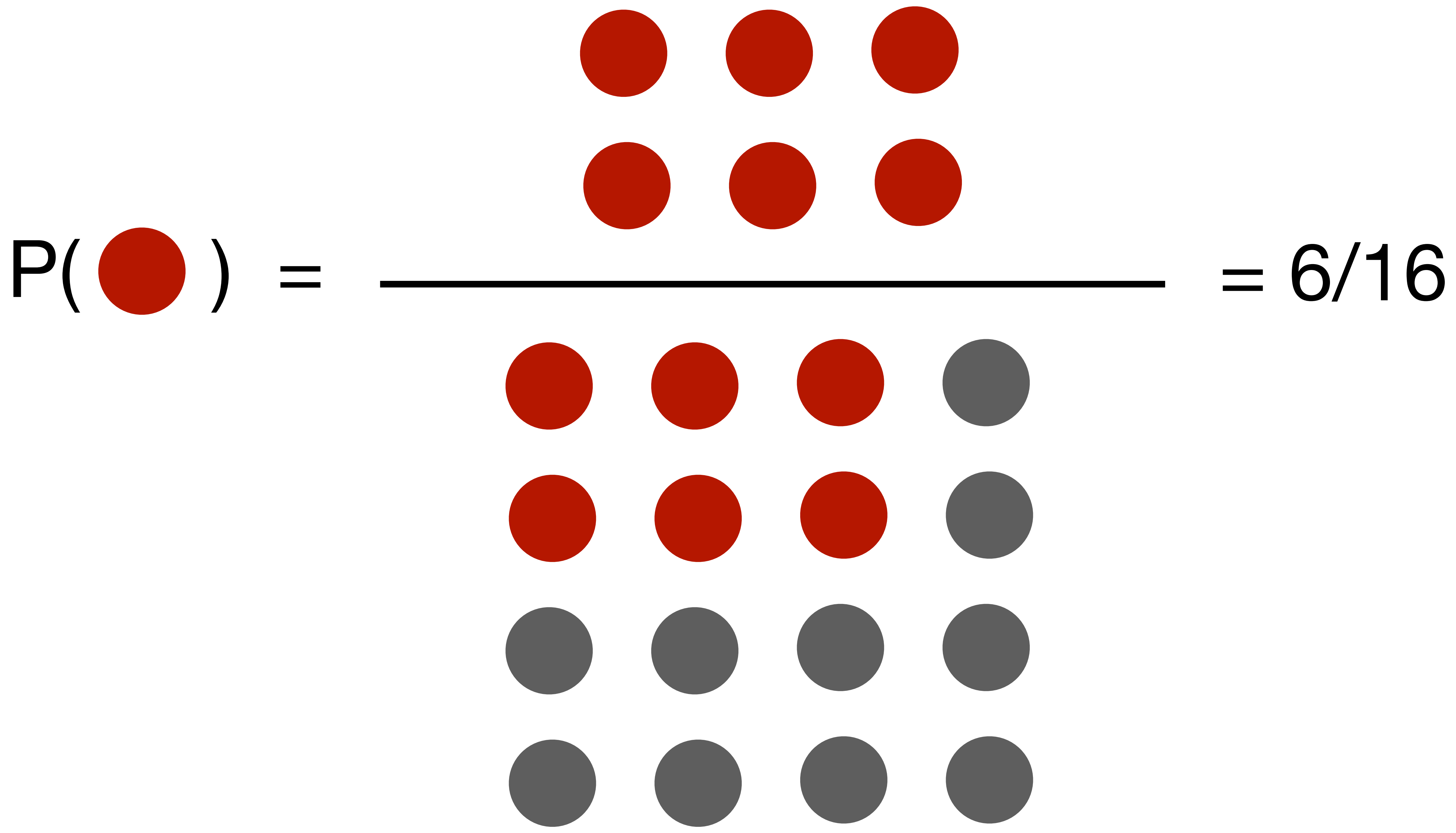




Probability as counting of outcomes

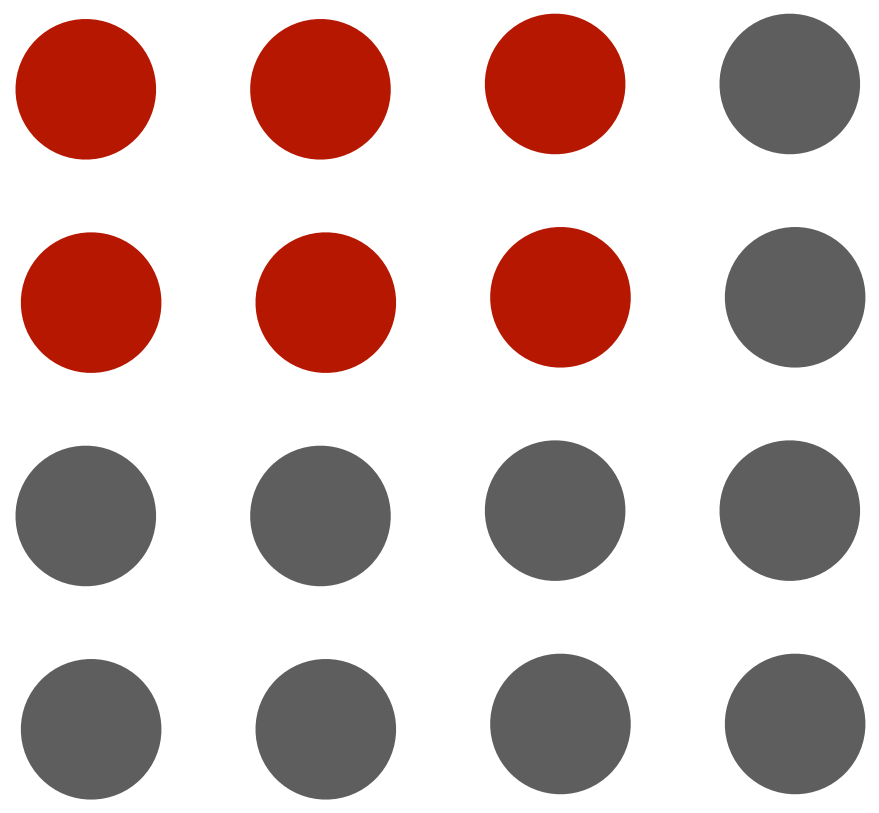
$$\frac{\#\ \text{interesting outcomes}}{\#\ \text{possible outcomes}}$$





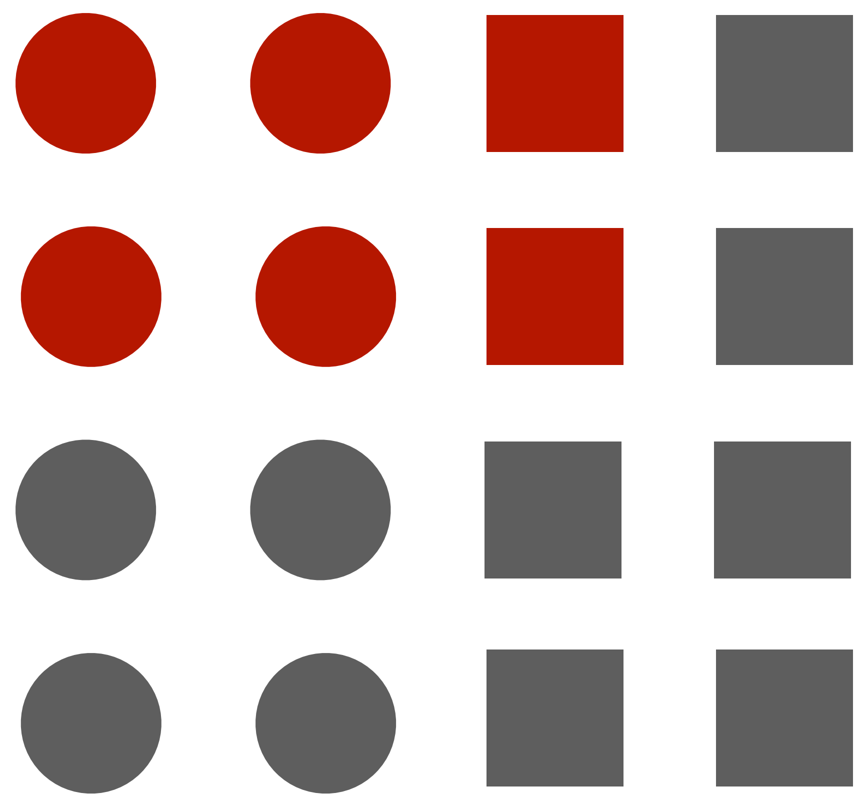
Conditional probability

Counting outcomes in a specific subset



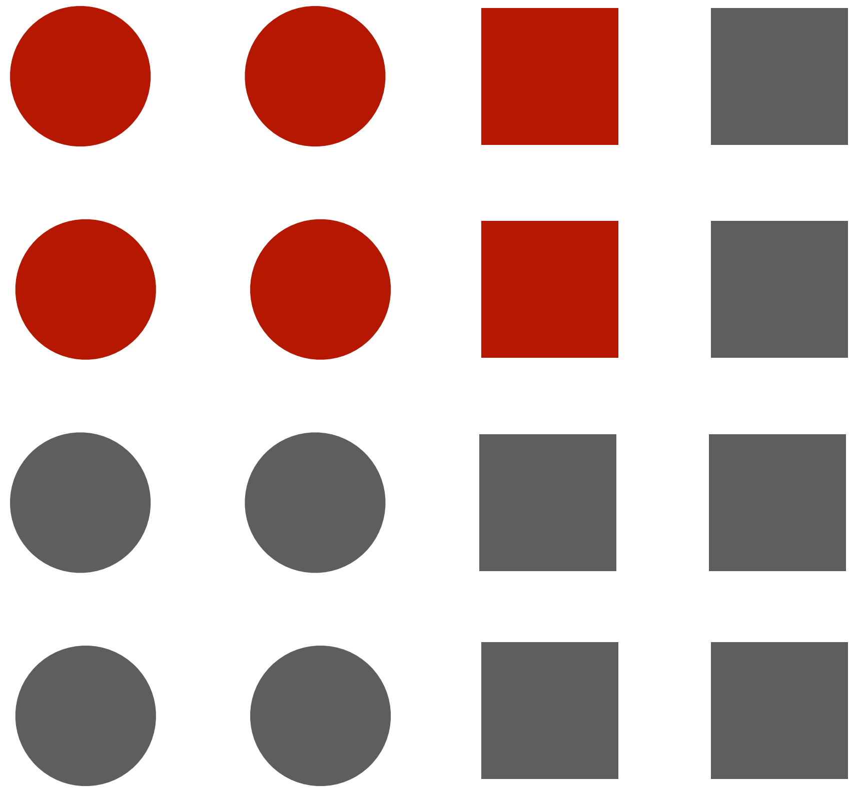
Conditional probability

Counting outcomes in a specific subset



Conditional probability

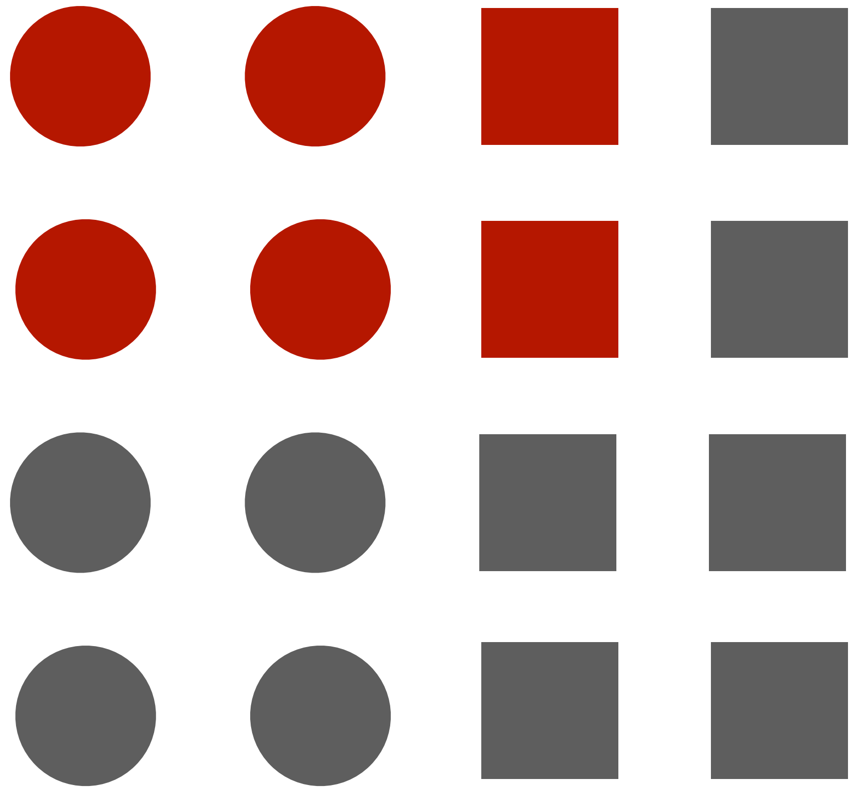
Counting outcomes in a specific subset



P(

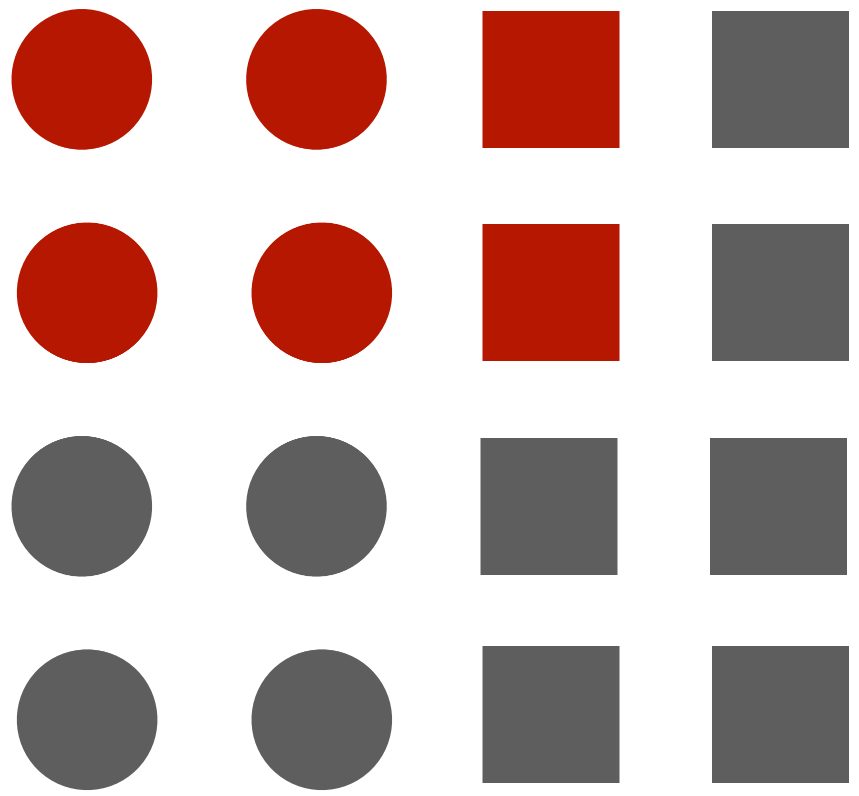
Conditional probability

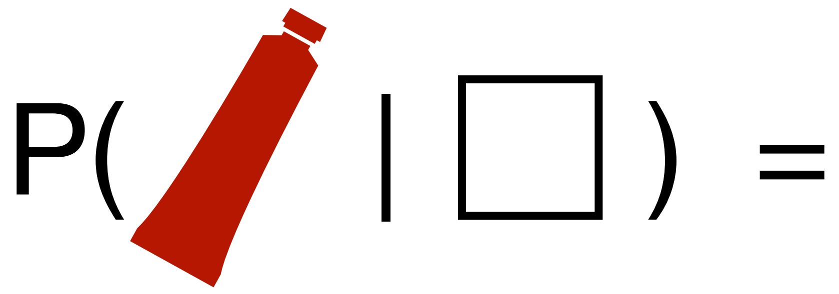
Counting outcomes in a specific subset



Conditional probability

Counting outcomes in a specific subset

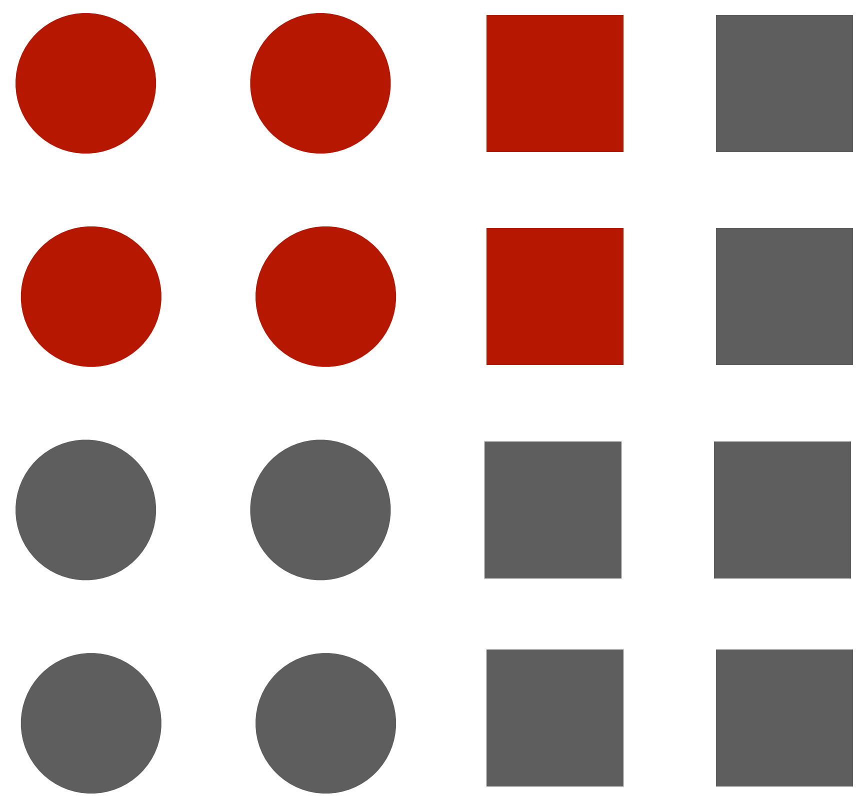


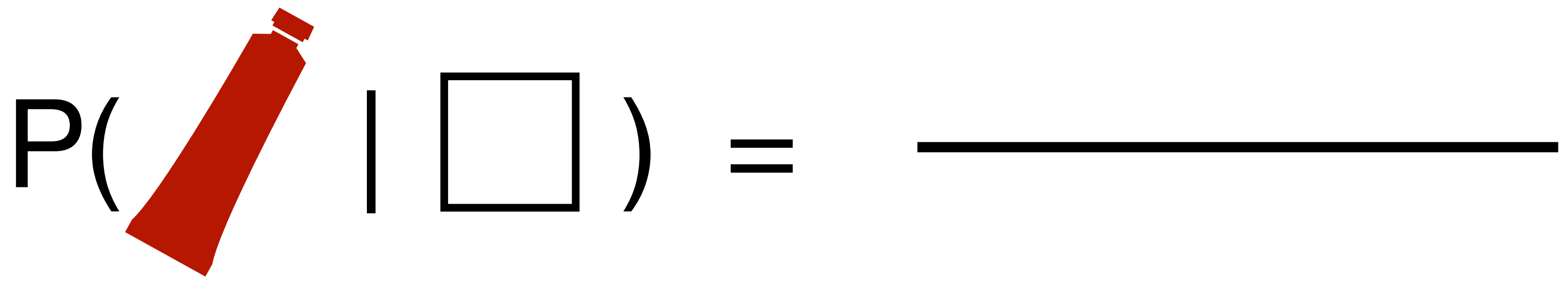


"Probability of red color given square shape" or "probability of red color conditional on square shape"

Conditional probability

Counting outcomes in a specific subset

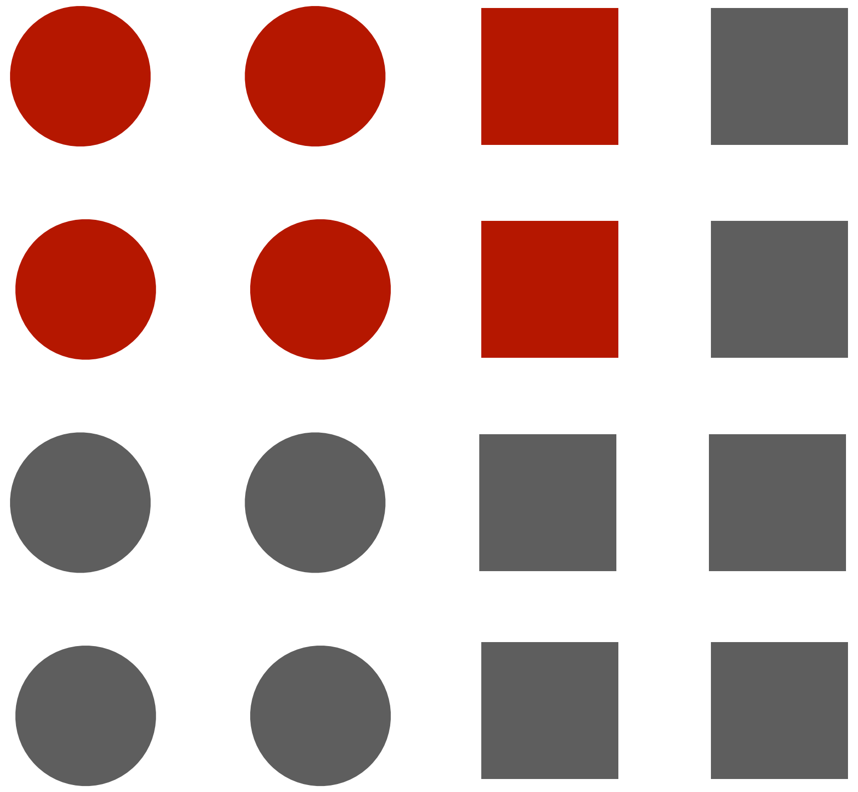


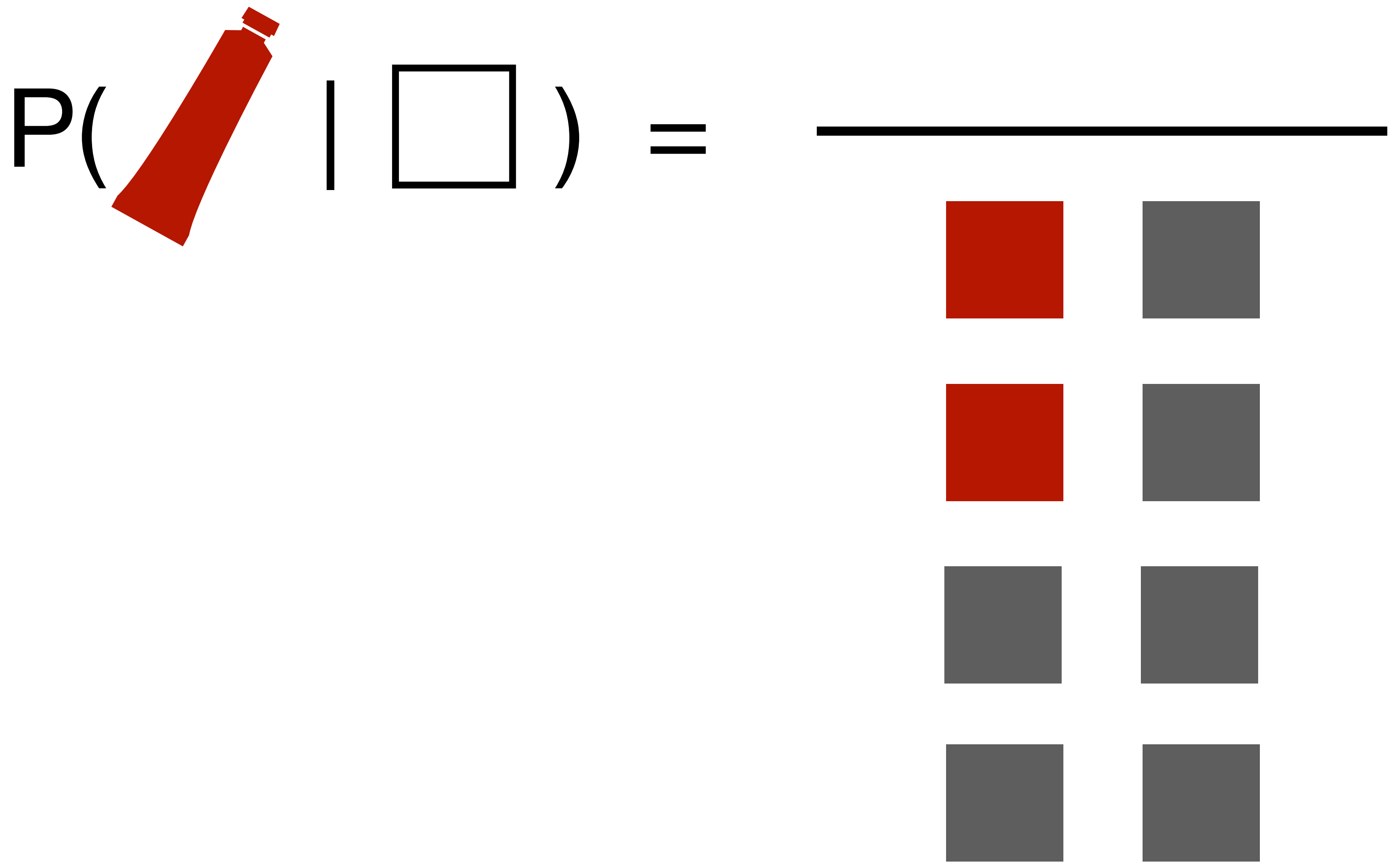


“Probability of red color given square shape” or “probability of red color conditional on square shape”

Conditional probability

Counting outcomes in a specific subset

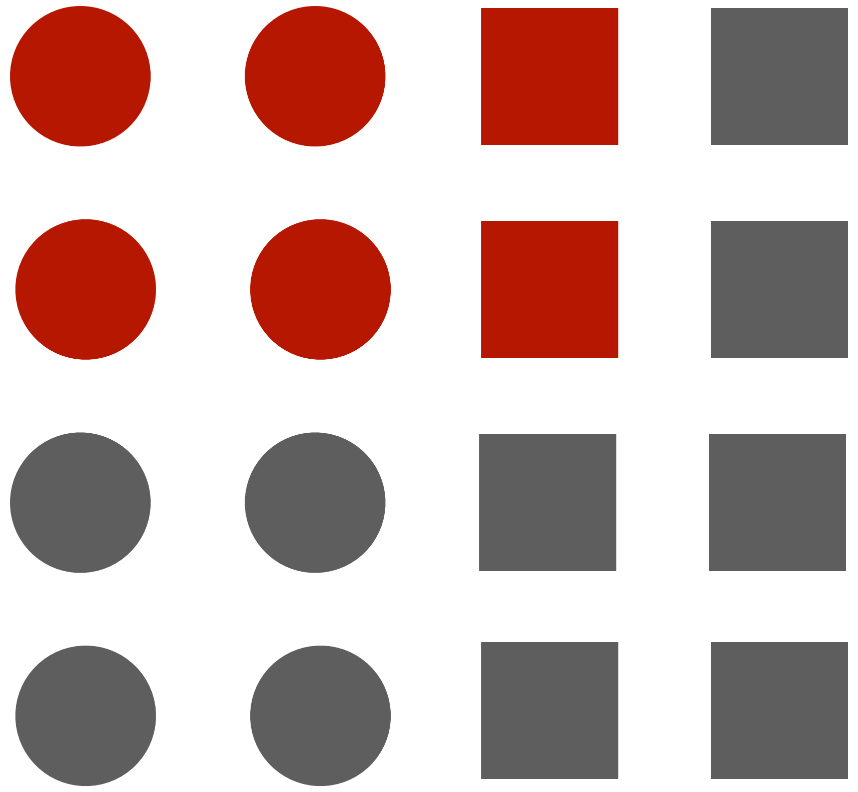


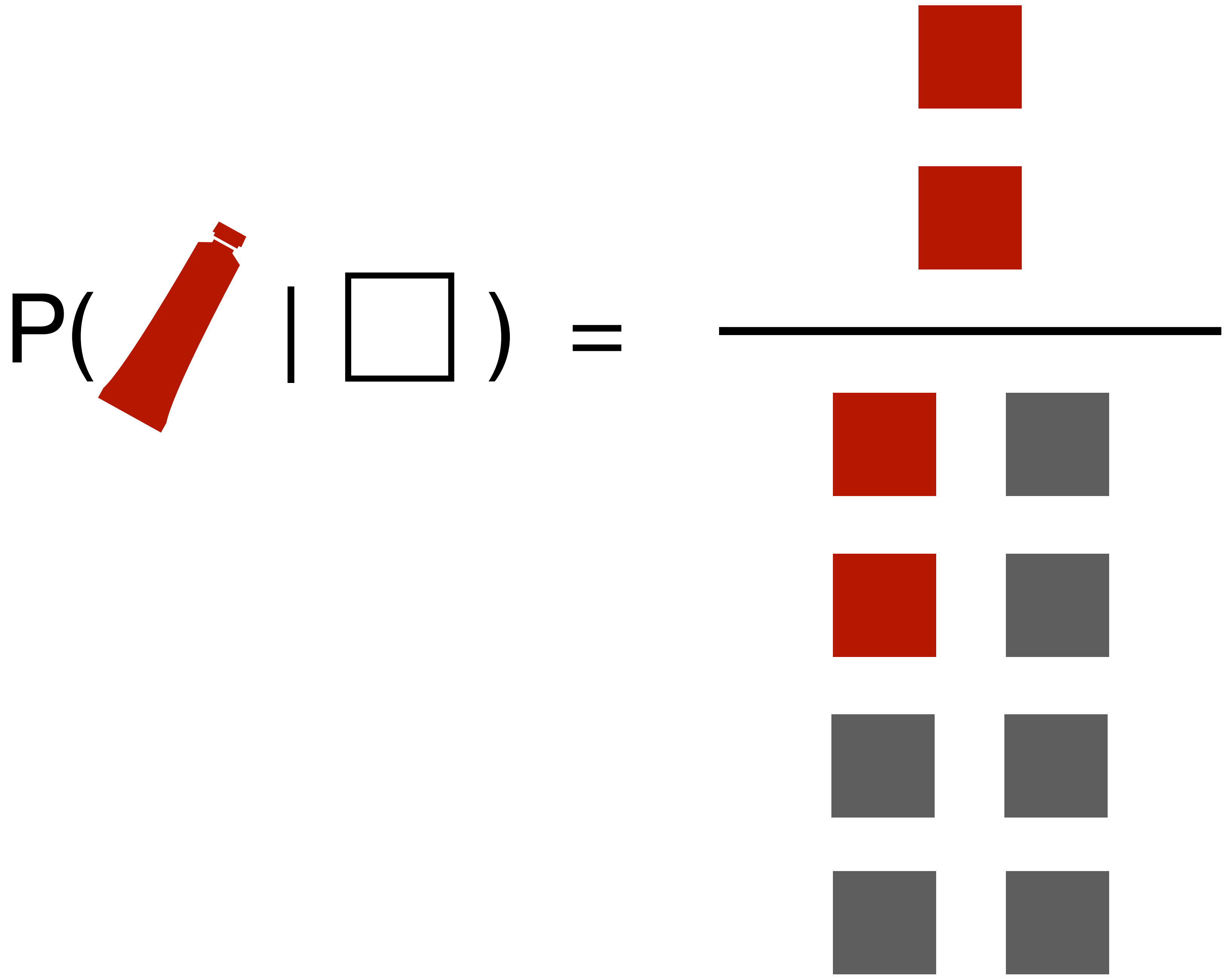


“Probability of red color given square shape” or “probability of red color conditional on square shape”

Conditional probability

Counting outcomes in a specific subset

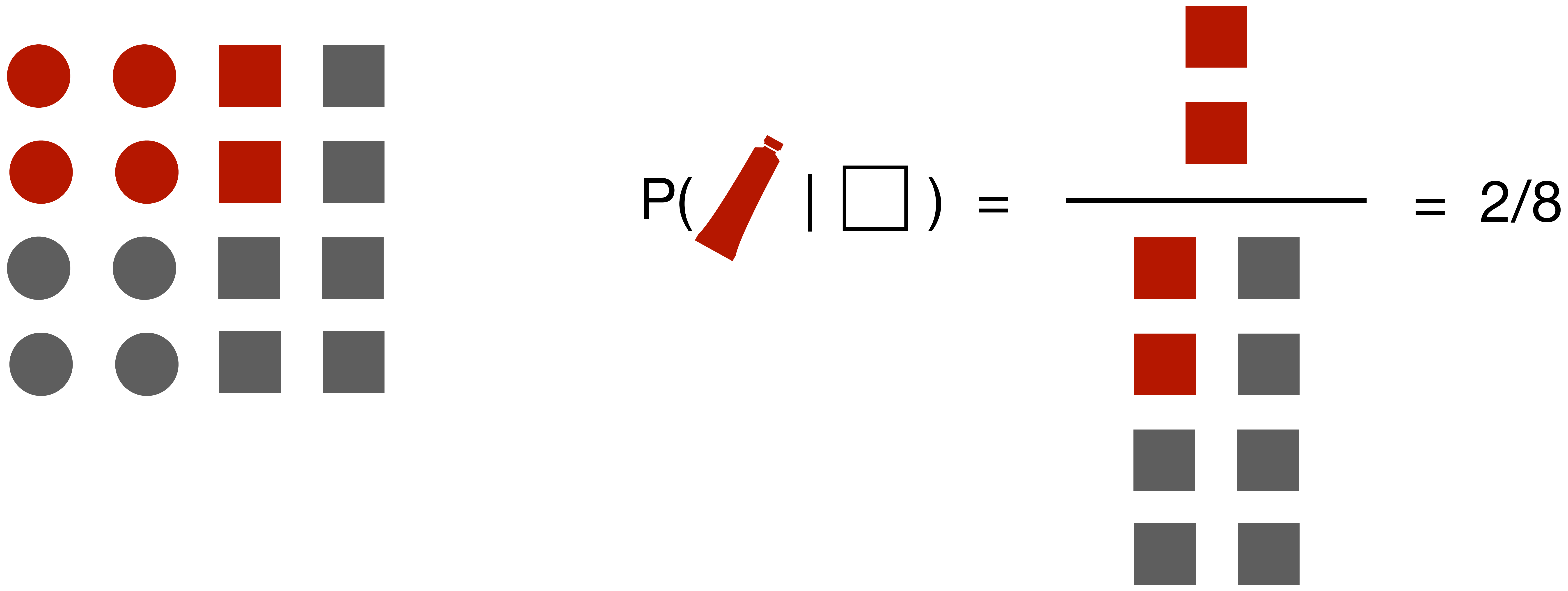




“Probability of red color given square shape” or “probability of red color conditional on square shape”

Conditional probability

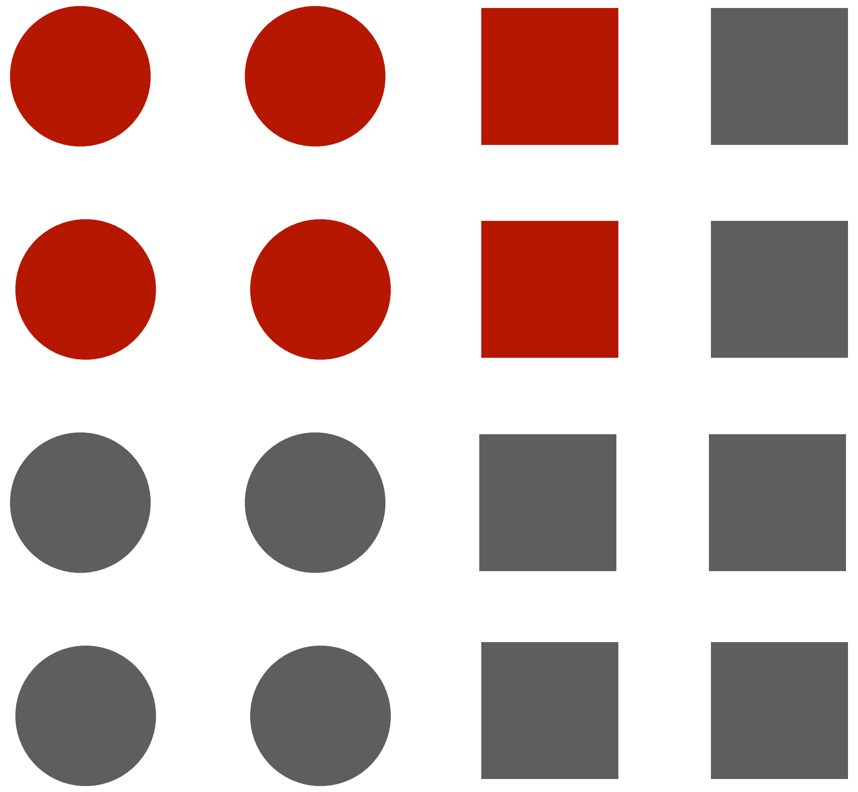
Counting outcomes in a specific subset

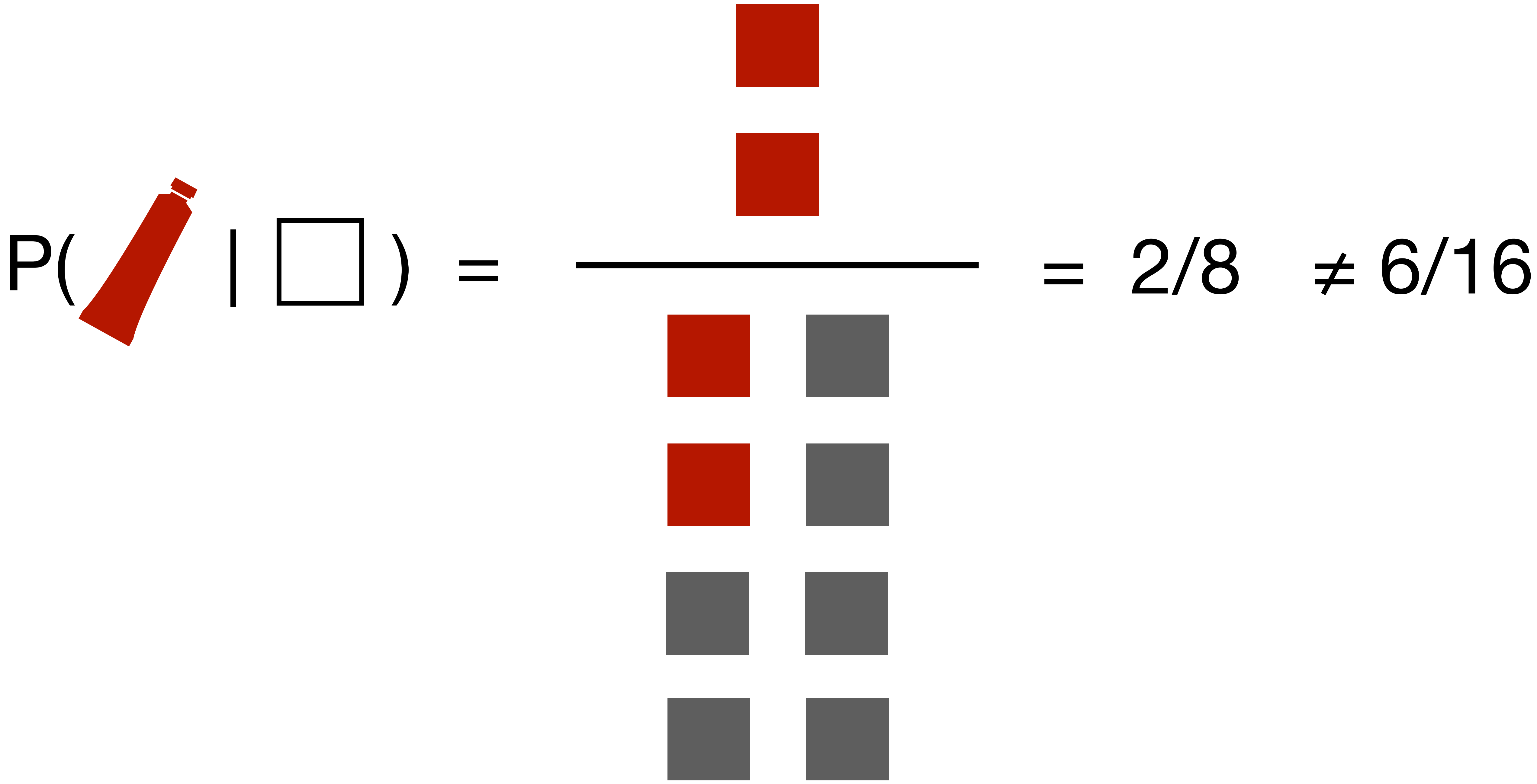


“Probability of red color given square shape” or “probability of red color conditional on square shape”

Conditional probability

Counting outcomes in a specific subset

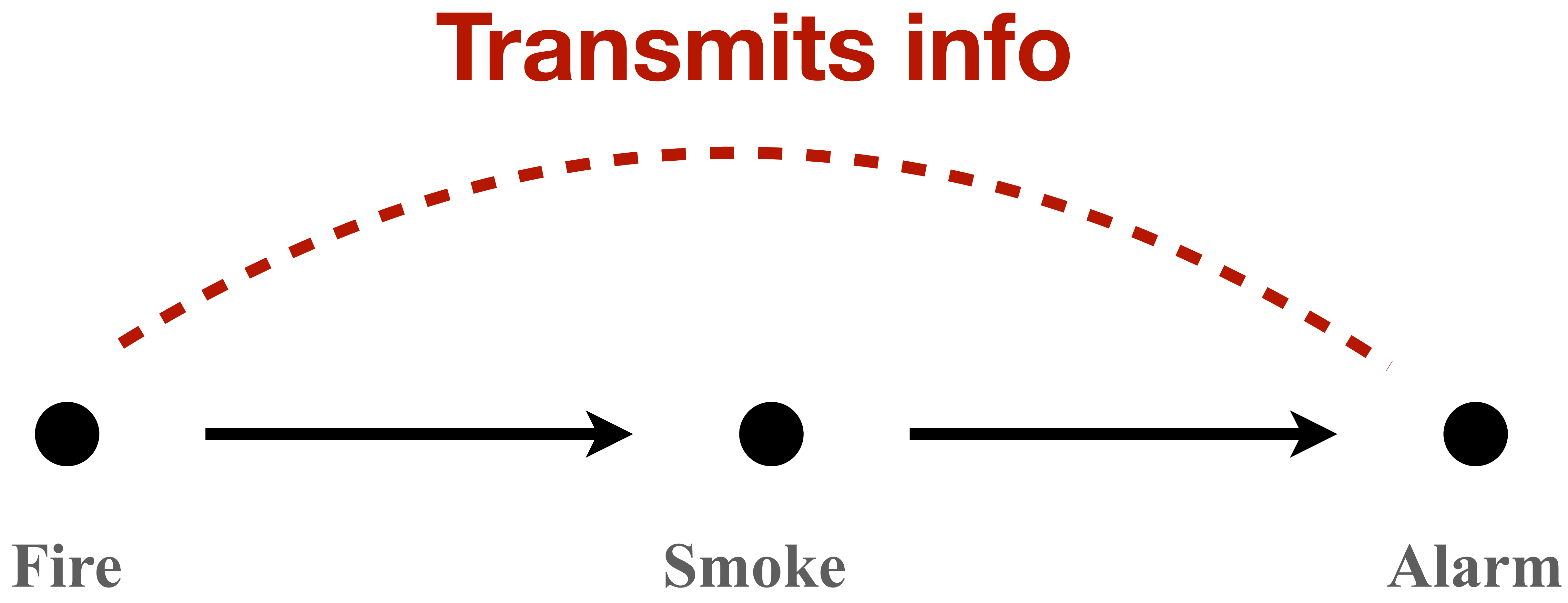




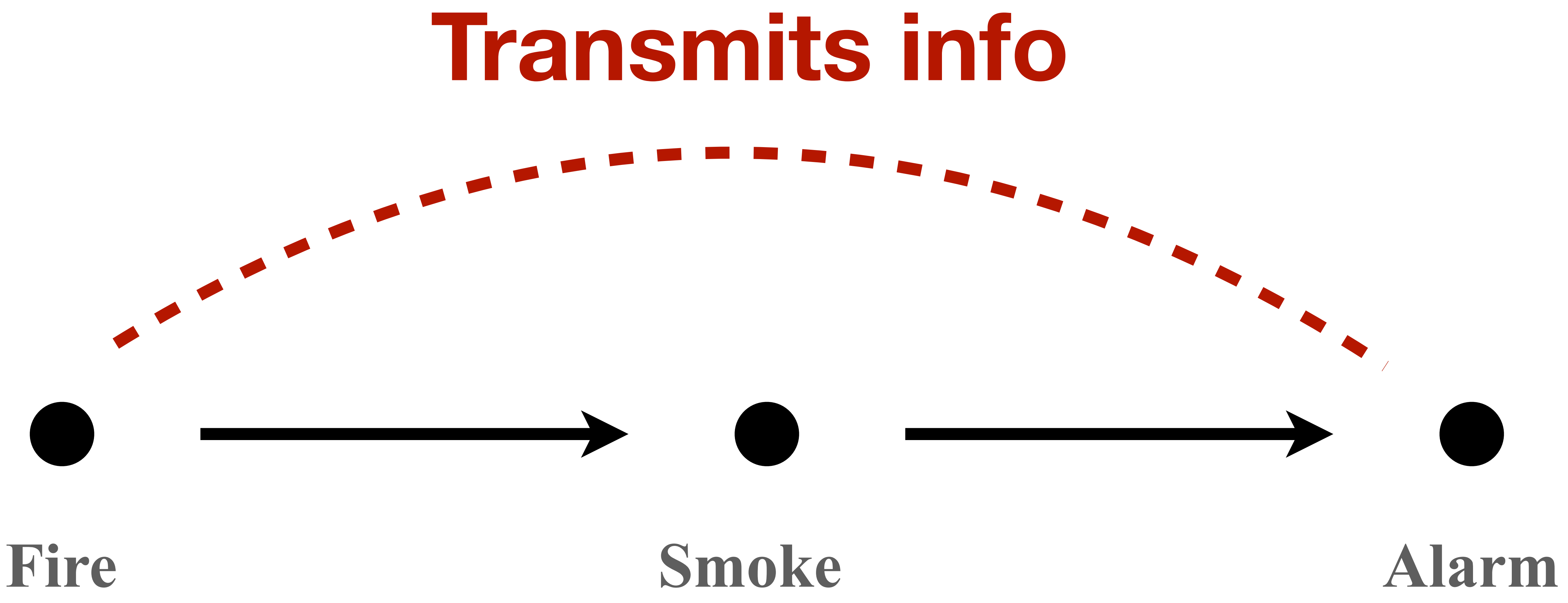
“Probability of red color given square shape” or “probability of red color conditional on square shape”

Conditioning alters the flow of information in the 3 basic junctions

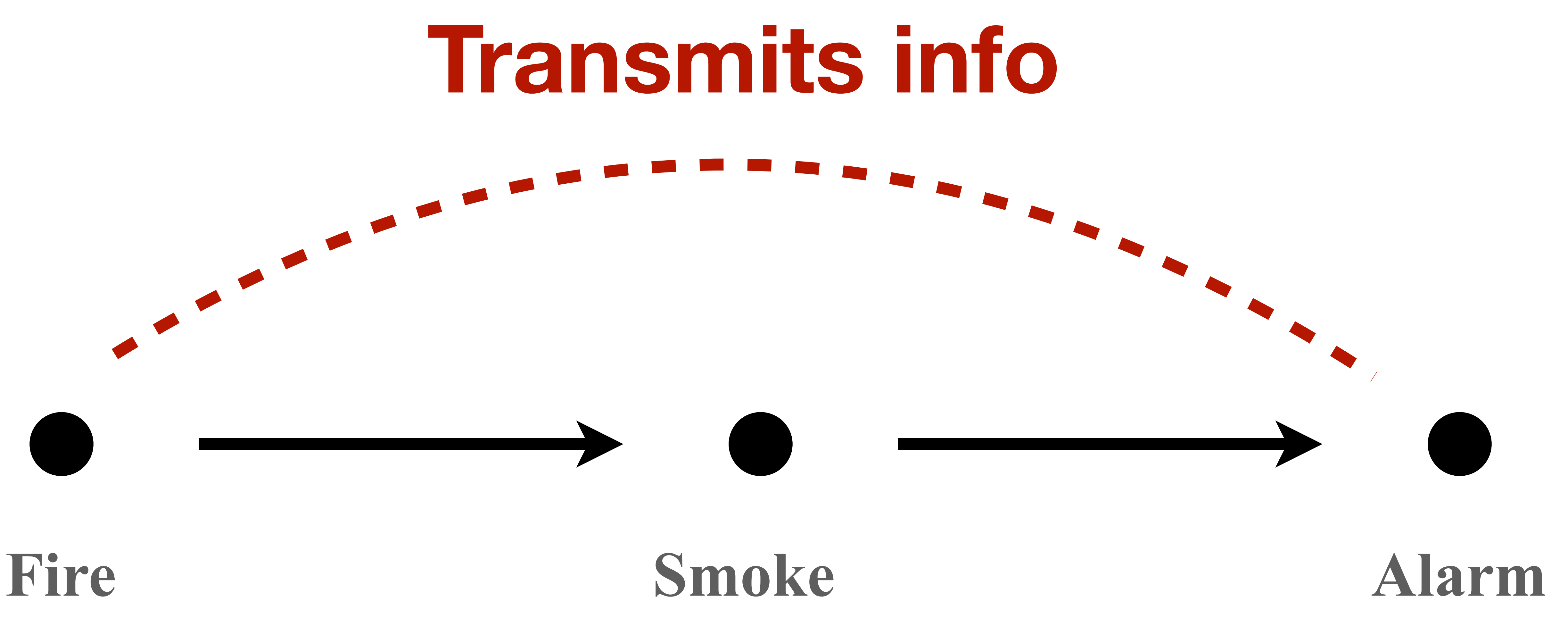
NB: the three data sets I use in this section are fake — I generated them with a computer to illustrate my points



Transmits info
Fire
Smoke
Alarm

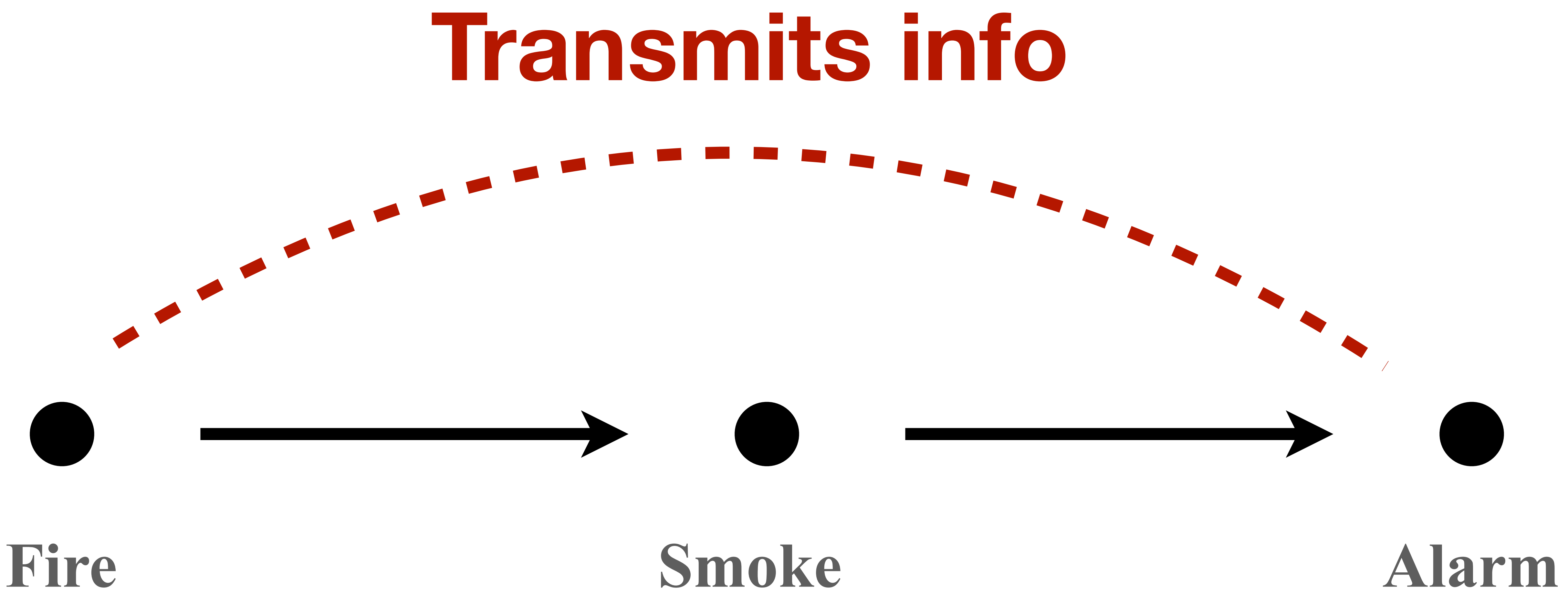


	No alarm	Alarm
No fire	9338	153
Fire	9	500



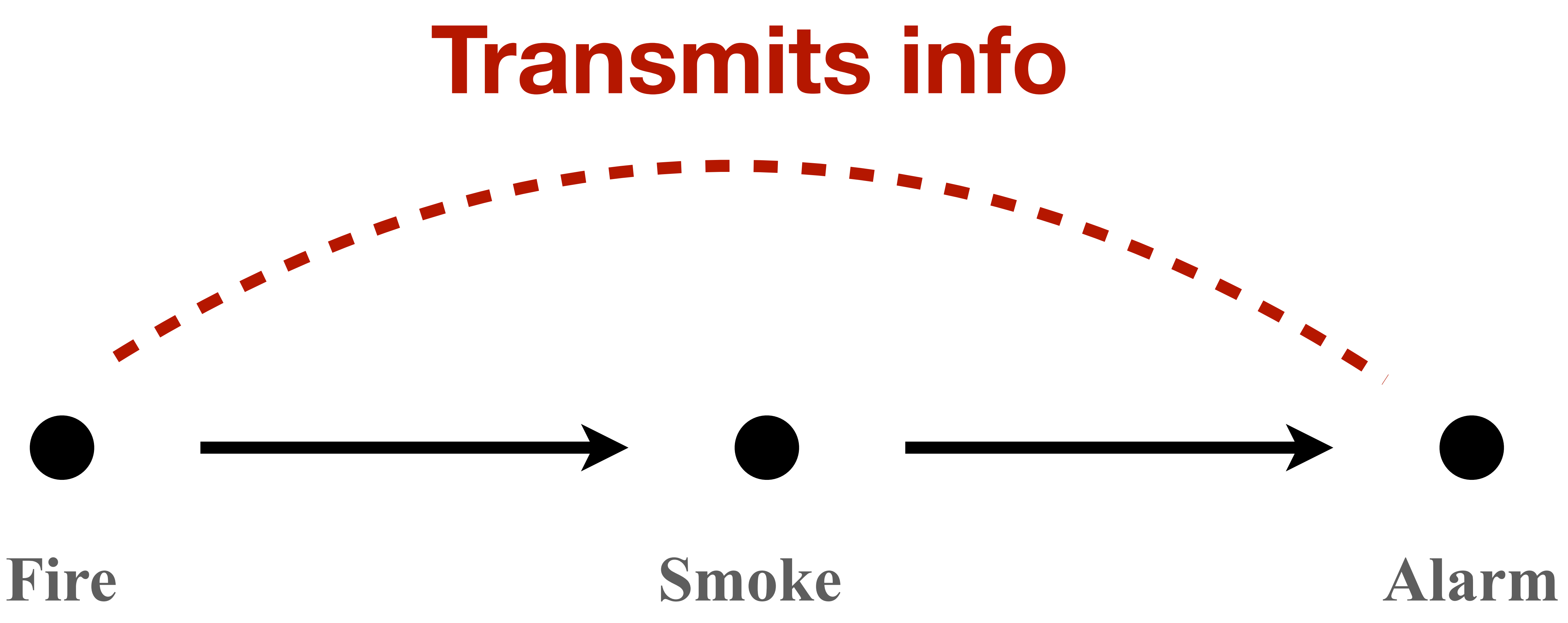
	No alarm	Alarm
No fire	9338	153
Fire	9	500

P(fire) = ________________



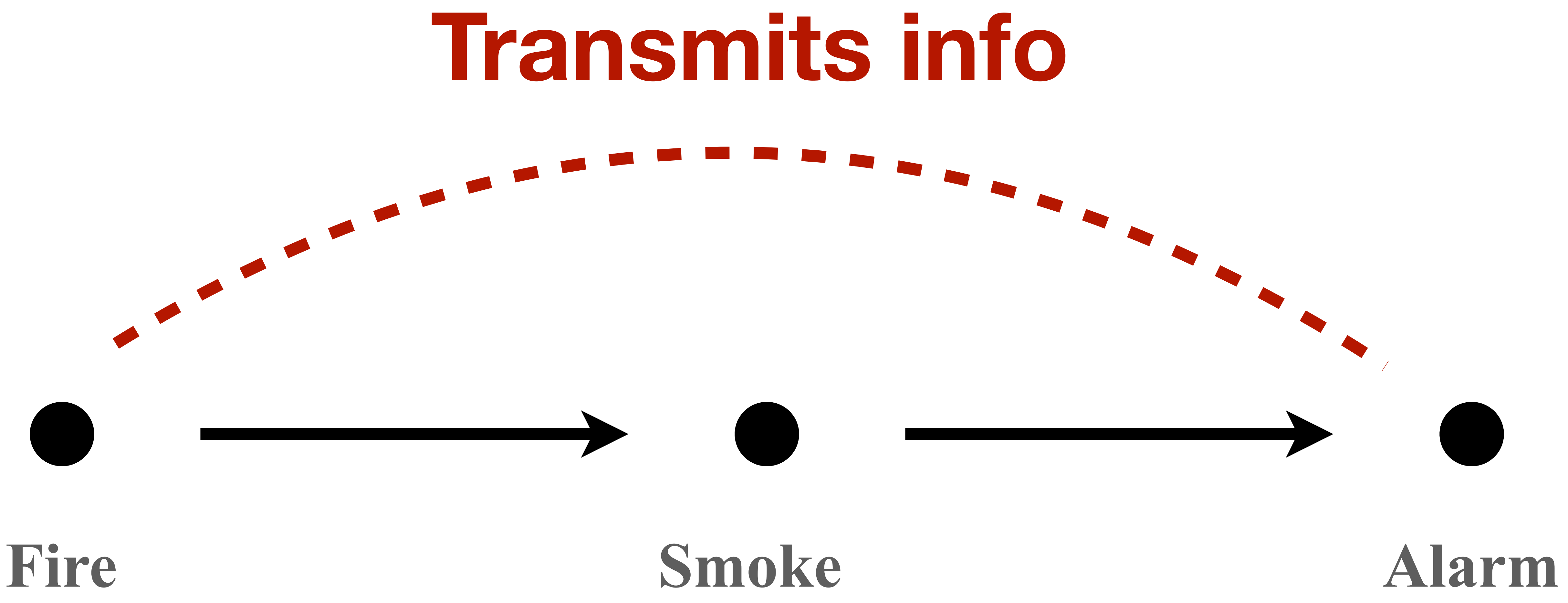
	No alarm	Alarm
No fire	9338	153
Fire	9	500

$$P(\text{ fire }) = \frac{9 + 500}{\qquad\qquad}$$



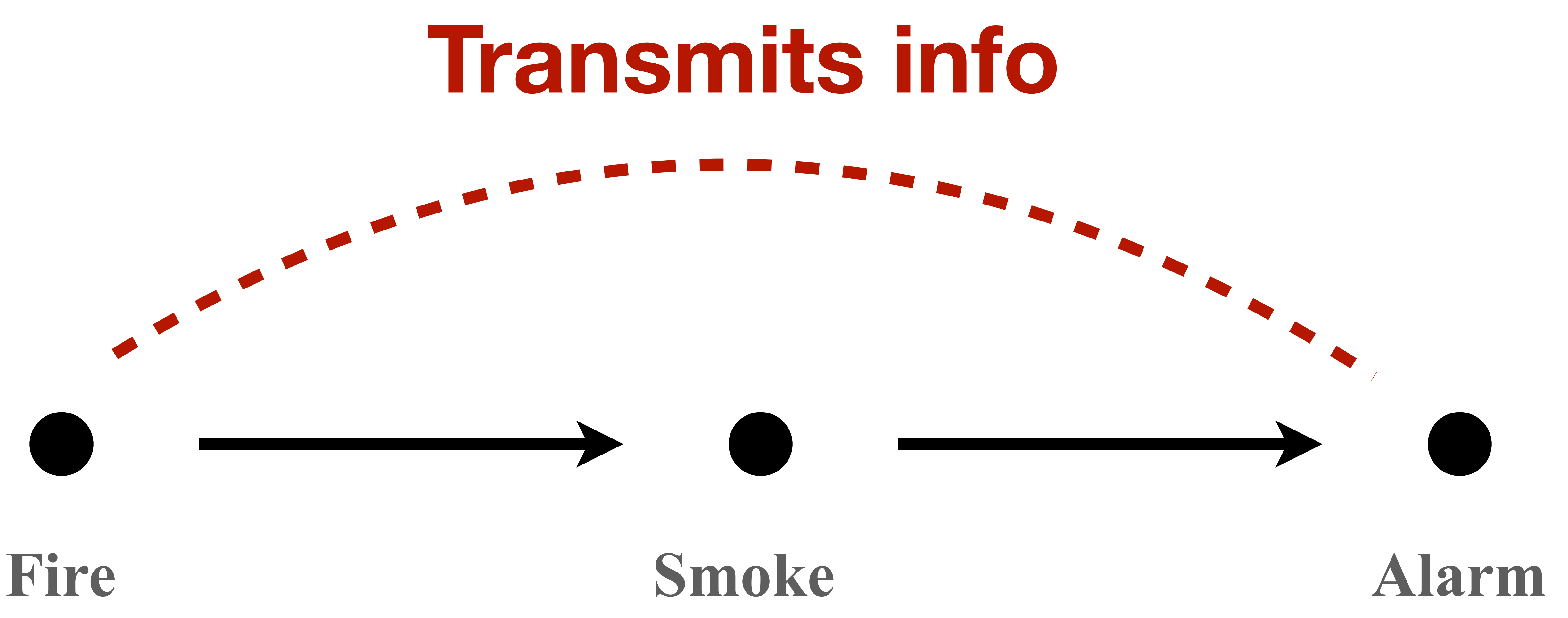
	No alarm	Alarm
No fire	9338	153
Fire	9	500

$$P(\text{ fire }) = \frac{9 + 500}{9 + 500 + 9338 + 153} =$$



	No alarm	Alarm
No fire	9338	153
Fire	9	500

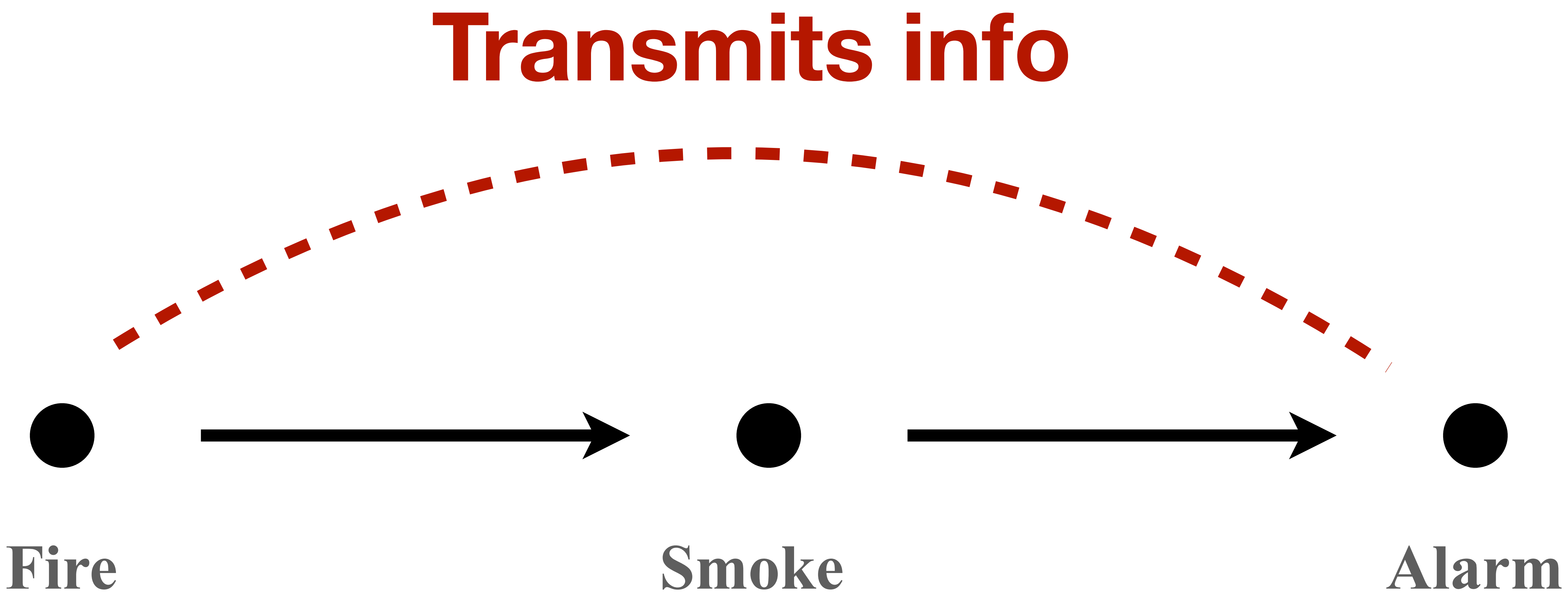
$$P(\text{ fire }) = \frac{9 + 500}{9 + 500 + 9338 + 153} = \frac{509}{10\,000} \approx 0.05$$



	No alarm	Alarm
No fire	9338	153
Fire	9	500

$$P(\text{ fire }) = \frac{9 + 500}{9 + 500 + 9338 + 153} = \frac{509}{10\,000} \approx 0.05$$

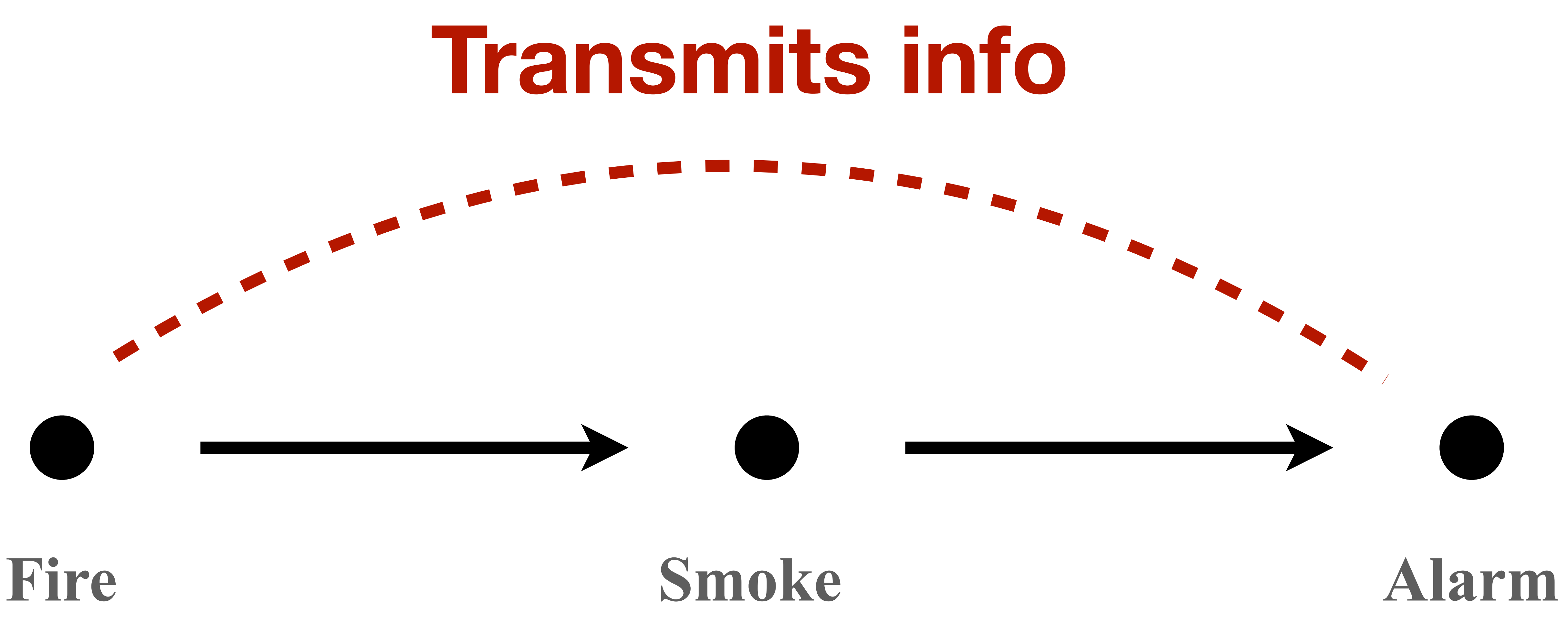
$$P(\text{ fire} \mid \text{alarm }) = \underline{\qquad\qquad}$$



	Alarm
No fire	153
Fire	500

$$P(\text{ fire }) = \frac{9 + 500}{9 + 500 + 9338 + 153} = \frac{509}{10\,000} \approx 0.05$$

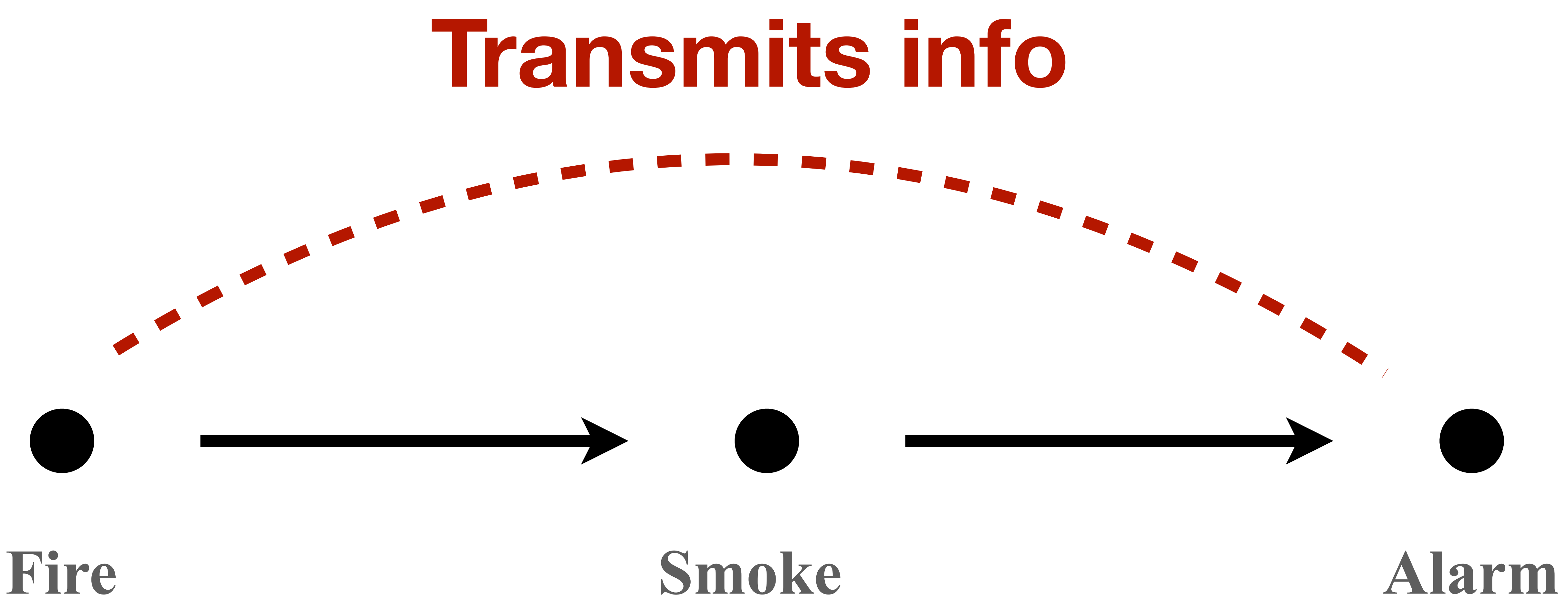
$$P(\text{ fire } \mid \text{ alarm }) = \underline{\qquad\qquad}$$



	Alarm
No fire	153
Fire	500

$$P(\text{ fire }) = \frac{9 + 500}{9 + 500 + 9338 + 153} = \frac{509}{10\,000} \approx 0.05$$

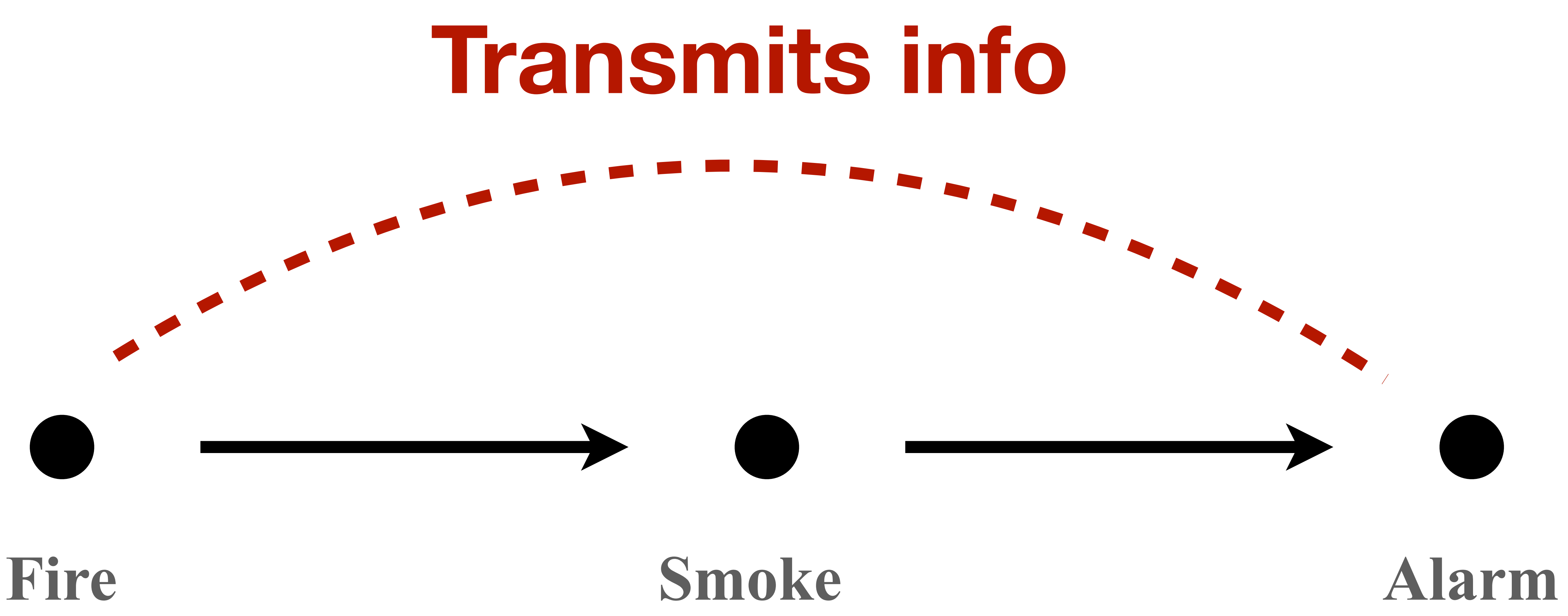
$$P(\text{ fire } \mid \text{ alarm }) = \frac{500}{}$$



	Alarm
No fire	153
Fire	500

$$P(\text{ fire }) = \frac{9 + 500}{9 + 500 + 9338 + 153} = \frac{509}{10\ 000} \approx 0.05$$

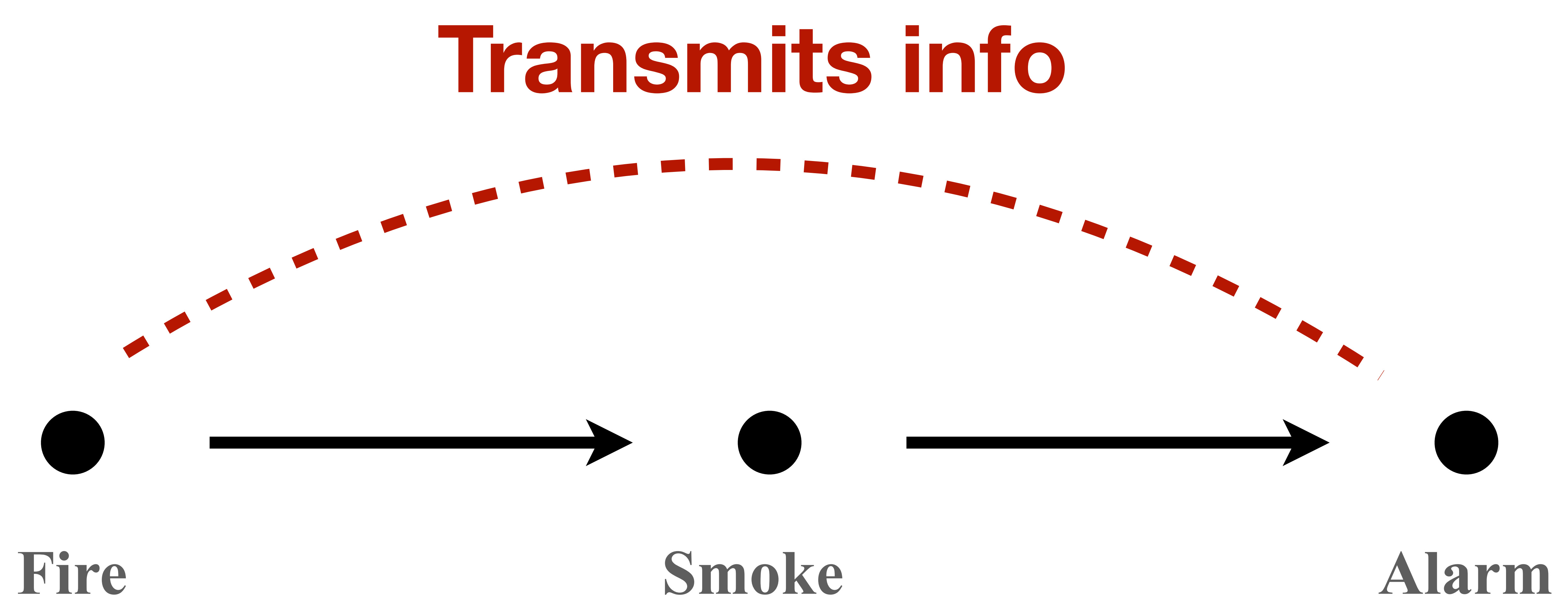
$$P(\text{ fire } | \text{ alarm }) = \frac{500}{500 + 153}$$



	Alarm
No fire	153
Fire	500

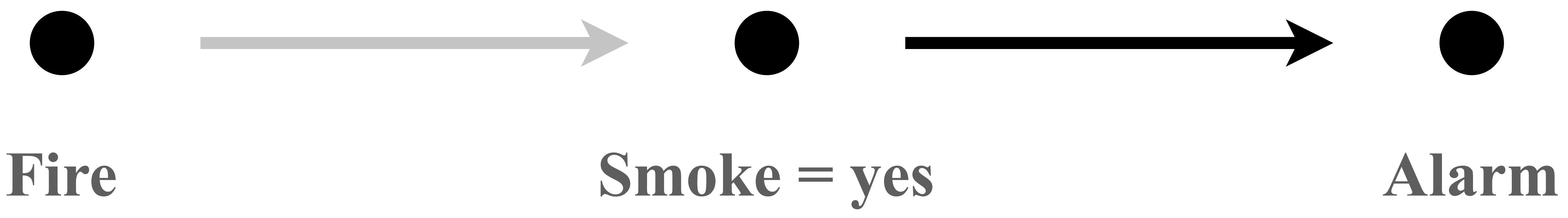
$$P(\text{ fire }) = \frac{9 + 500}{9 + 500 + 9338 + 153} = \frac{509}{10\,000} \approx 0.05$$

$$P(\text{ fire} \mid \text{alarm }) = \frac{500}{500 + 153} \approx 0.77$$



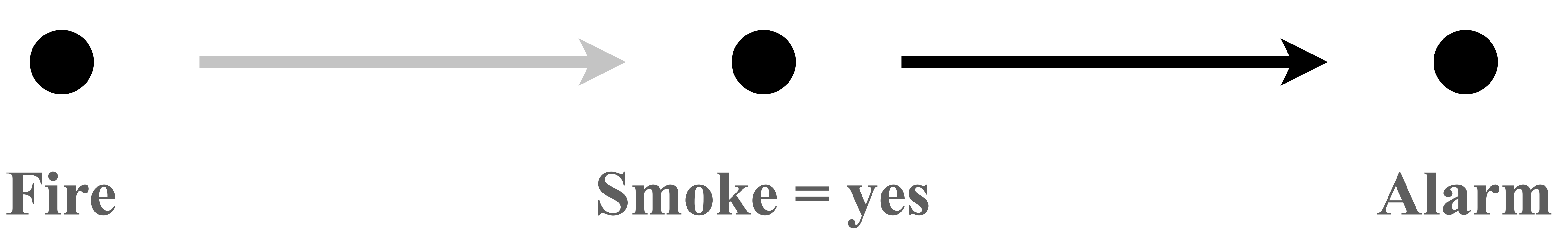
POINT: hearing the alarm gives us information about the likelihood of any ongoing fires

No longer transmits info



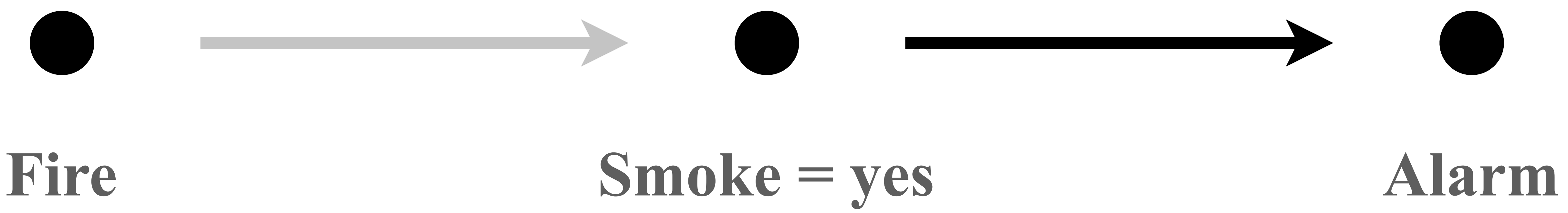
	No alarm	Alarm
No fire	9338	153
Fire	9	500

No longer transmits info



	No alarm	Alarm	No alarm	Alarm
No fire	1	92	9337	61
Fire	6	500	3	0
	Smoke = yes		Smoke = no	

No longer transmits info

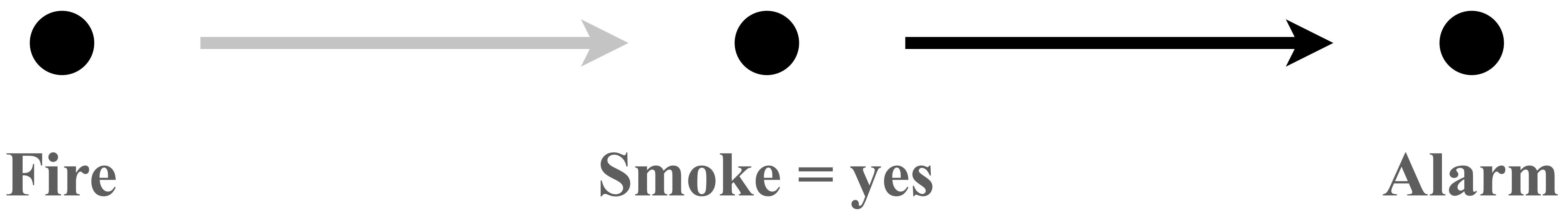


	No alarm	Alarm
No fire	1	92
Fire	6	500

Smoke = yes

P(fire | smoke) =

No longer transmits info

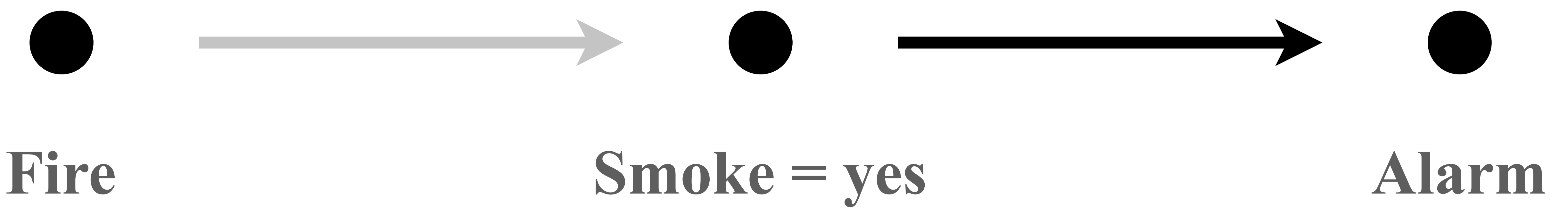


	No alarm	Alarm
No fire	1	92
Fire	6	500

Smoke = yes

P(fire | smoke) = 506/599

No longer transmits info

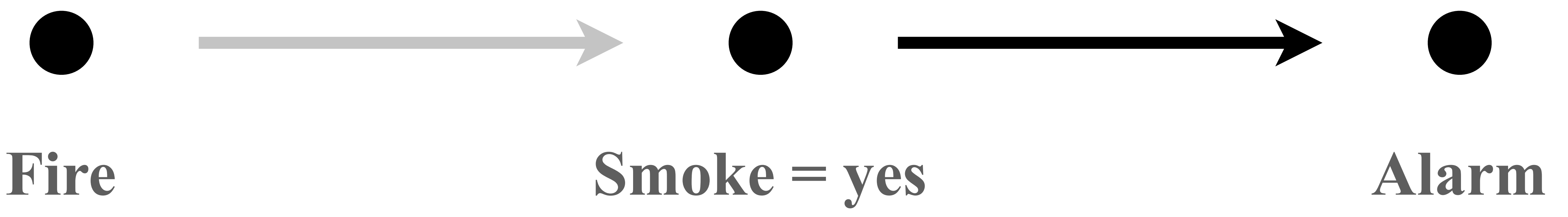


	No alarm	Alarm
No fire	1	92
Fire	6	500

Smoke = yes

$$P(\text{ fire } | \text{ smoke }) = 506/599 \approx 0.84$$

No longer transmits info



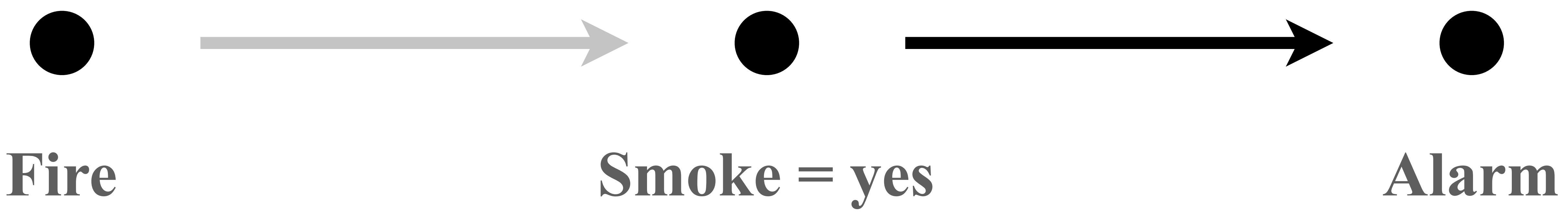
	No alarm	Alarm
No fire	1	92
Fire	6	500

Smoke = yes

$$P(\text{ fire} \mid \text{smoke }) = 506/599 \approx 0.84$$

$$P(\text{ fire} \mid \text{alarm, smoke }) =$$

No longer transmits info



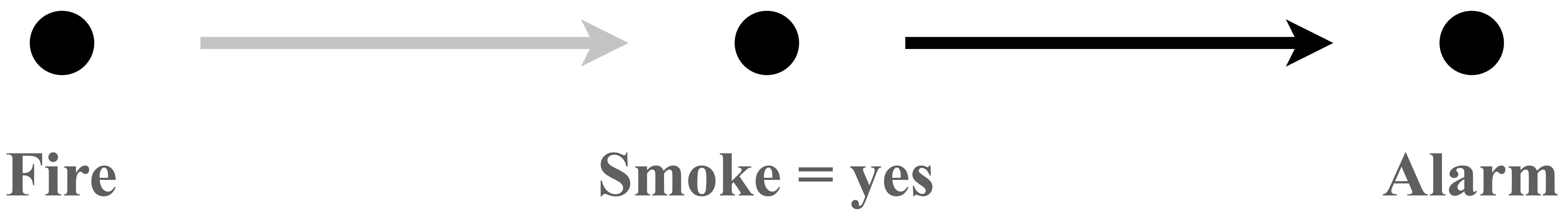
	No alarm	Alarm
No fire		92
Fire		500

Smoke = yes

P(fire | smoke) = 506/599 ≈ 0.84

P(fire | alarm, smoke) =

No longer transmits info



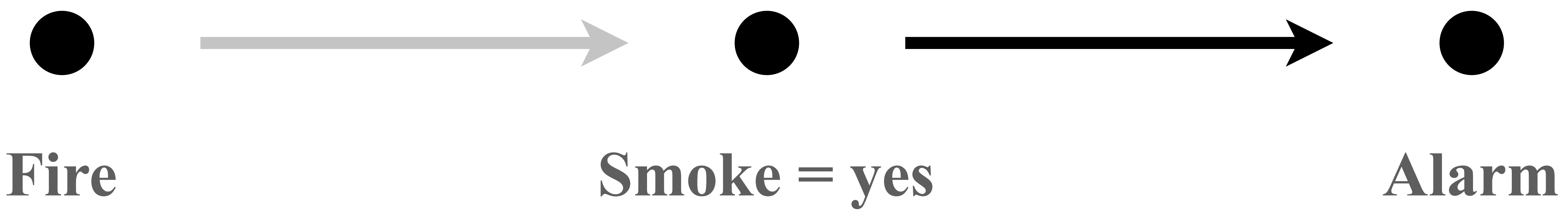
	No alarm	Alarm
No fire		92
Fire		500

Smoke = yes

$$P(\text{ fire} \mid \text{smoke }) = 506/599 \approx 0.84$$

$$P(\text{ fire} \mid \text{alarm, smoke }) = 500/592$$

No longer transmits info



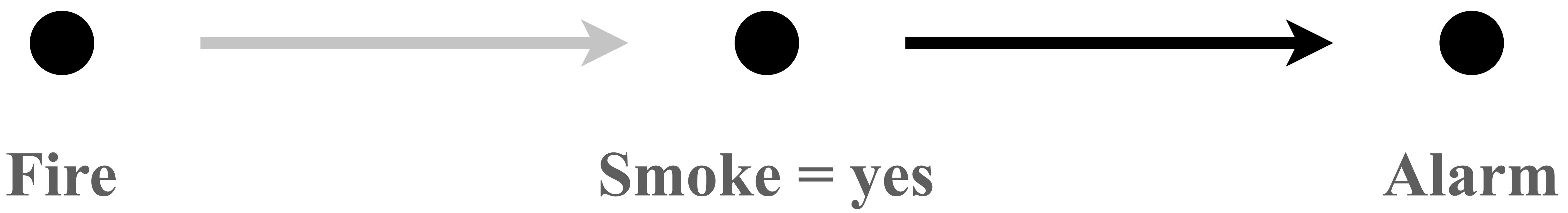
	No alarm	Alarm
No fire		92
Fire		500

Smoke = yes

$$P(\text{ fire} \mid \text{smoke }) = 506/599 \approx 0.84$$

$$P(\text{ fire} \mid \text{alarm, smoke }) = 500/592 \approx 0.84$$

No longer transmits info



	No alarm	Alarm
No fire	1	92
Fire	6	500

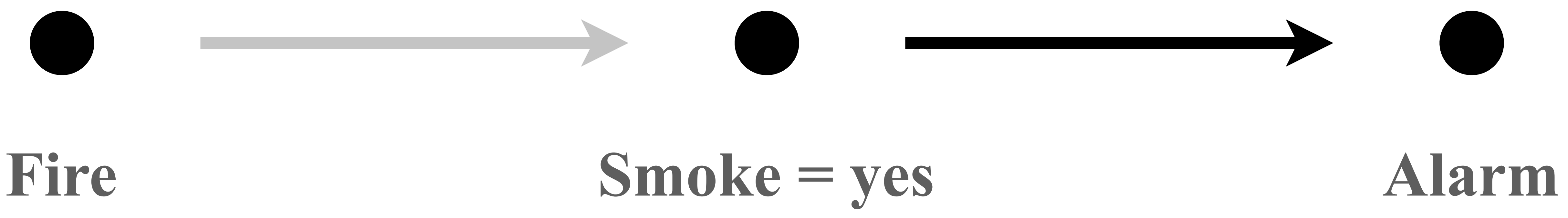
Smoke = yes

P(fire | smoke) = 506/599 ≈ 0.84

Conditional on smoke, the alarm gives no extra information!

P(fire | alarm, smoke) = 500/592 ≈ 0.84

No longer transmits info



	No alarm	Alarm
No fire	1	92
Fire	6	500

Smoke = yes

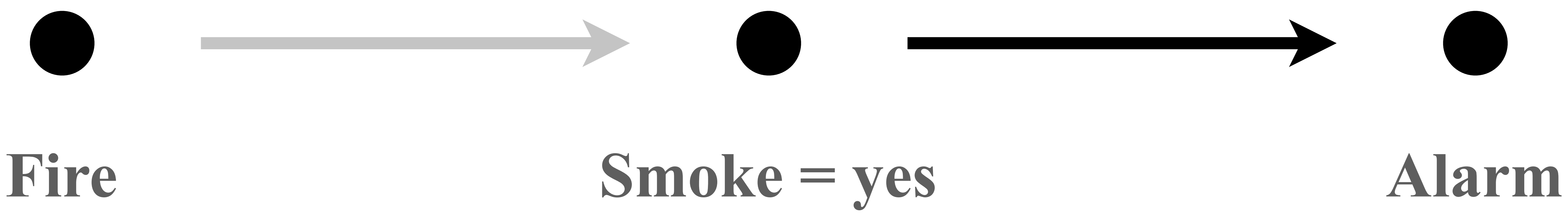
P(fire | smoke) = 506/599 ≈ 0.84

P(fire | alarm, smoke) = 500/592 ≈ 0.84

Conditional on smoke, the alarm gives no extra information!

If I knew nothing about fire, smoke, or alarms, and analyzed these data haphazardly, I would conclude that smoke alarms are useless

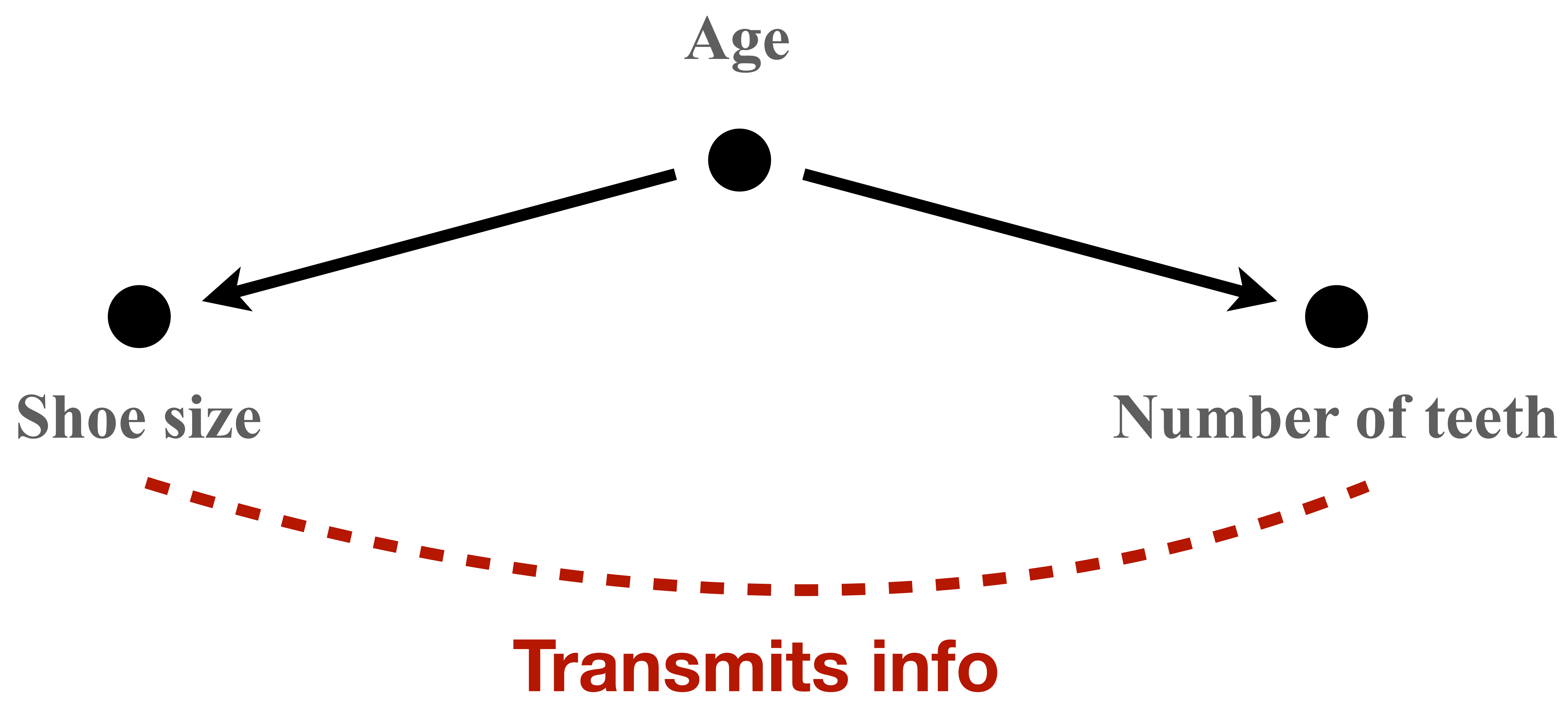
No longer transmits info

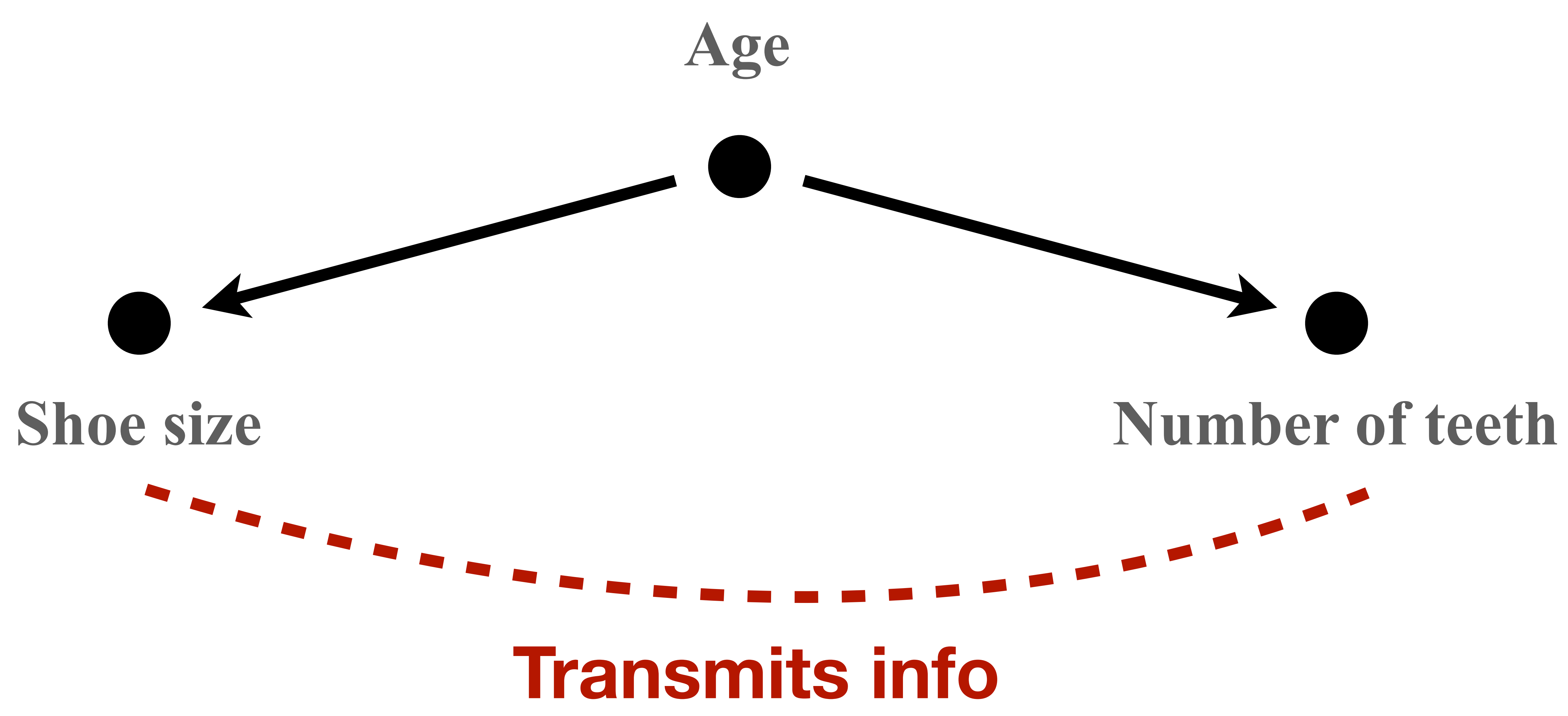


Conditional on smoke, the alarm gives no extra information!

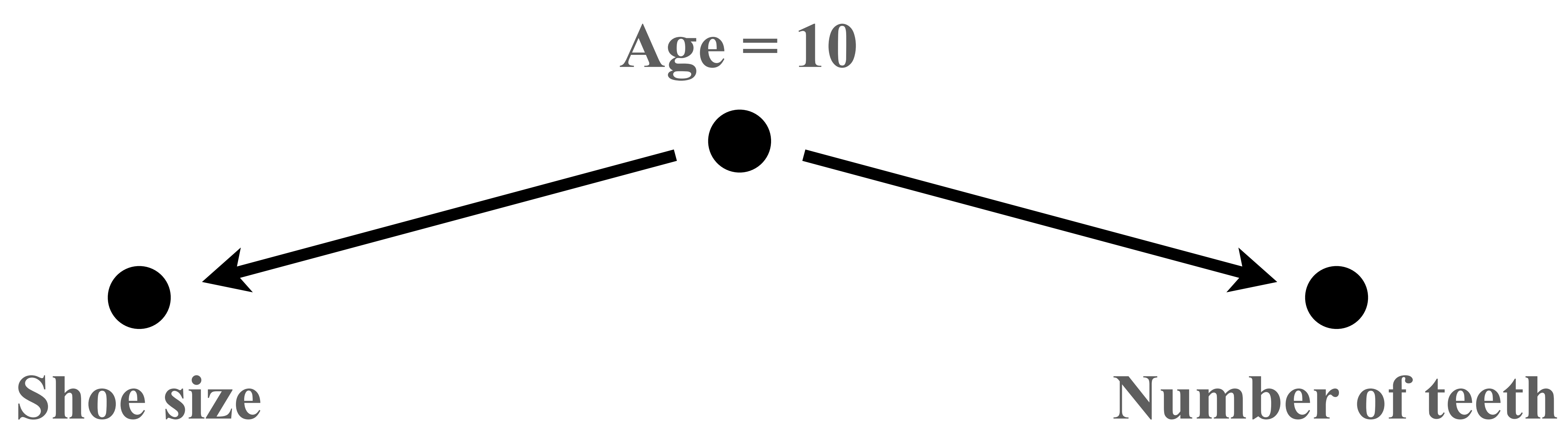
If I knew nothing about fire, smoke, or alarms, and analyzed these data haphazardly, I would conclude that smoke alarms are useless

“Conditioning on a mediator” — adjusting away the effect of interest

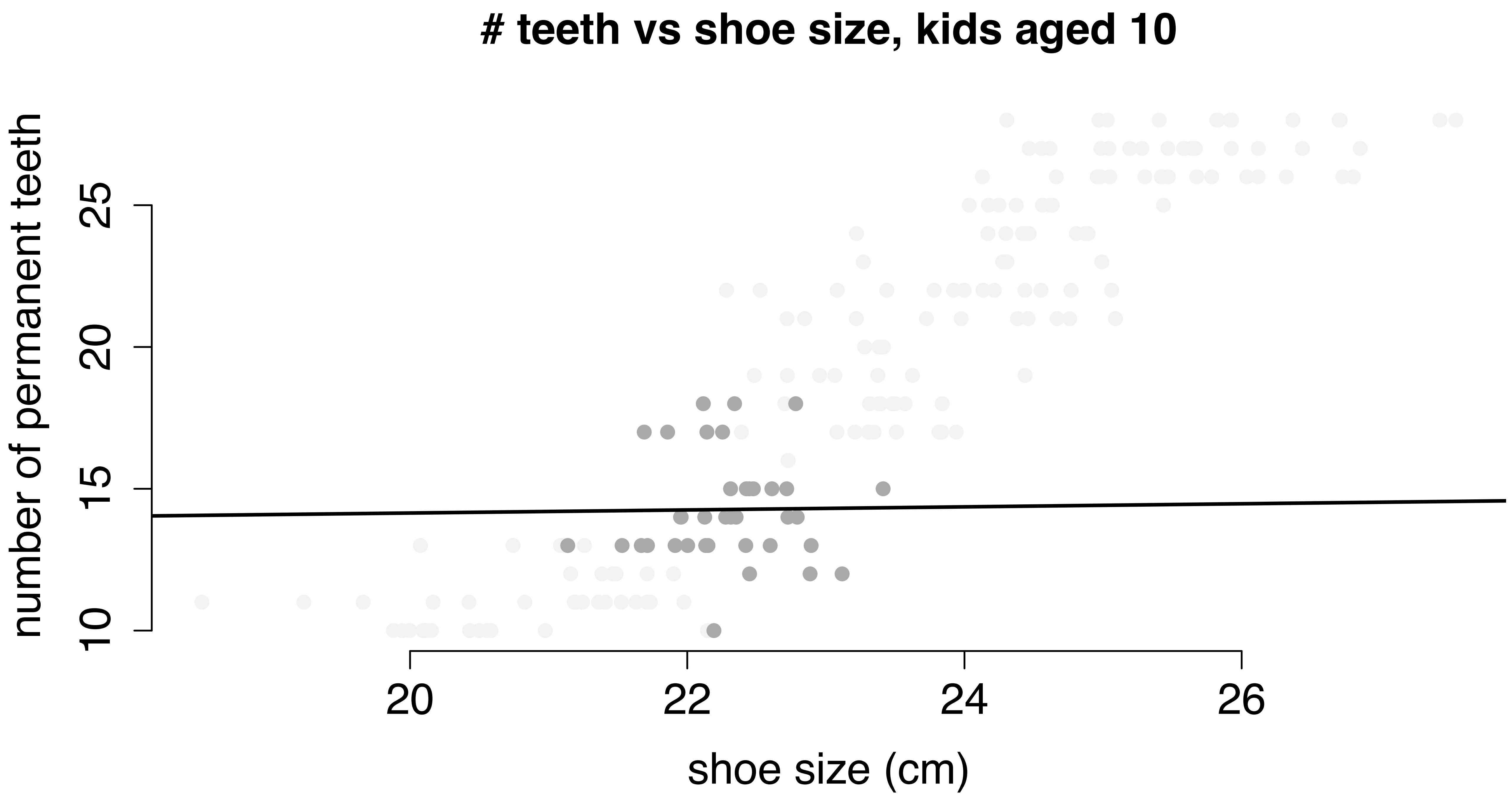

Age
Shoe size
Number of teeth
Transmits info

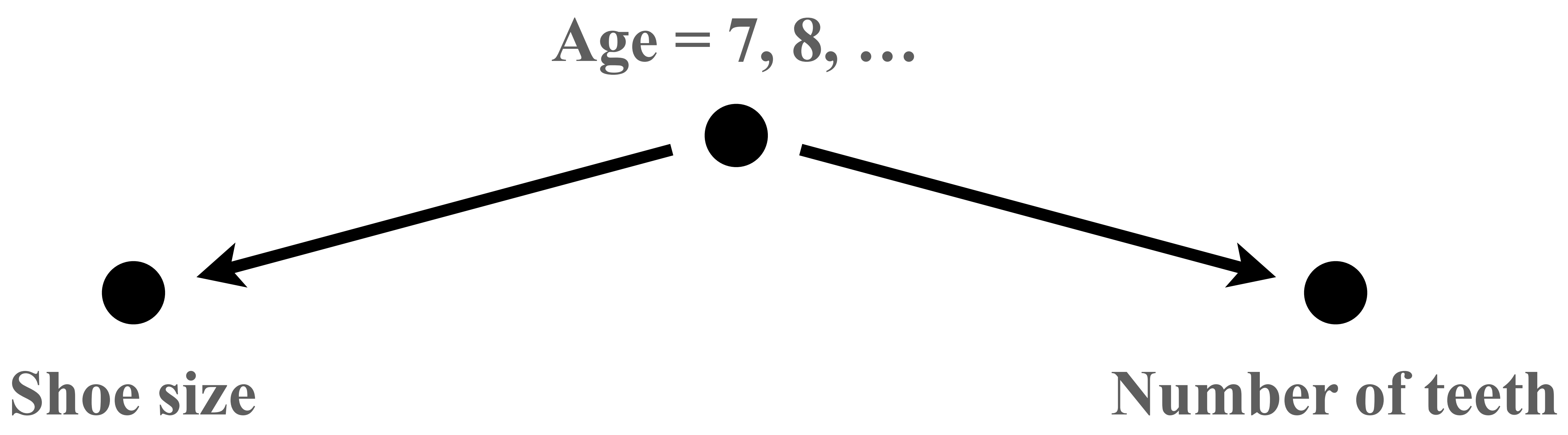

Age
Shoe size
Number of teeth
Transmits info


teeth vs shoe size, kids aged 7–15
number of permanent teeth
10
15
20
25
20
22
24
26
shoe size (cm)

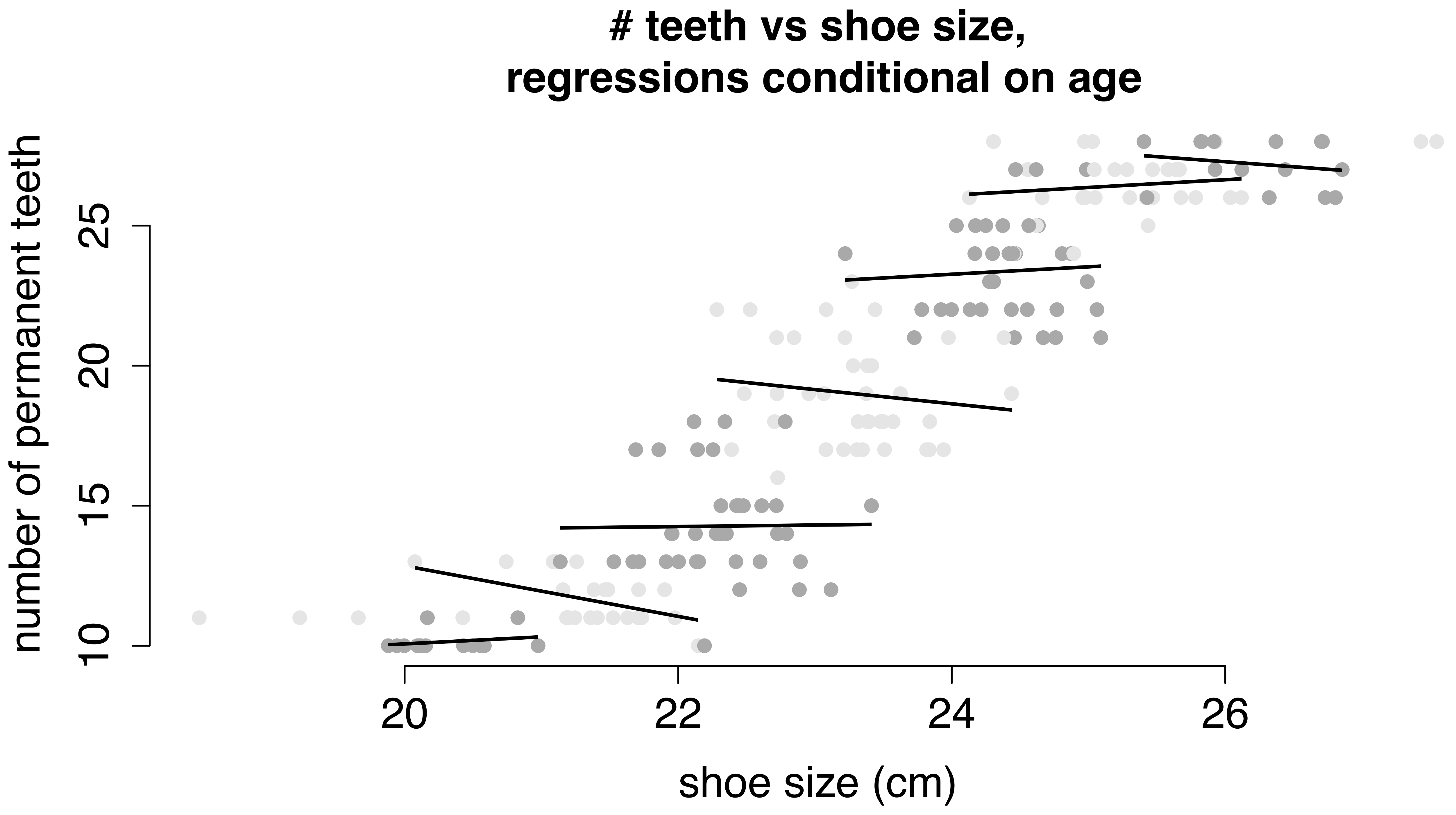


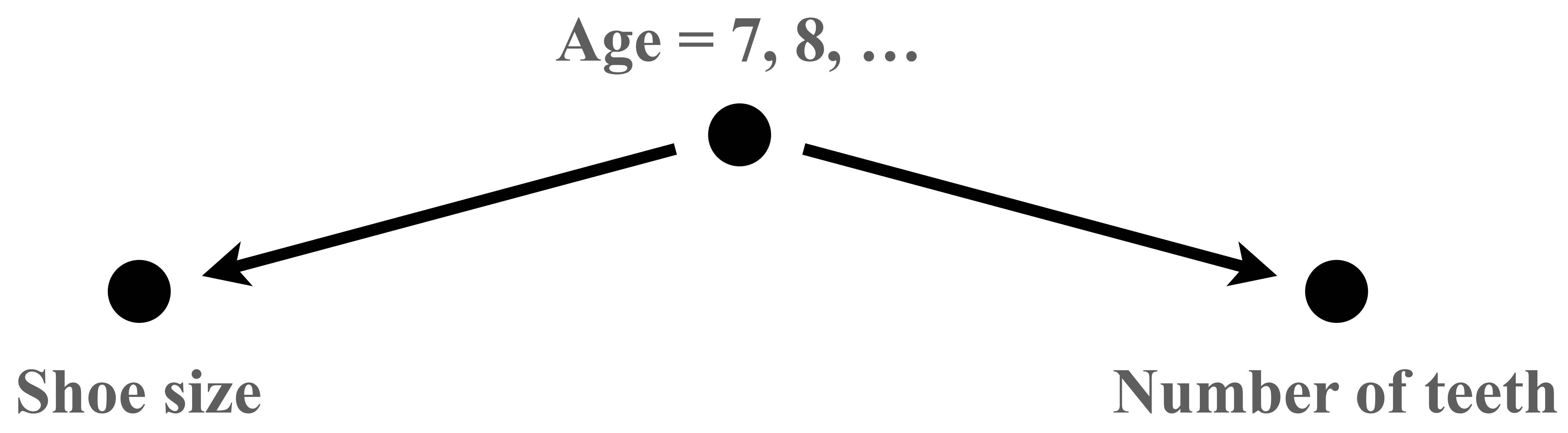
Does not transmit info



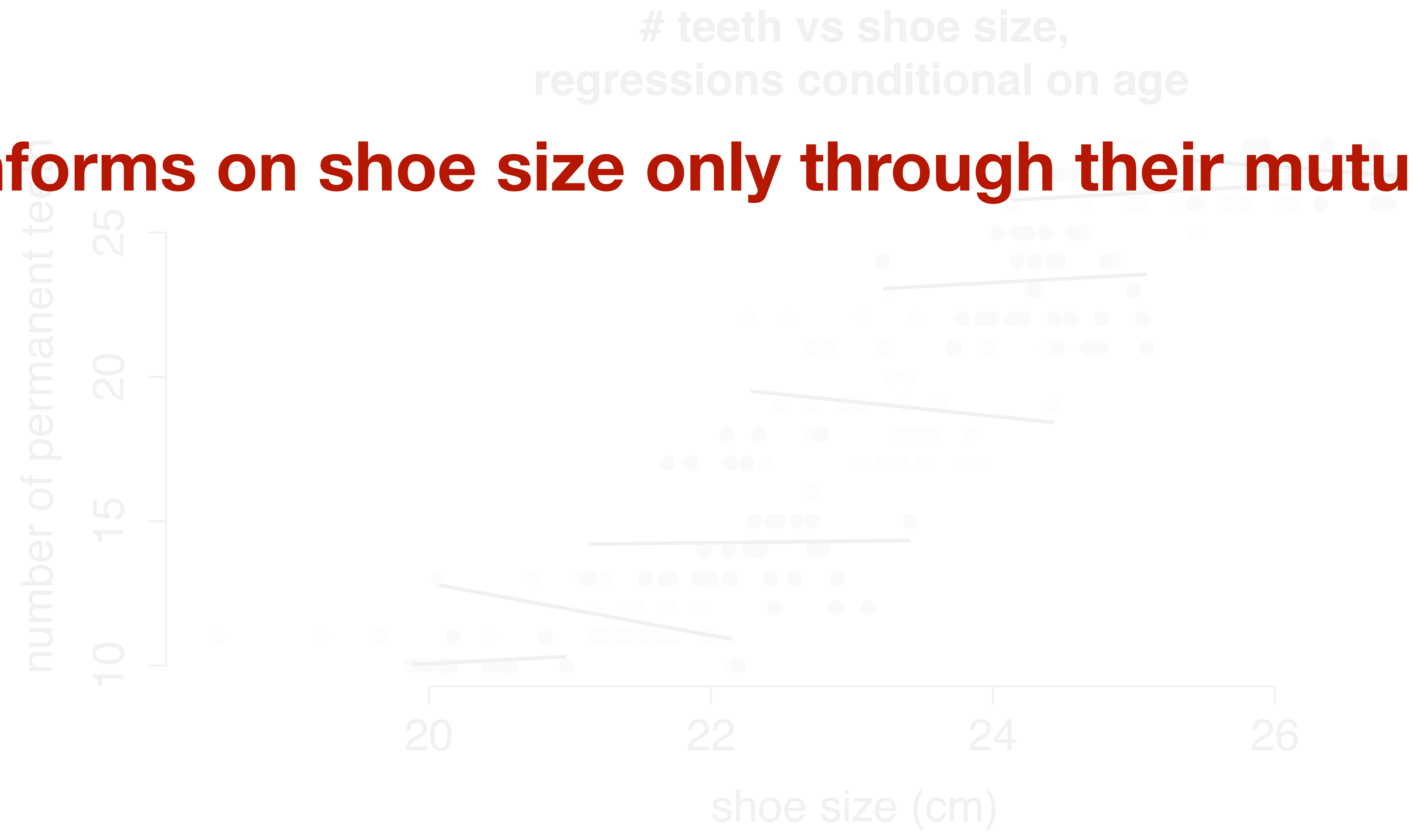


Does not transmit info

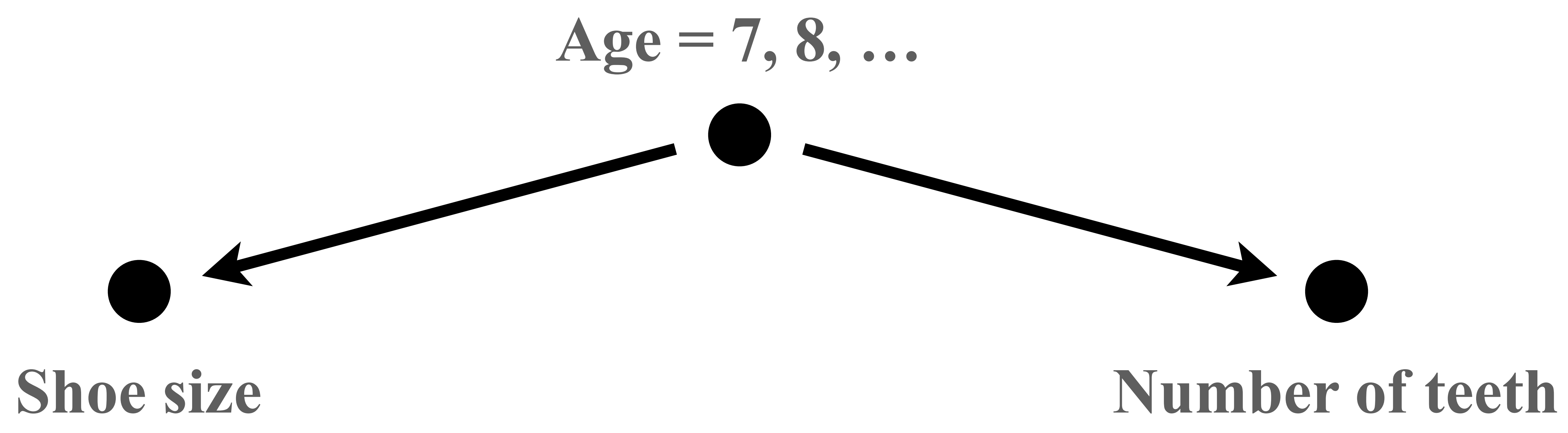




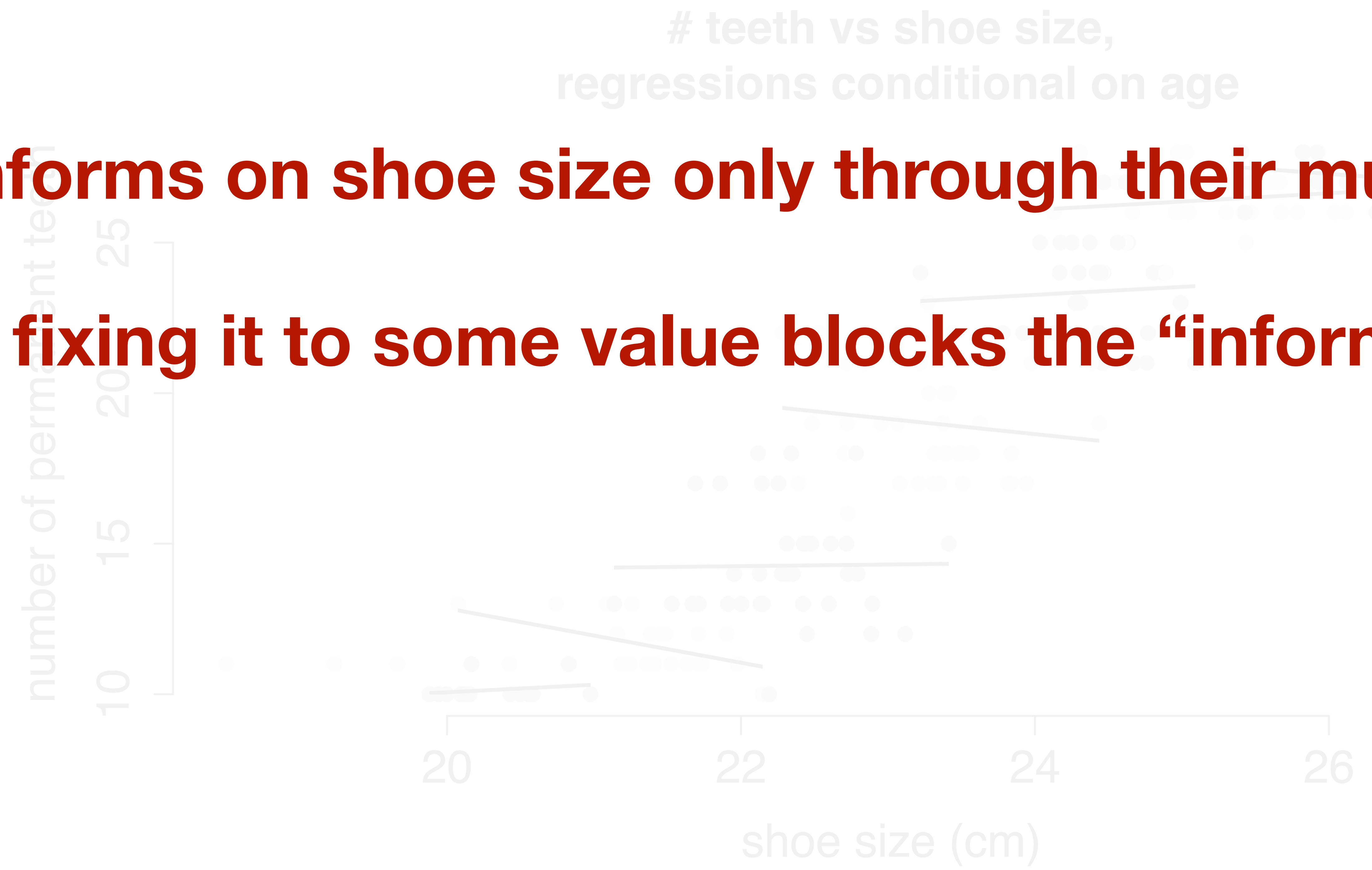
Does not transmit info



Number of teeth informs on shoe size only through their mutual association with age

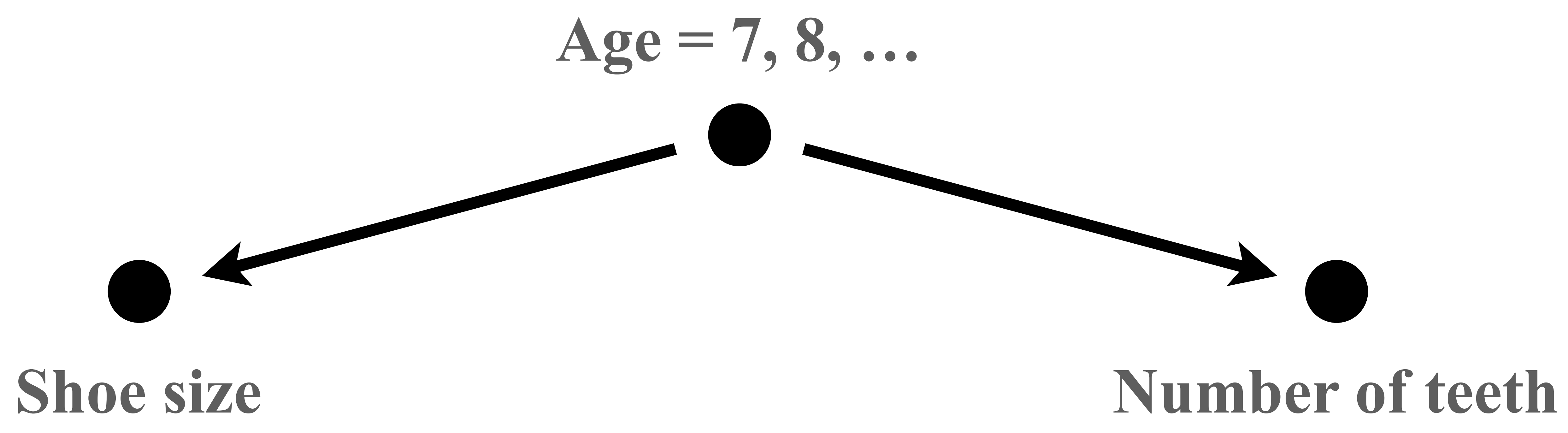


Does not transmit info



Number of teeth informs on shoe size only through their mutual association with age

Controlling age by fixing it to some value blocks the “information flow”



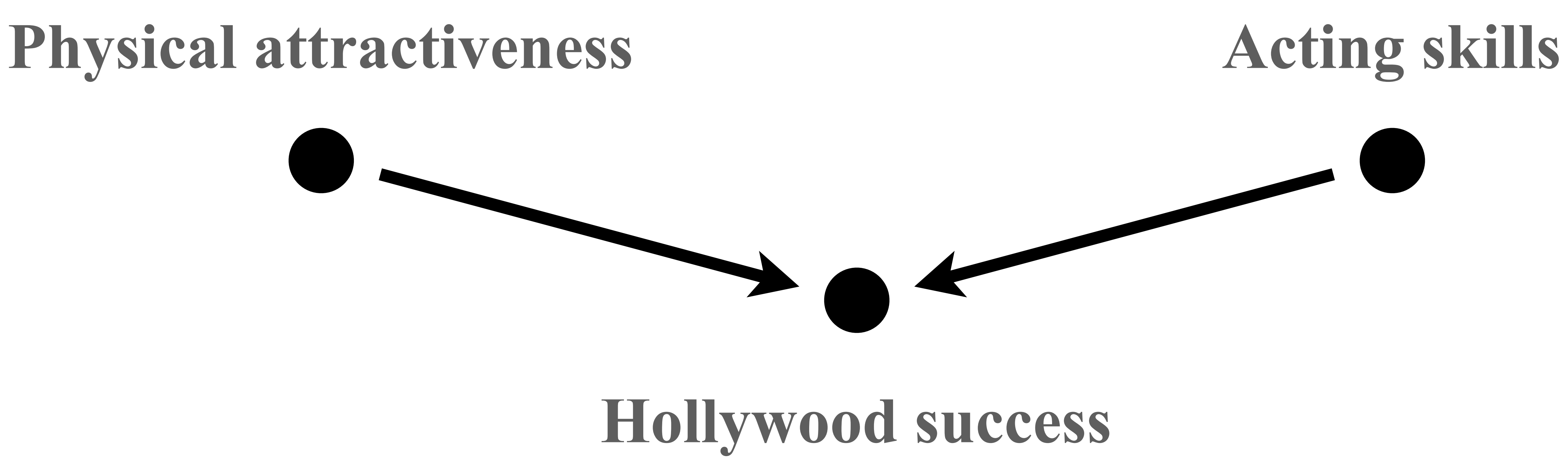
Does not transmit info

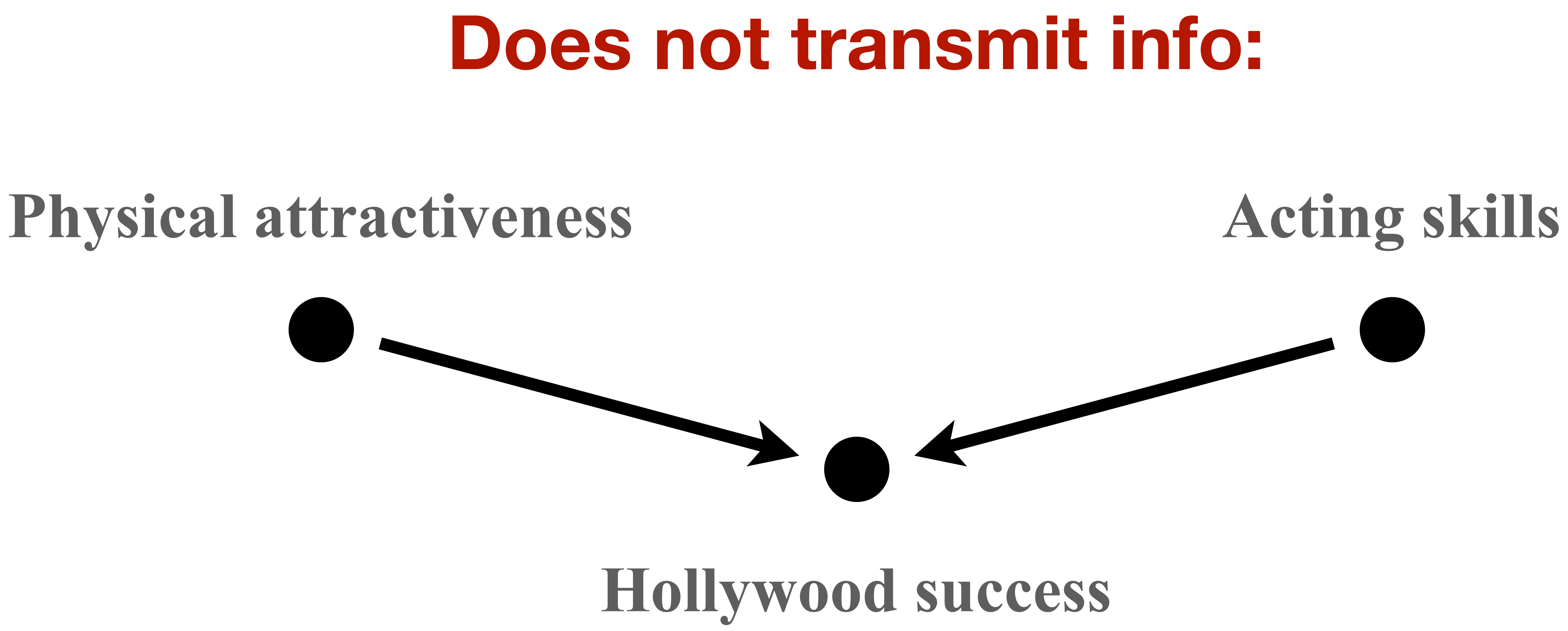
Number of teeth informs on shoe size only through their mutual association with age

Controlling age by fixing it to some value blocks the “information flow”

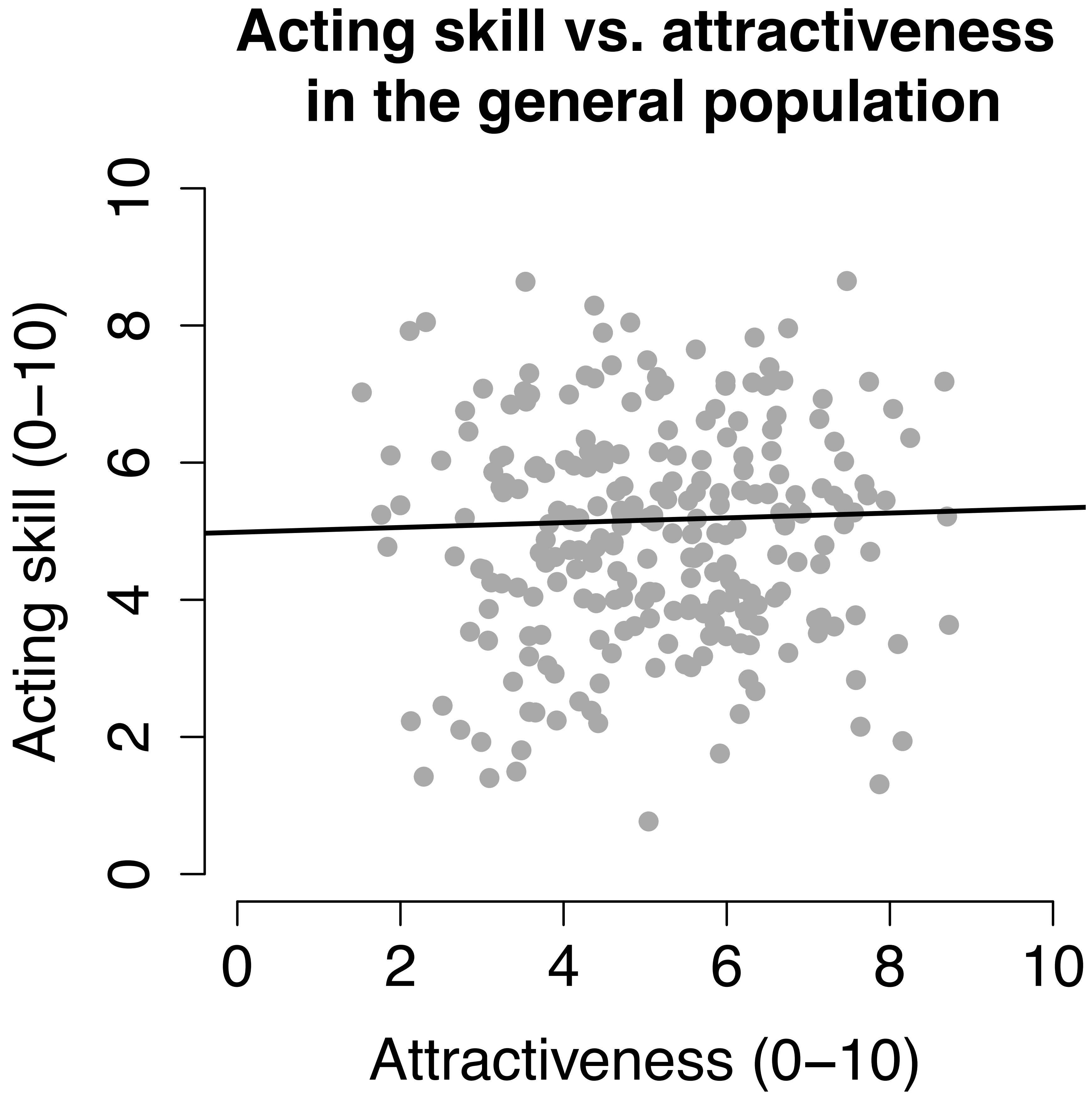
Ignoring age leaves “uncontrolled confounding”

Does not transmit info:

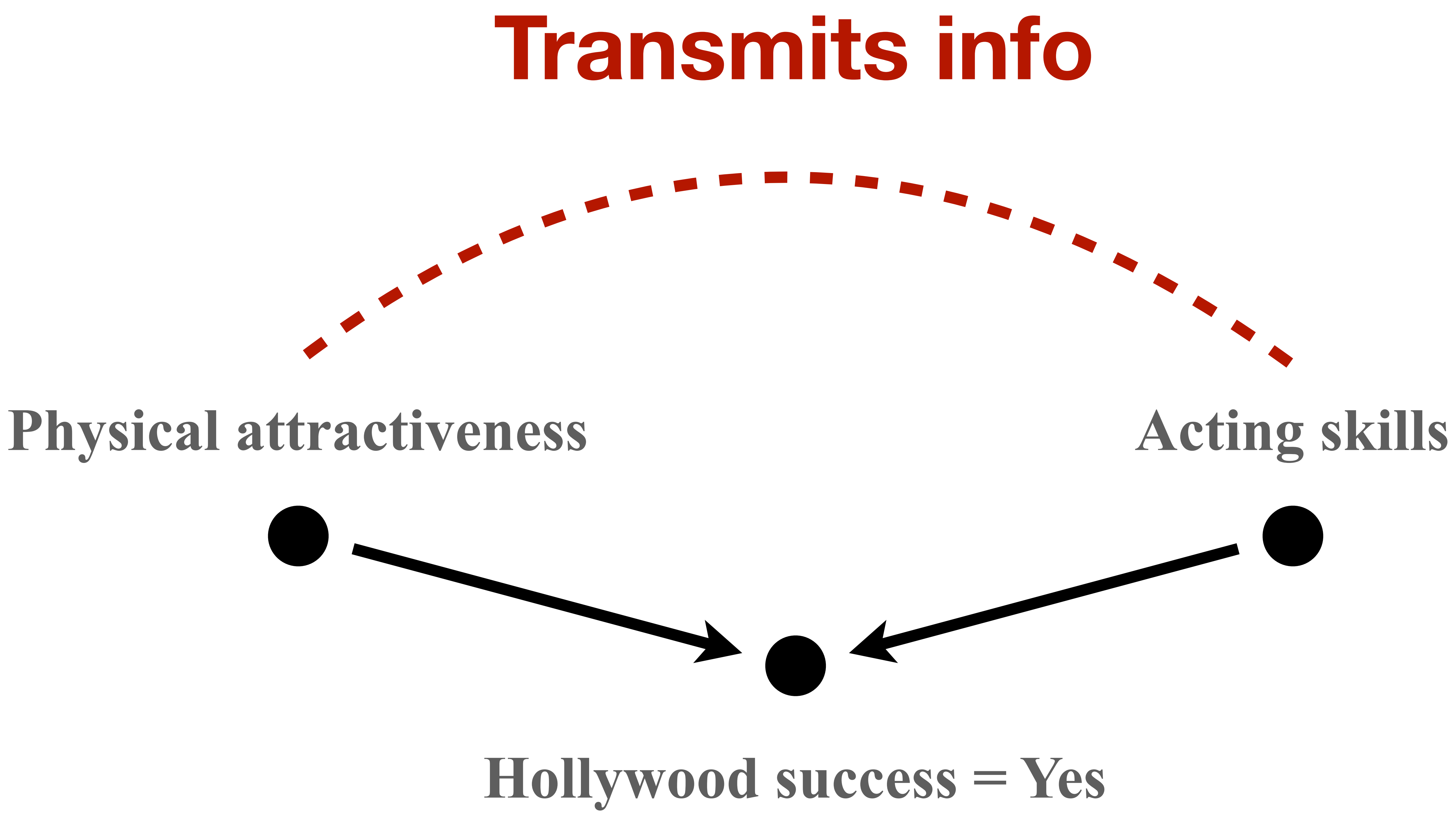


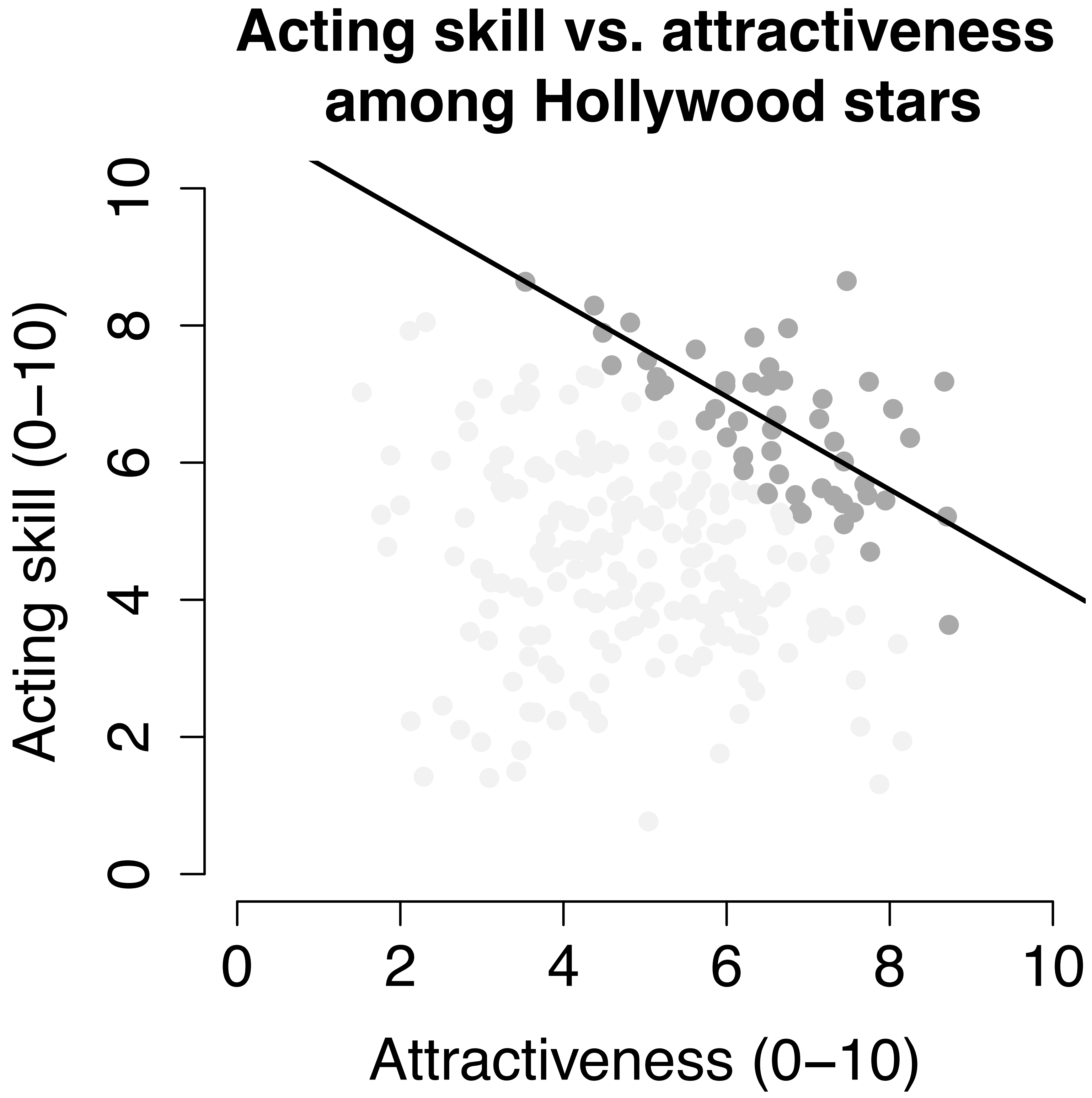


Does not transmit info:
Physical attractiveness
Acting skills
Hollywood success

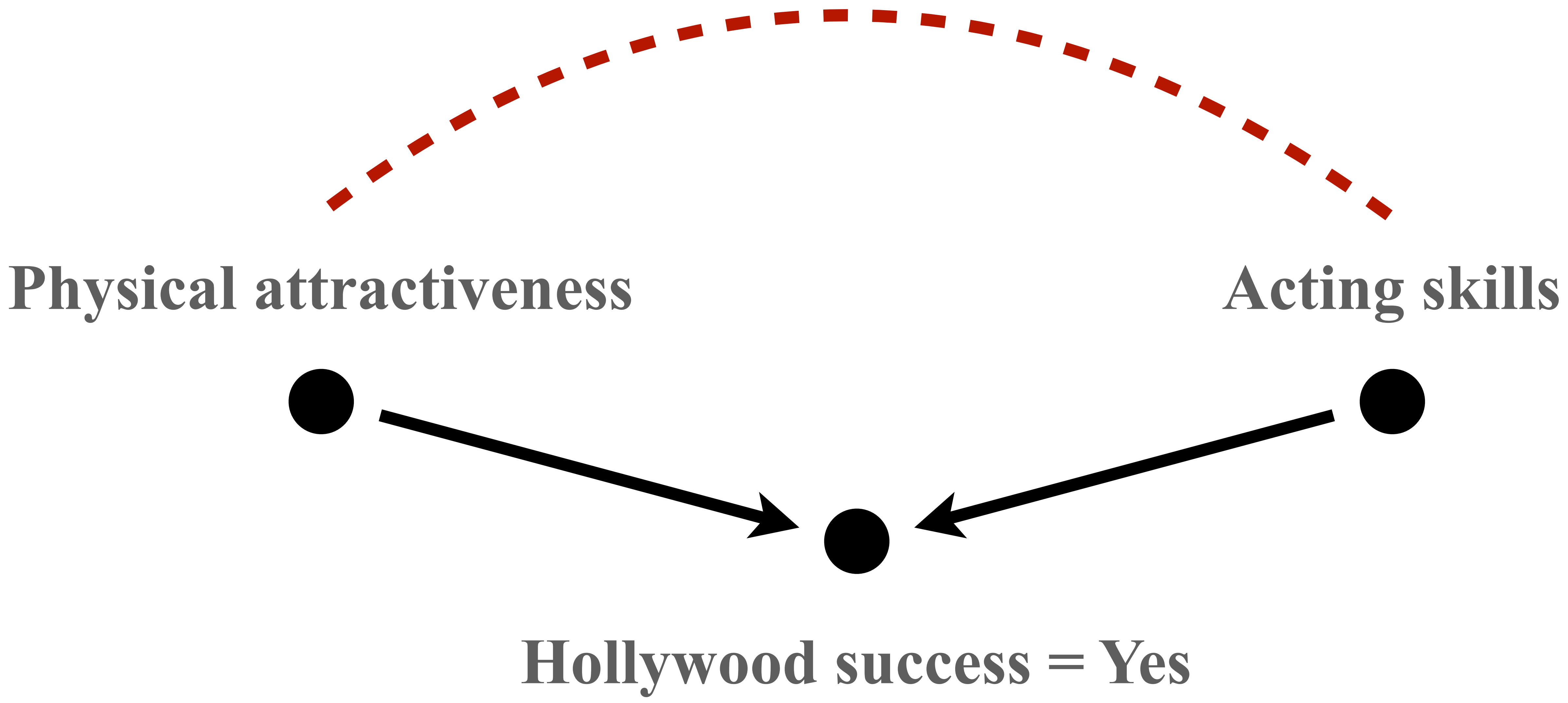


Acting skill vs. attractiveness in the general population
Acting skill (0–10)
0
2
4
6
8
10
Attractiveness (0–10)

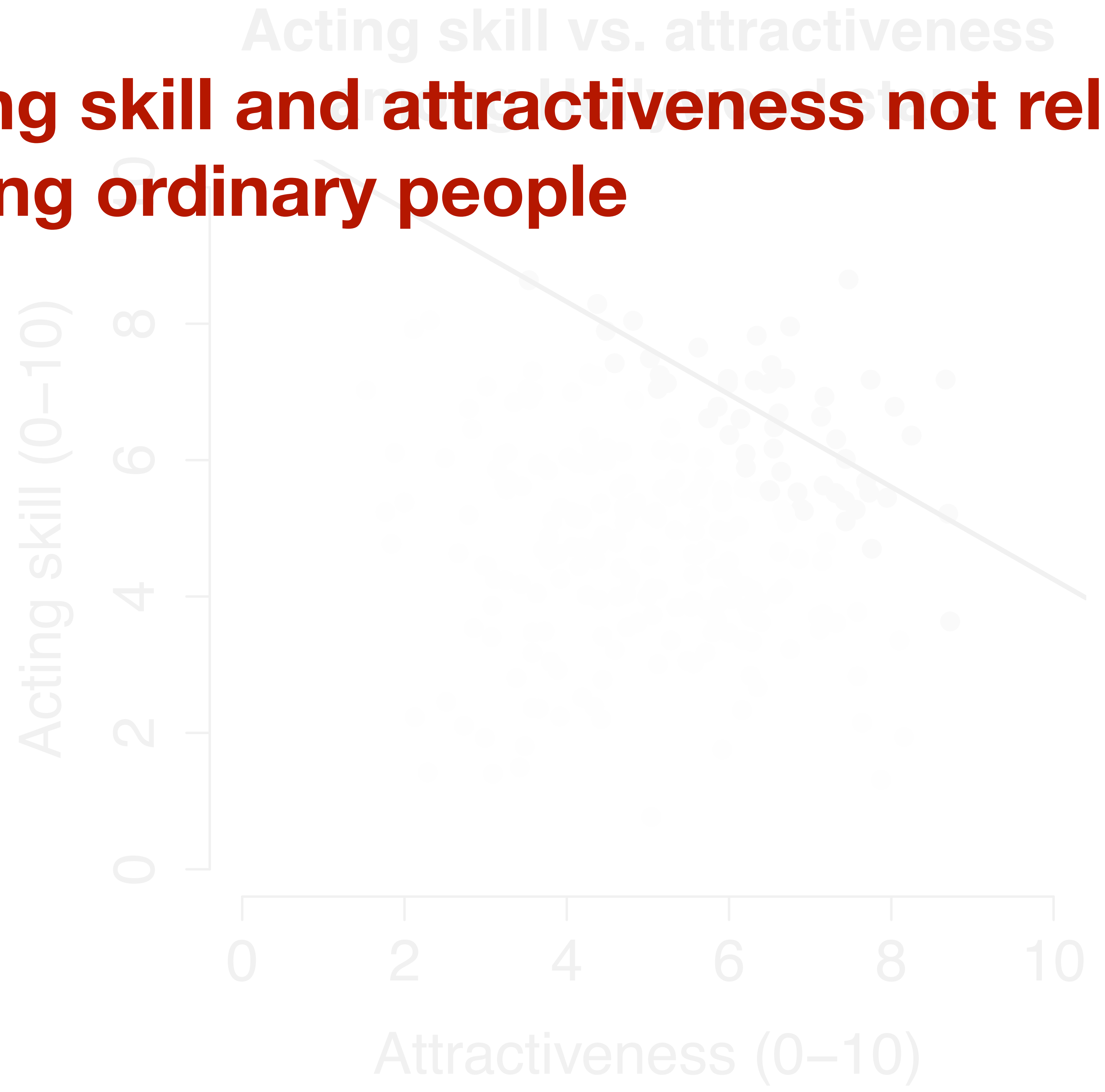

Transmits info
Physical attractiveness
Acting skills
Hollywood success = Yes

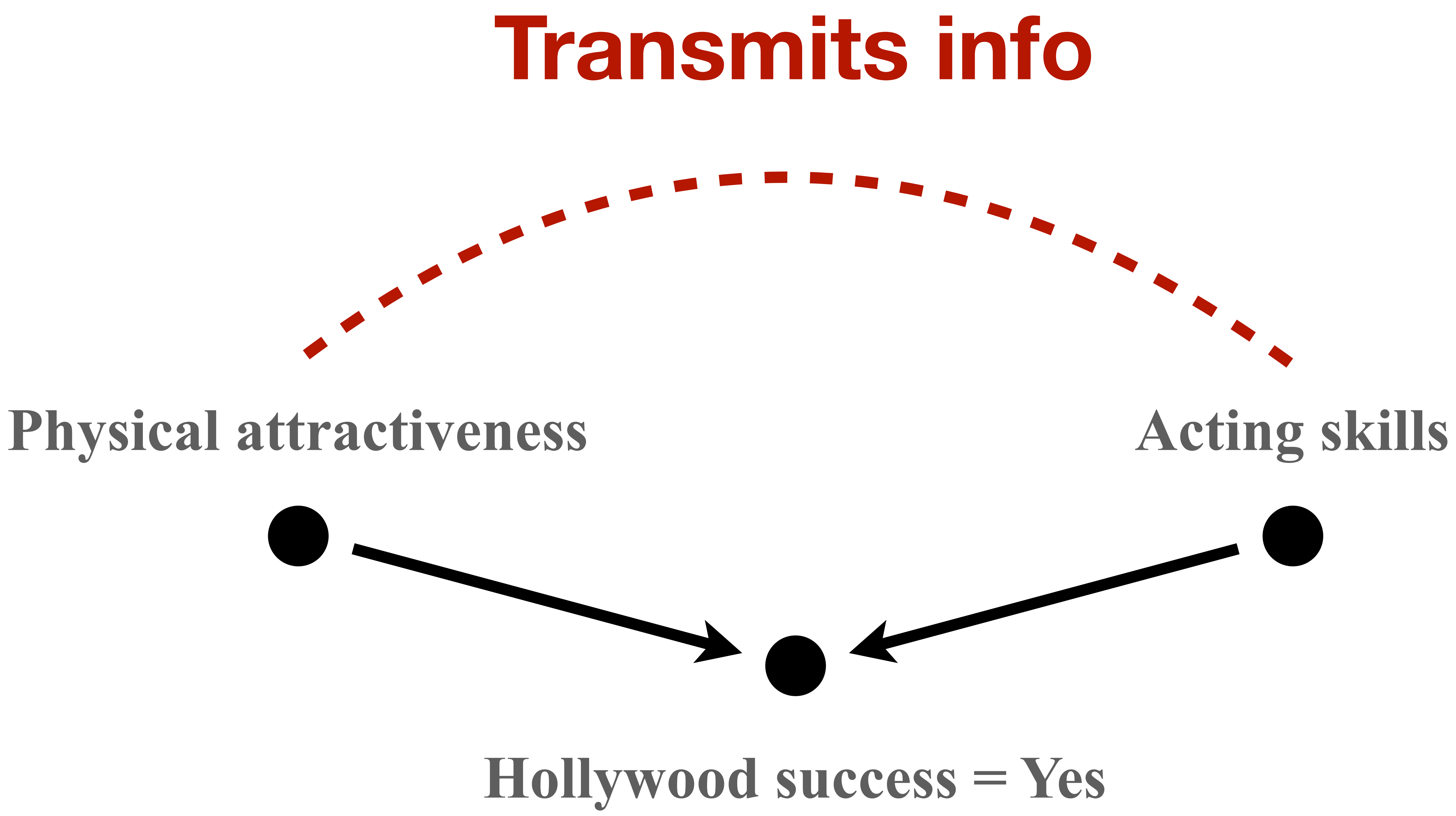

Acting skill vs. attractiveness
among Hollywood stars
Acting skill (0–10)
0
2
4
6
8
10
0
2
4
6
8
10
Attractiveness (0–10)

Transmits info



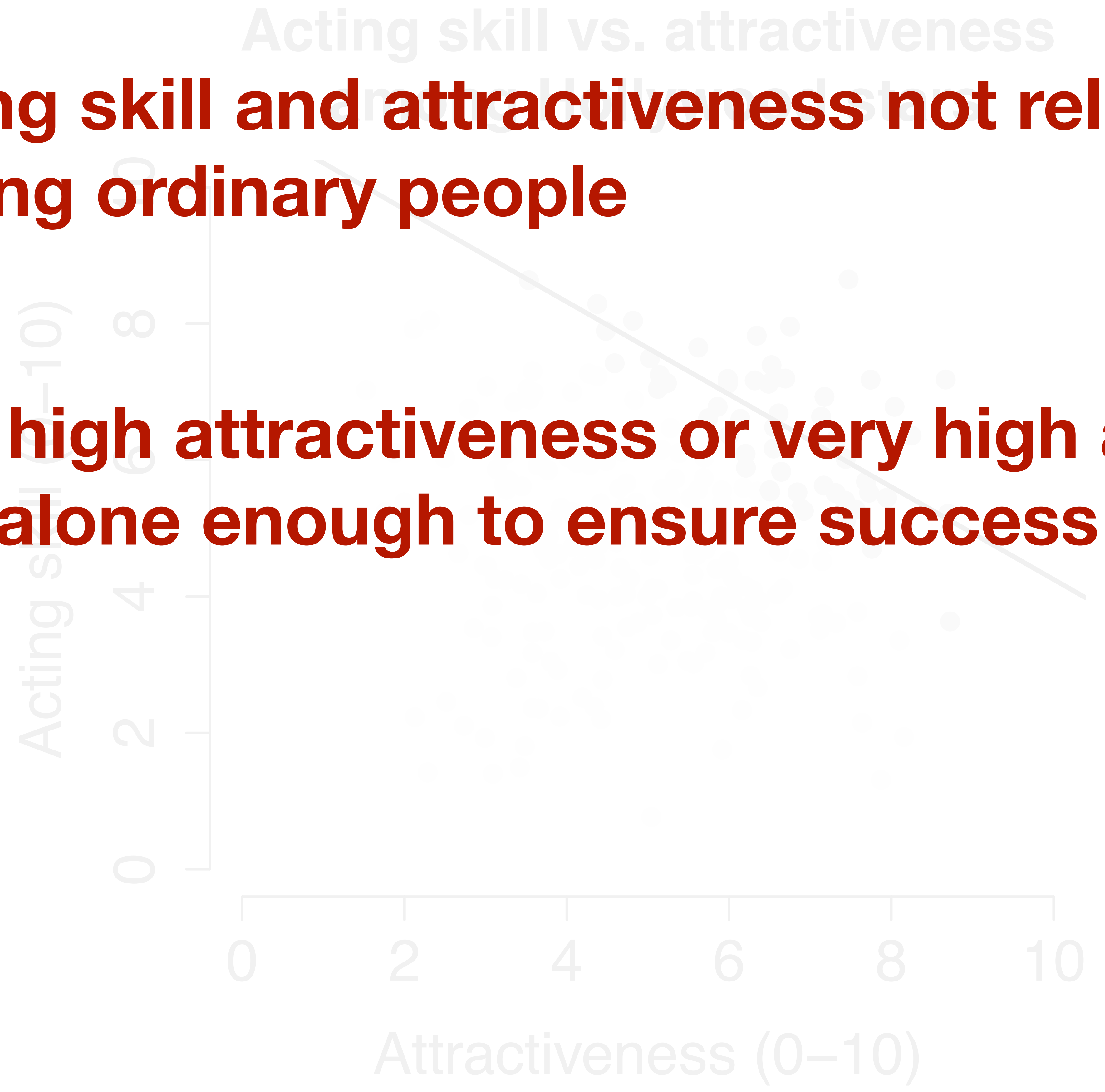
Acting skill and attractiveness not related among ordinary people

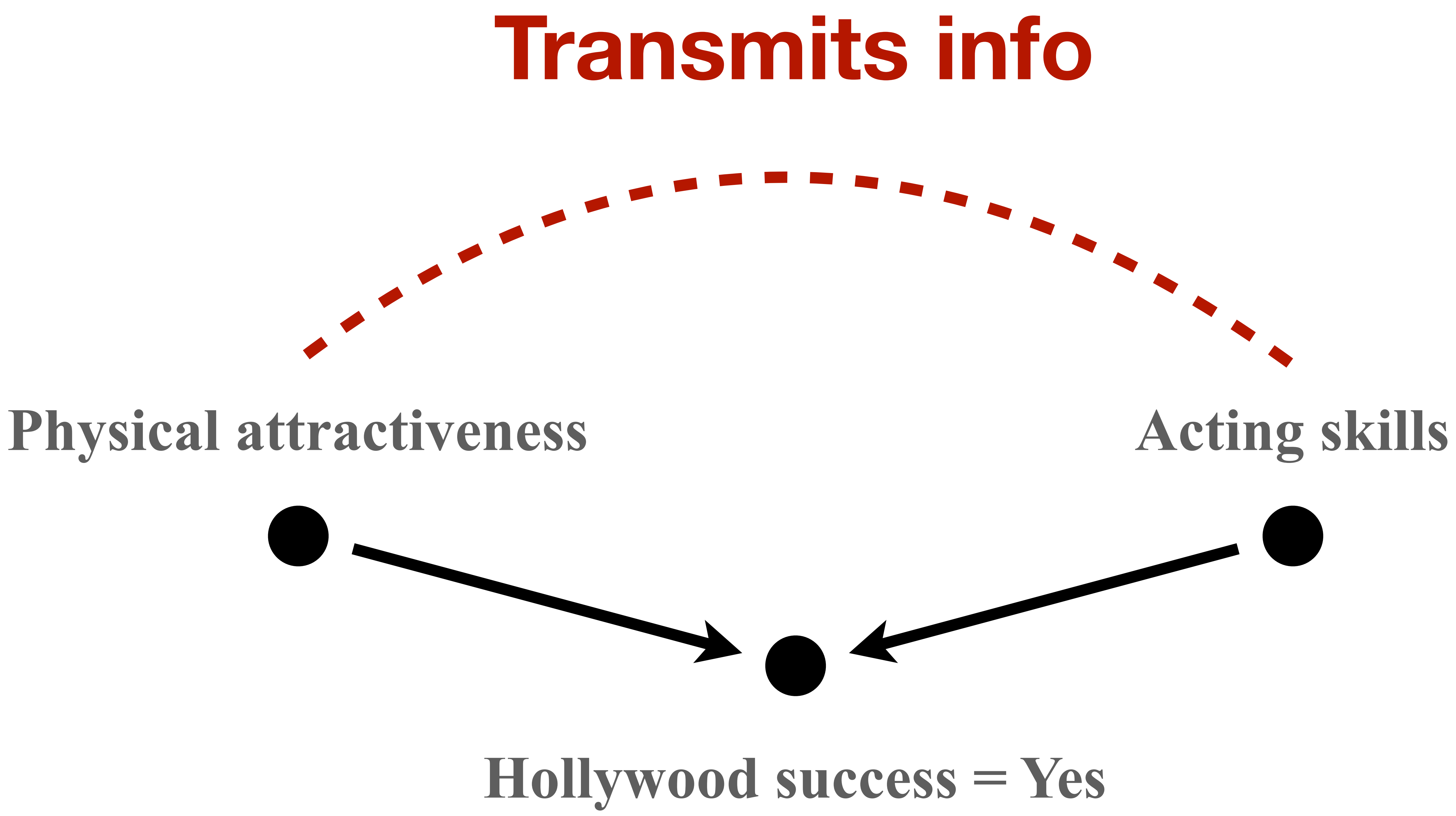




Acting skill and attractiveness not related among ordinary people

Very high attractiveness or very high acting skill alone enough to ensure success





Acting skill and attractiveness not related among ordinary people

Very high attractiveness or very high acting skill alone enough to ensure success

Conditioning on collider called “collider bias:” introduces an effect where there should be none

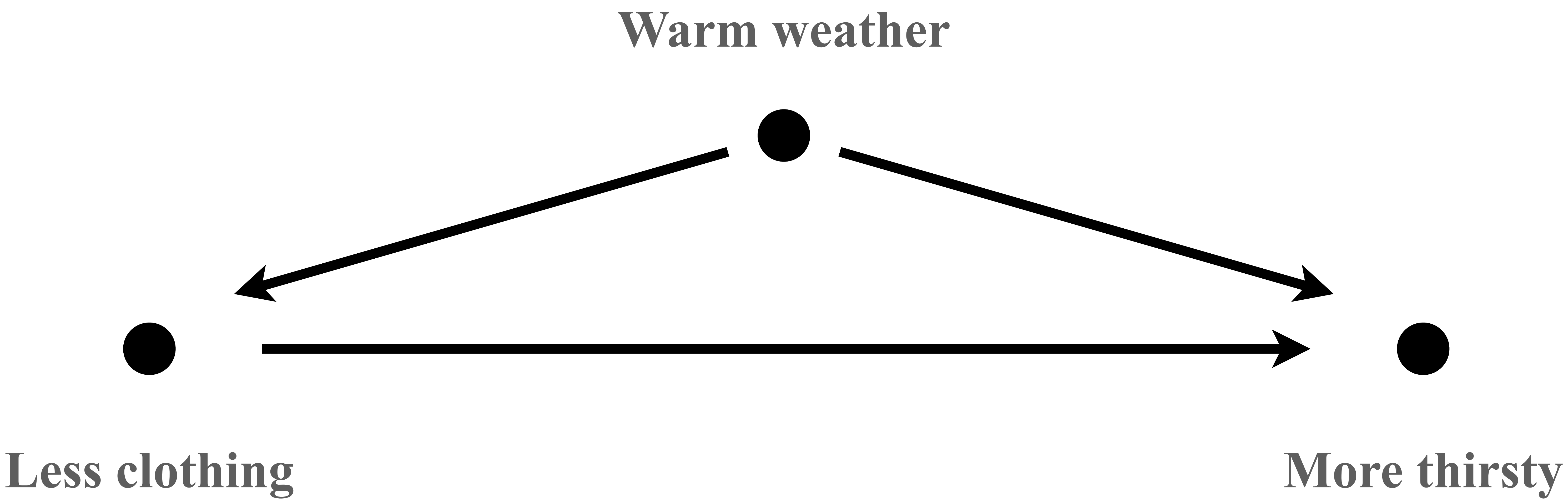
break probably

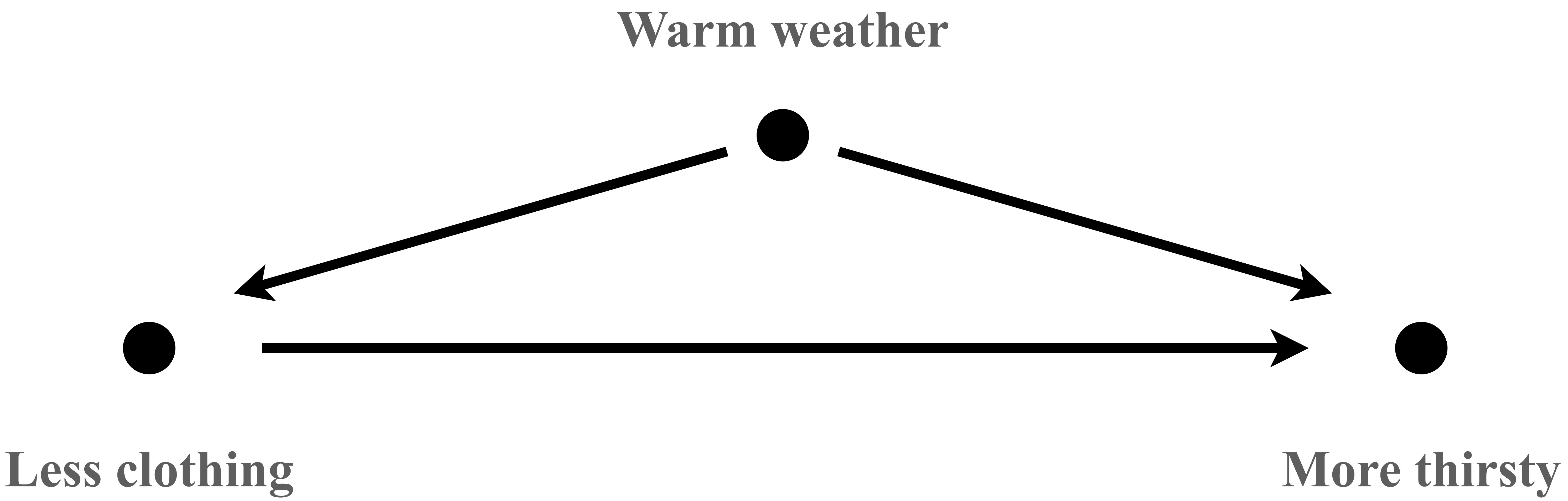
The Marko fallacy: confounding

Lessons so far

- The three basic junctures are like “information pipes”
- The chain and the fork are open, information flows through them
- The collider is closed, information does not flow through
- Conditioning on the middle node of a chain or fork closes the pipe
- Conditioning on the middle node of a collider opens the pipe

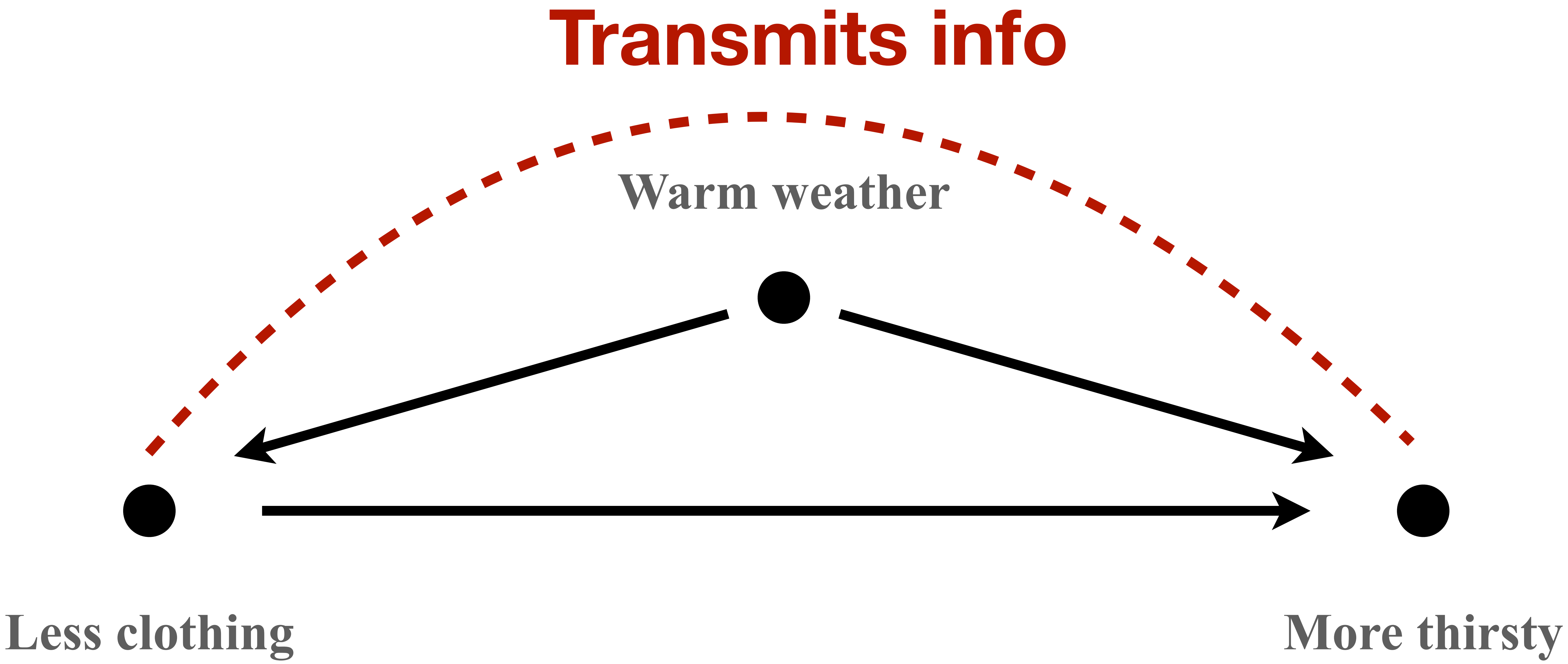
Please suggest a more sane model by applying one of the basic junctions to this mess

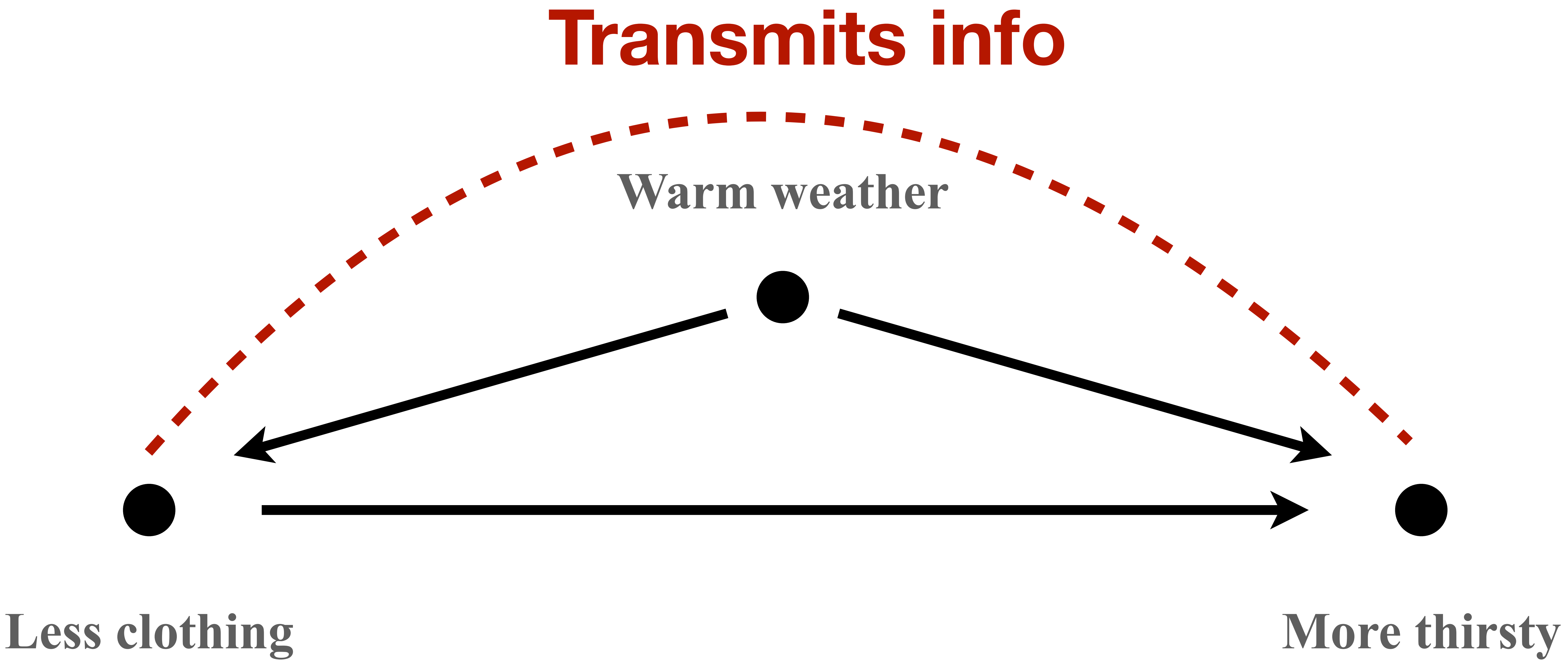

Warm weather
Less clothing
More thirsty



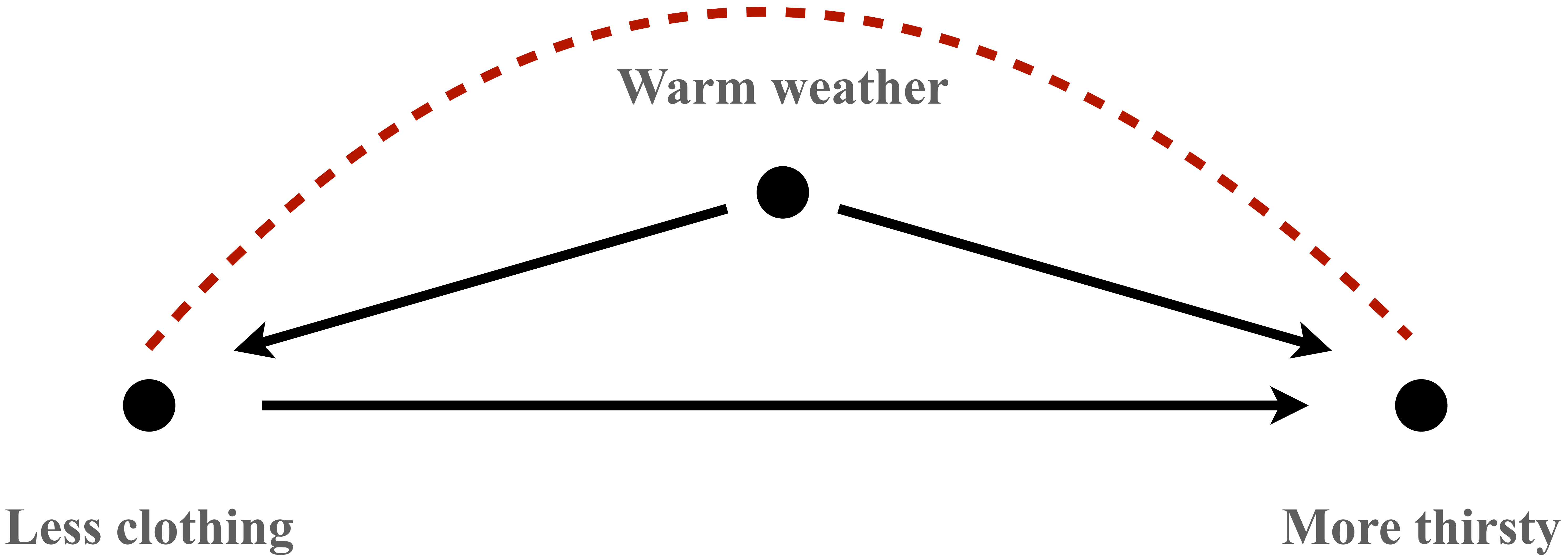
Note: leaving an arrow in is a weaker assumption than removing it.

assuming an effect might be anything (including 0)
VS assuming that it is 0 exactly

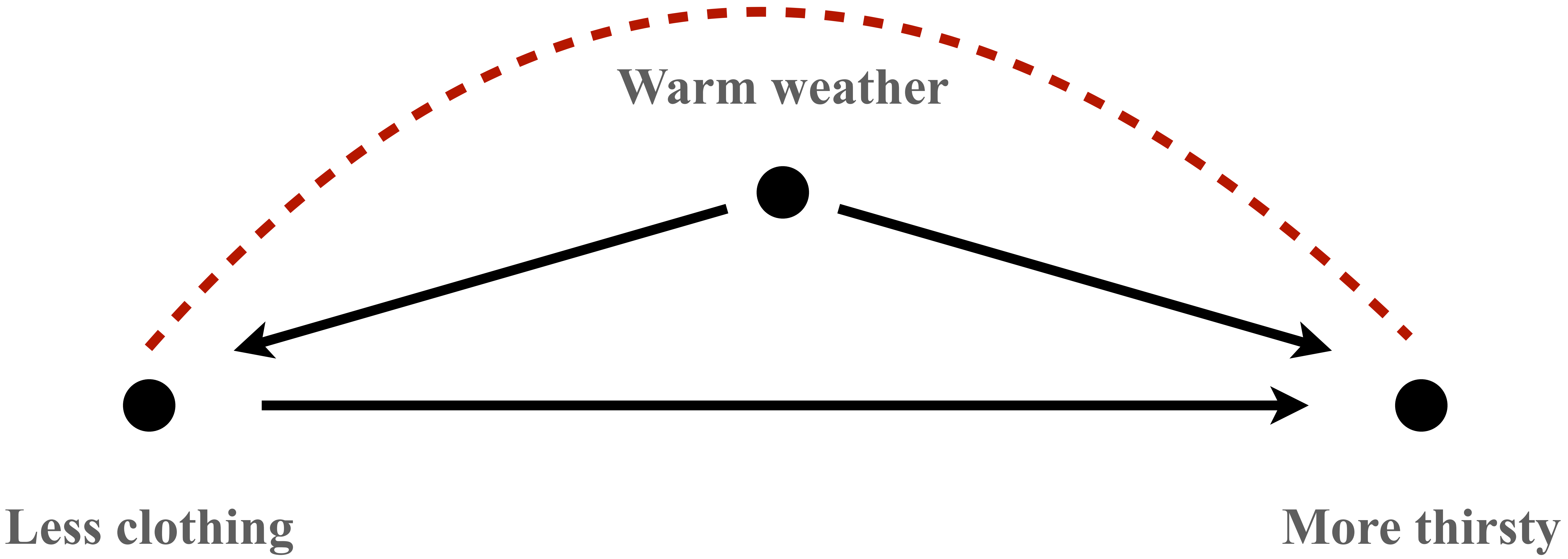

Transmits info
Warm weather
Less clothing
More thirsty



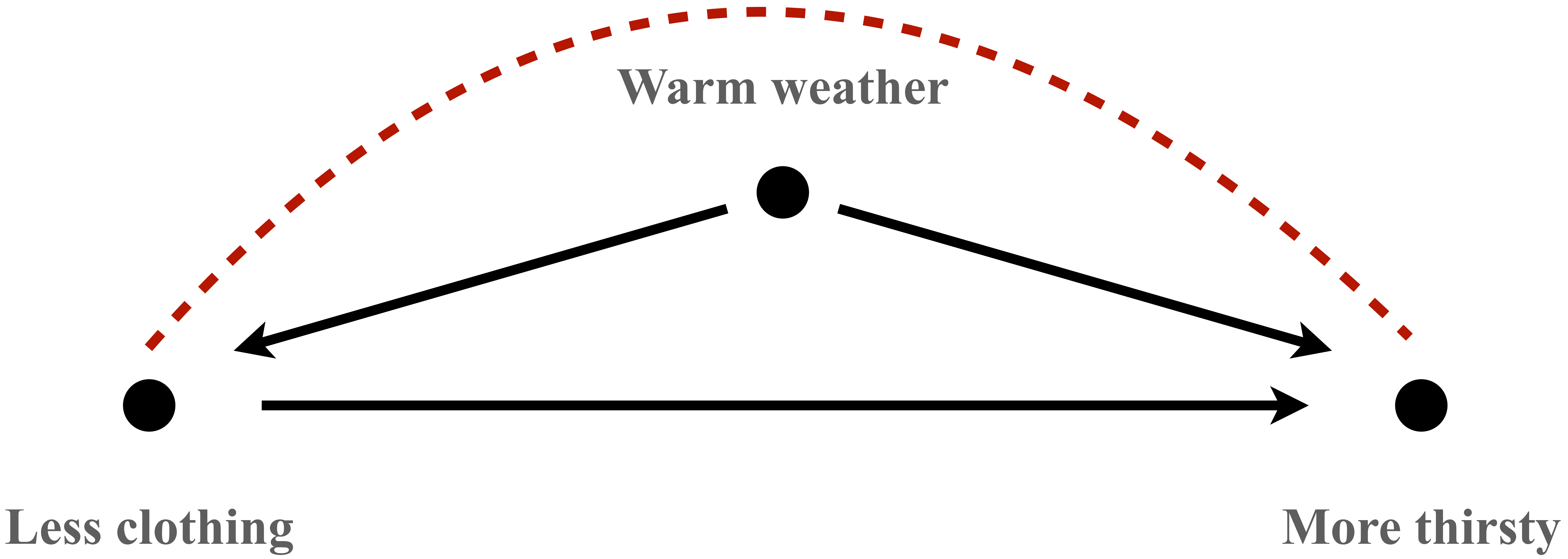
to confound: to mix or confuse



Confounding shows up as an open “back-door path” between exposure and outcome:

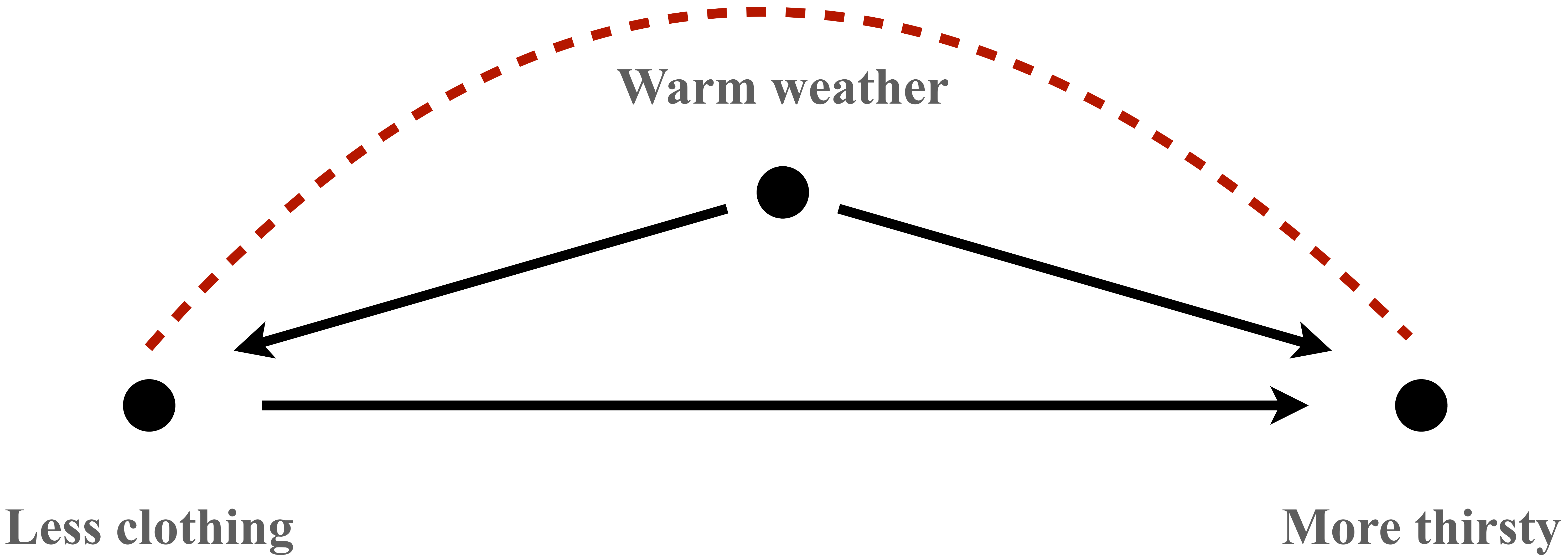


Confounding shows up as an open “back-door path” between exposure and outcome:



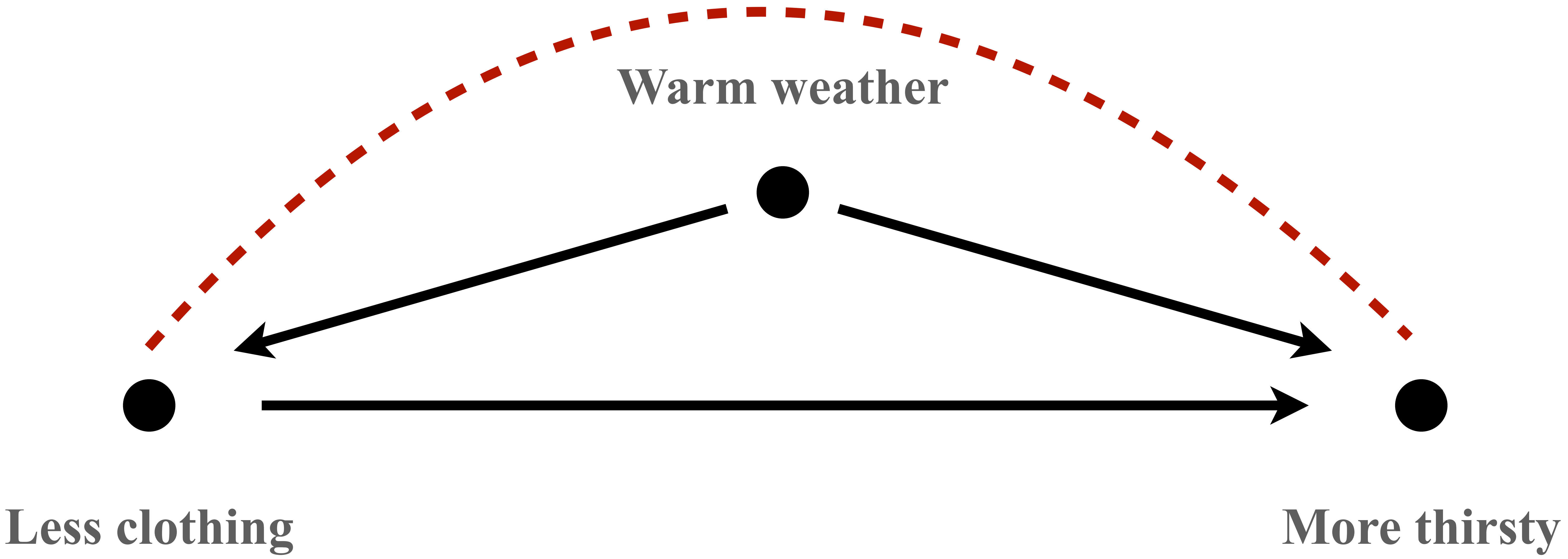
Confounding shows up as an open “back-door path” between exposure and outcome:

- **open:** no collider along the path



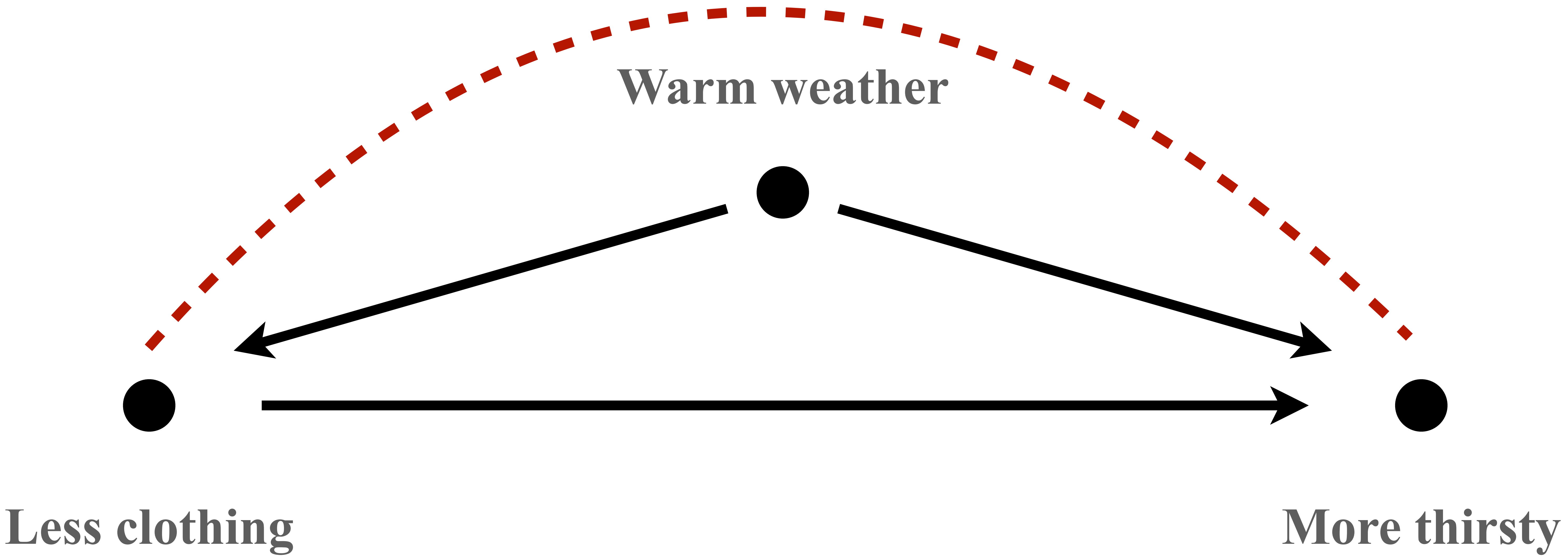
Confounding shows up as an open “back-door path” between exposure and outcome:

- **open:** no collider along the path
- **back-door:** an arrow goes *into* the exposure

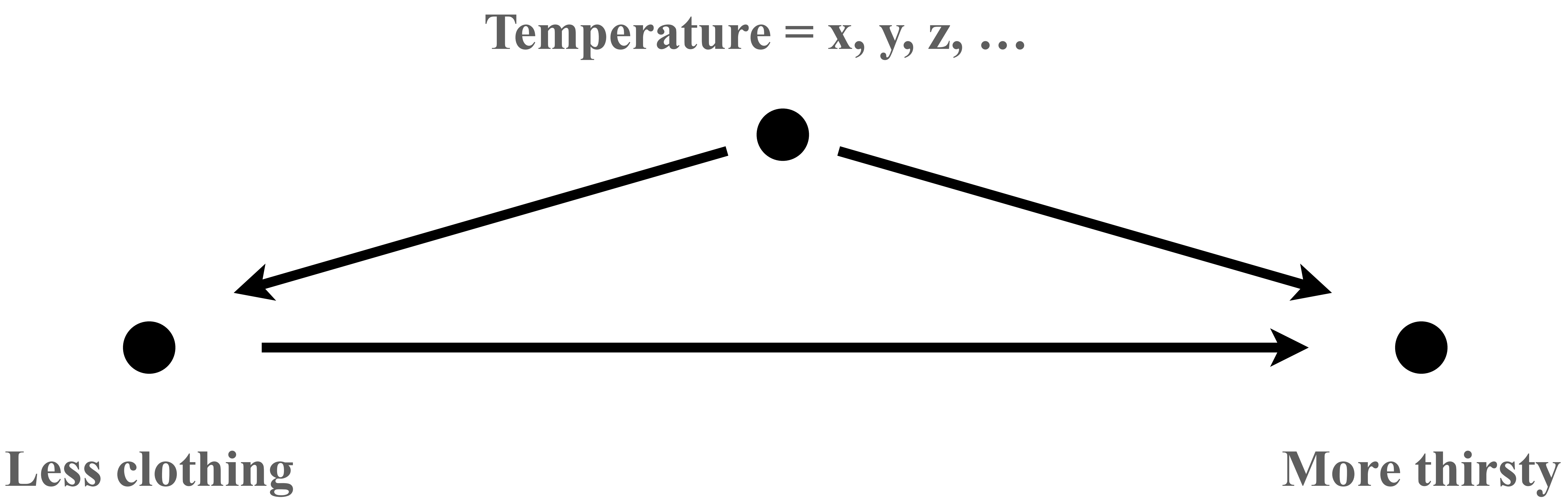


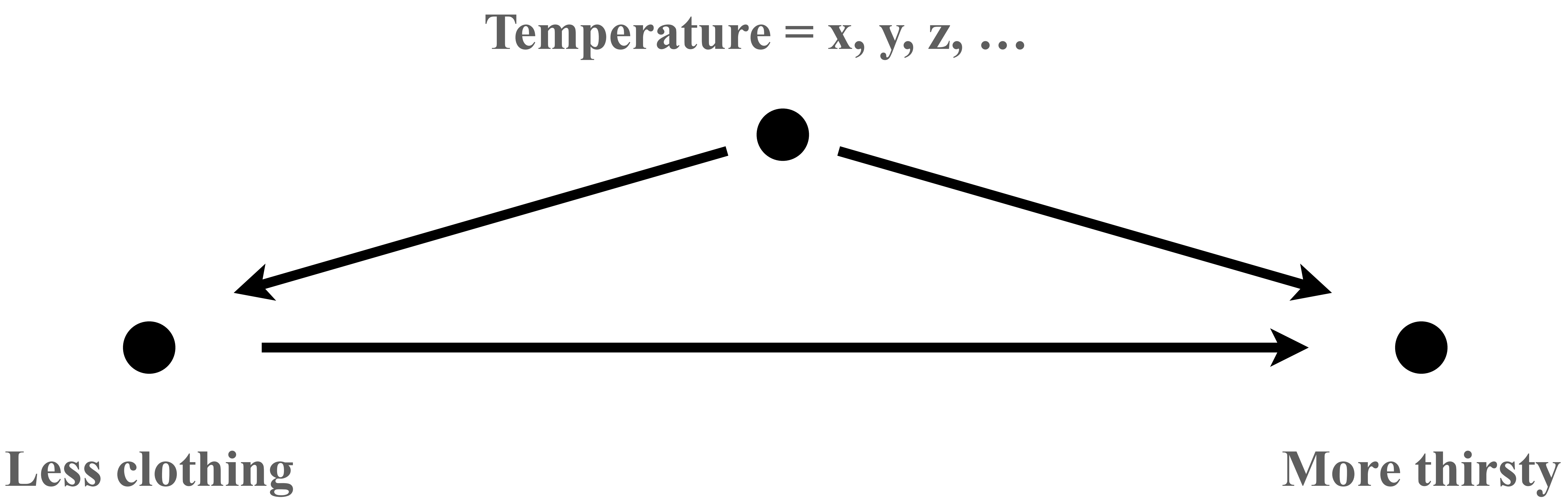
Confounding shows up as an open "back-door path" between exposure and outcome:

- **open:** no collider along the path
- **back-door:** an arrow goes *into* the exposure
- **The problem:** there is a mixing of the presumed causal relation along the direct path and the purely associational relation through the back-door path



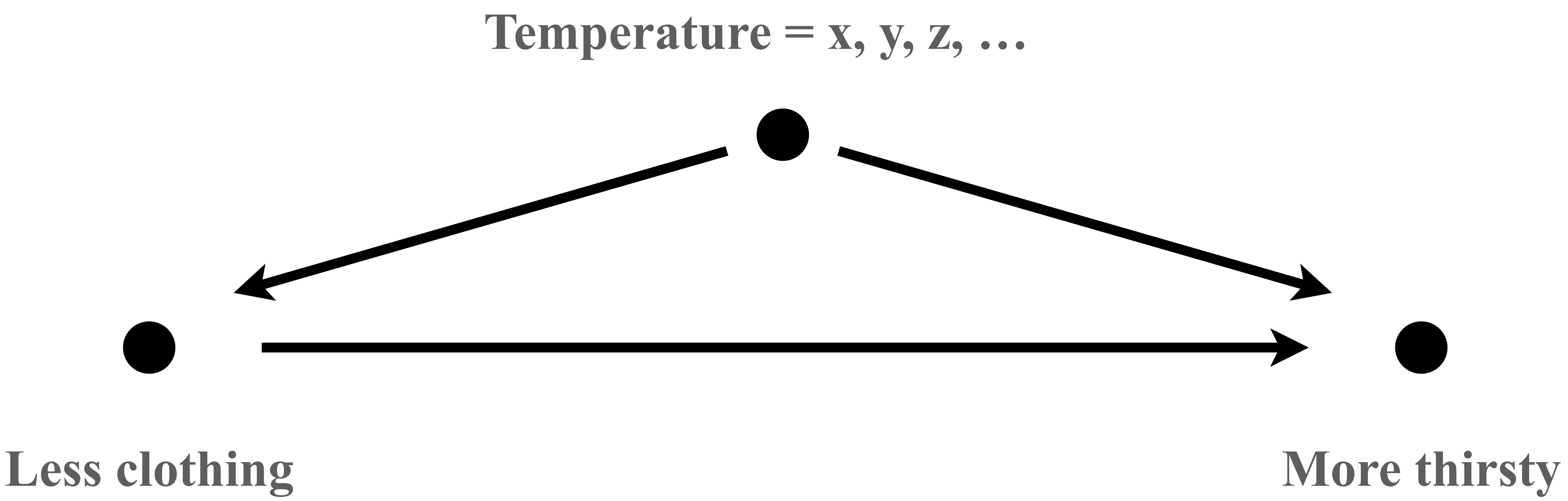
Please to tell me what I can do about this situation.

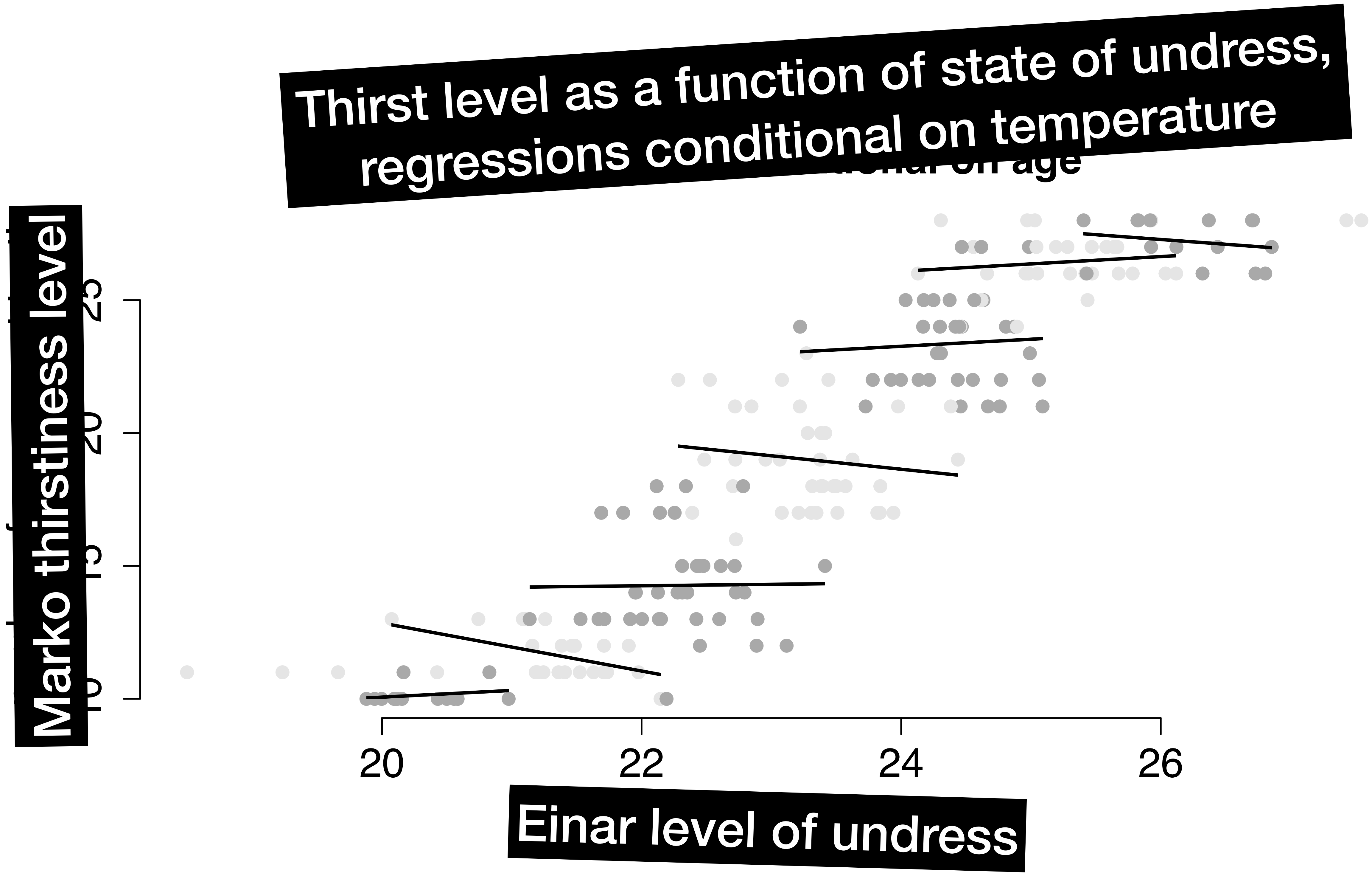

Temperature = x, y, z, …
Less clothing
More thirsty



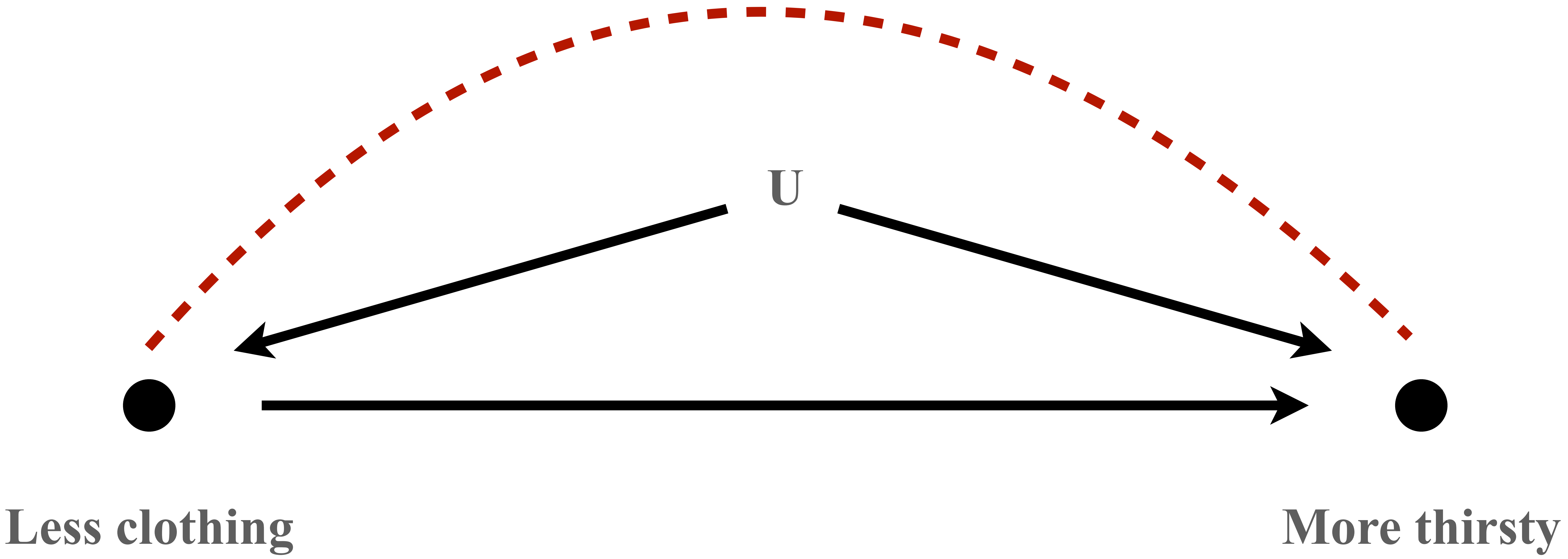
Temperature = x, y, z, …
Less clothing
More thirsty

teeth vs shoe size,
regressions conditional on age
number of permanent teeth
10
15
20
25
20
22
24
26
shoe size (cm)

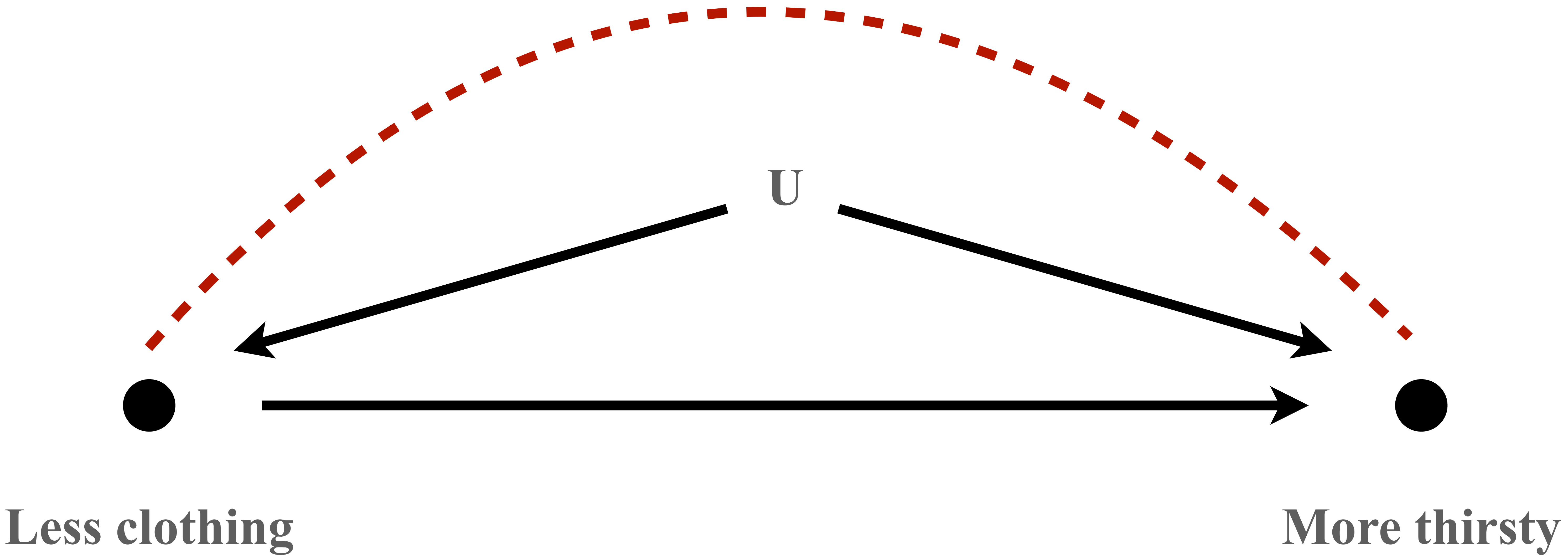

Temperature = x, y, z, …
Less clothing
More thirsty


Thirst level as a function of state of undress,
regressions conditional on temperature
Marko thirstiness level
20
22
24
26
Einar level of undress

Comparing 25 degree days with one another: comparing like with like

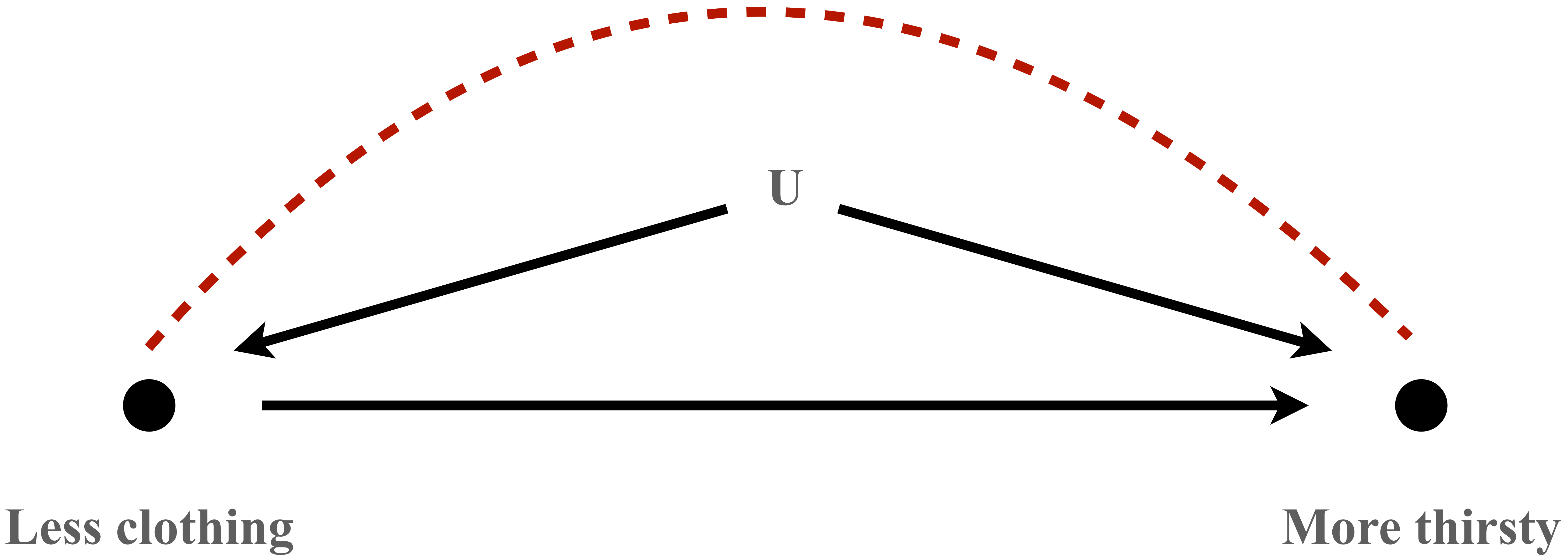


Temperature UNOBSERVED: what do we do???



Temperature UNOBSERVED: what do we do???

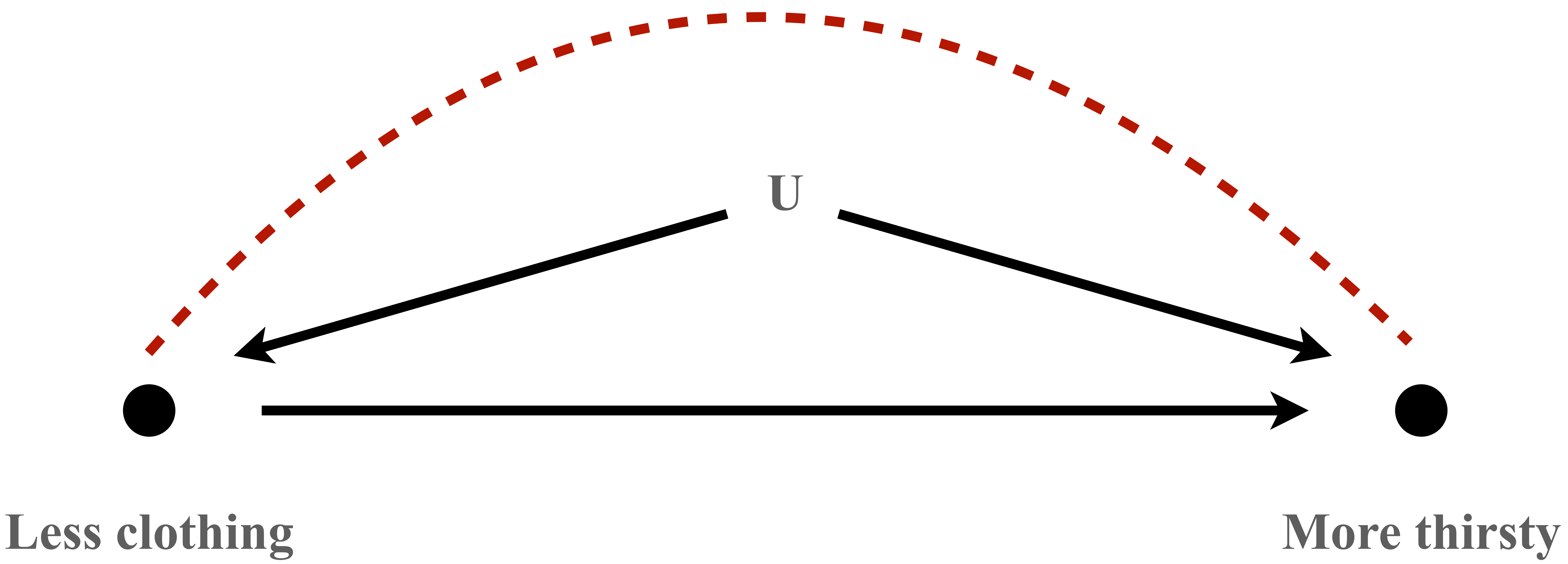
Not much to do: the effect is non-identifiable (get more data or do an actual experiment)



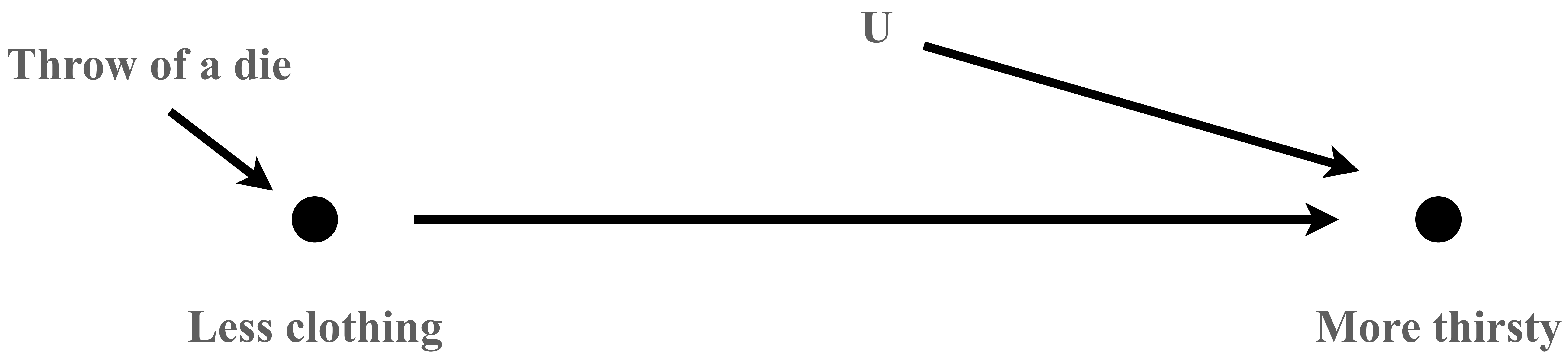
Temperature UNOBSERVED: what do we do???

Not much to do: the effect is **non-identifiable** (get more data or do an actual experiment)

An effect is **identifiable** if it is possible to close all back-door paths without opening new ones



Why randomization is so good

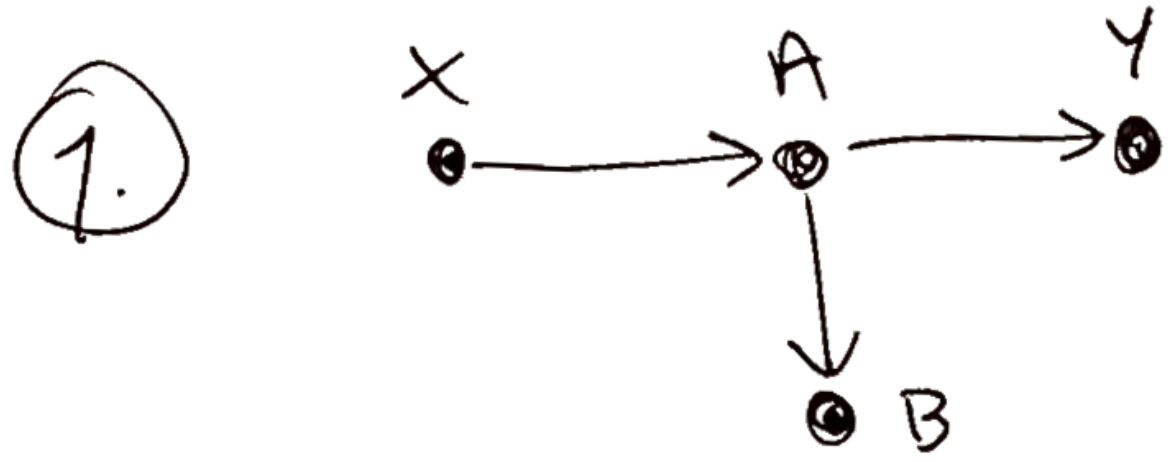


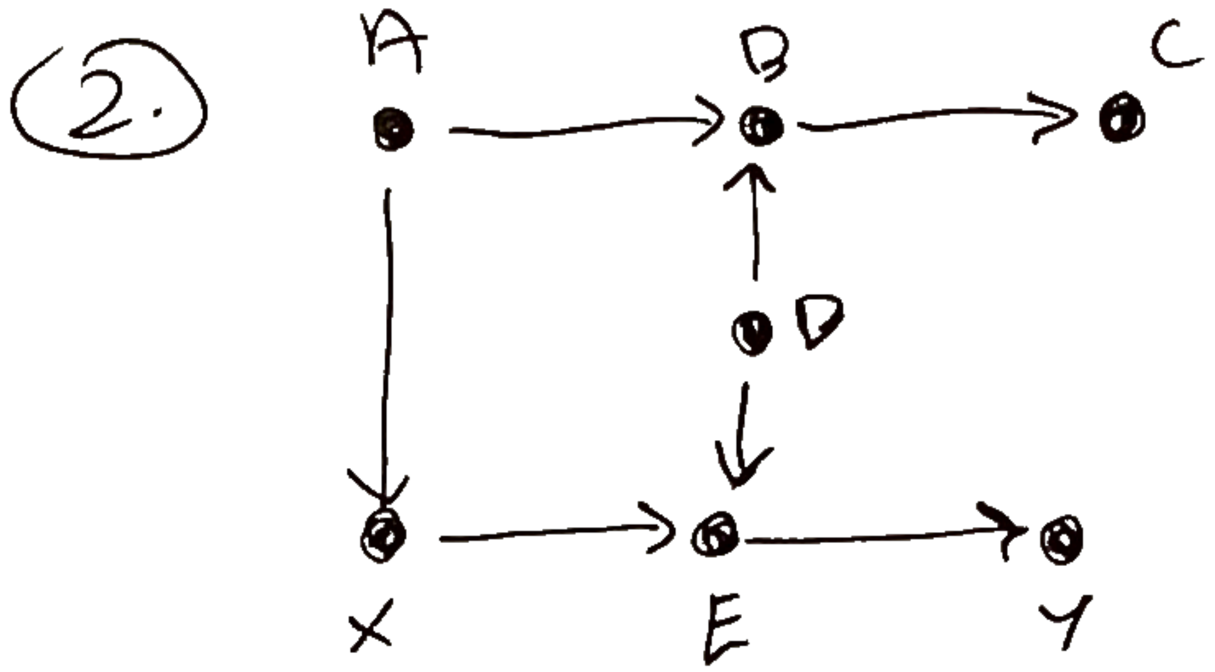
Why randomization is so good

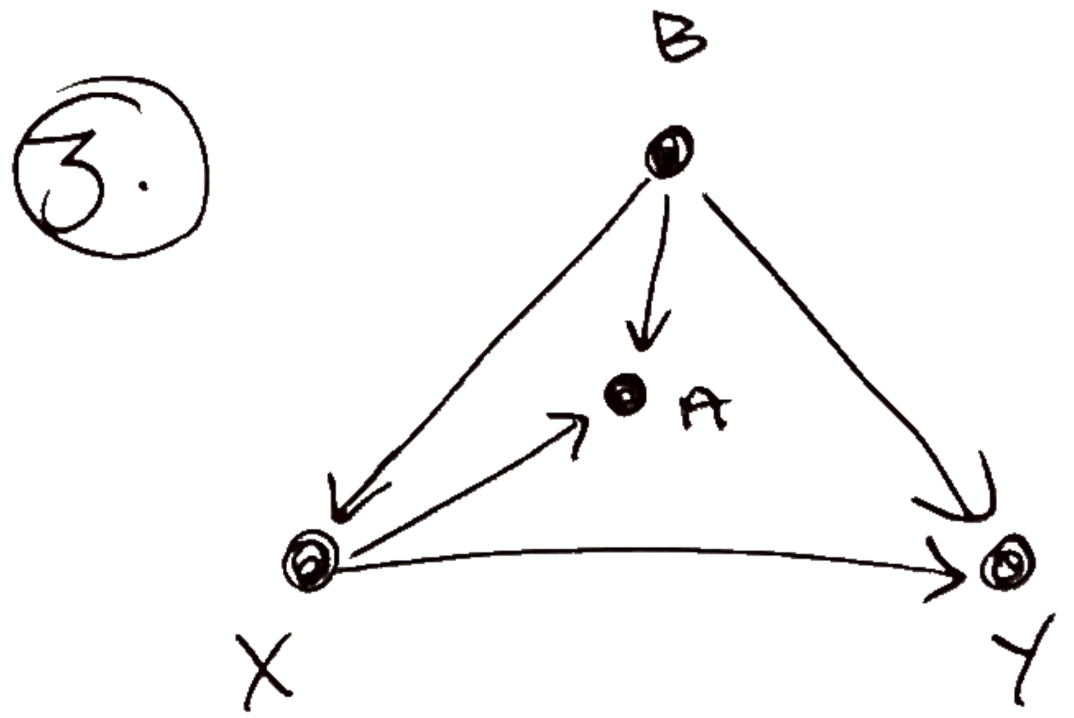
Point: For the price of making some causal assumptions, the rule about closing back-door paths tells us exactly what to adjust for

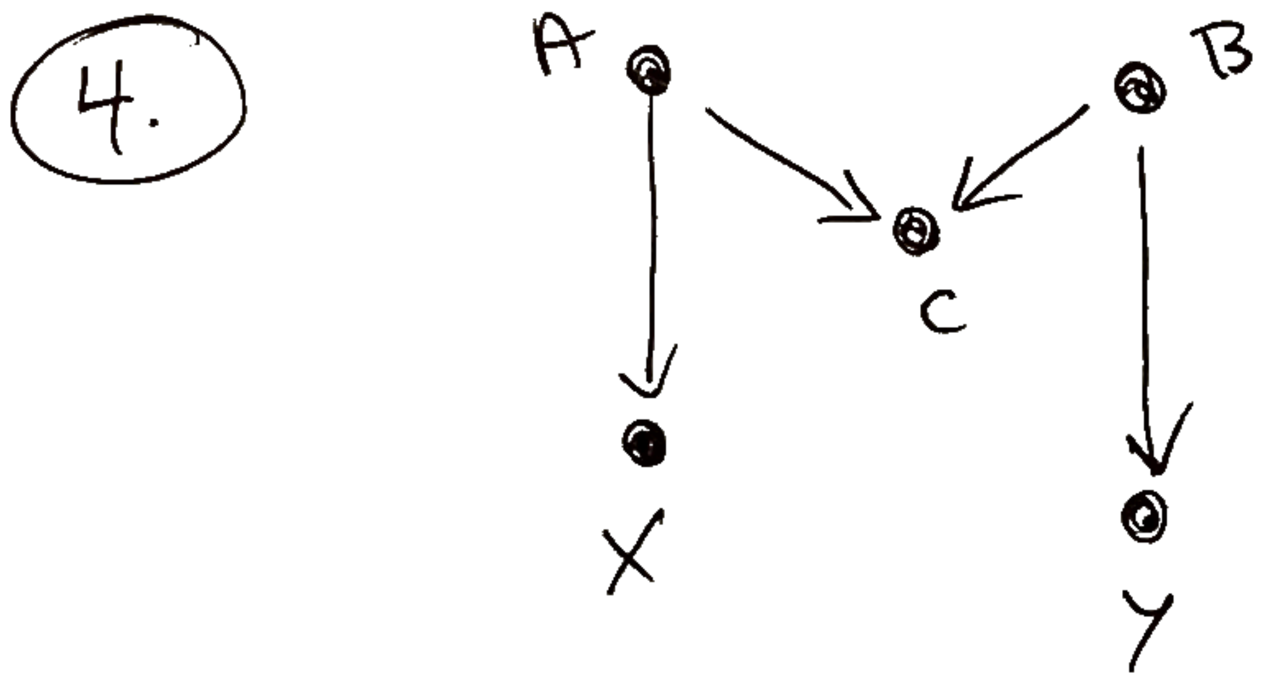
Point: For the price of making some causal assumptions, the rule about closing back-door paths tells us exactly what to adjust for

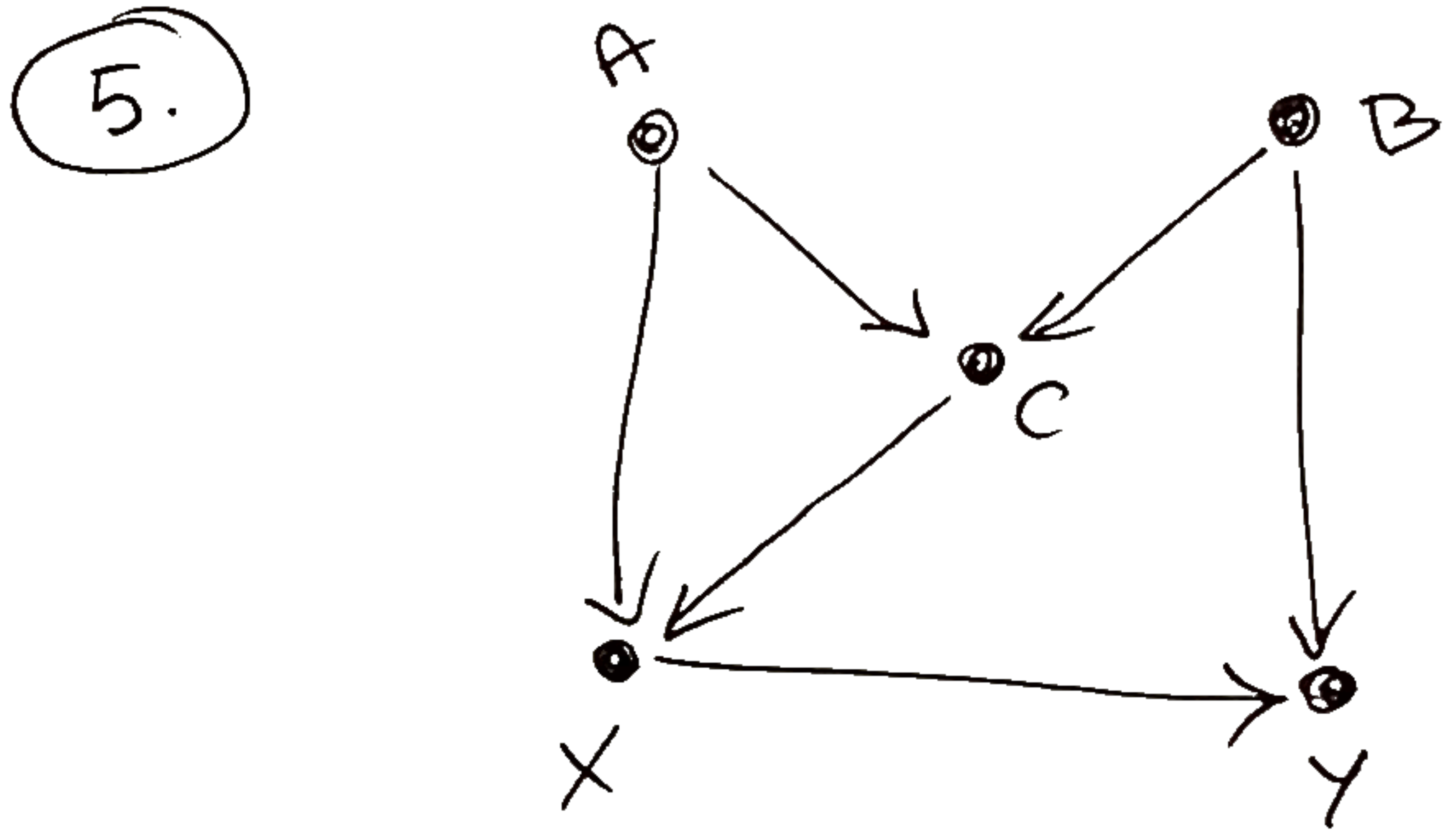
(Also you need to have made the right measurements)







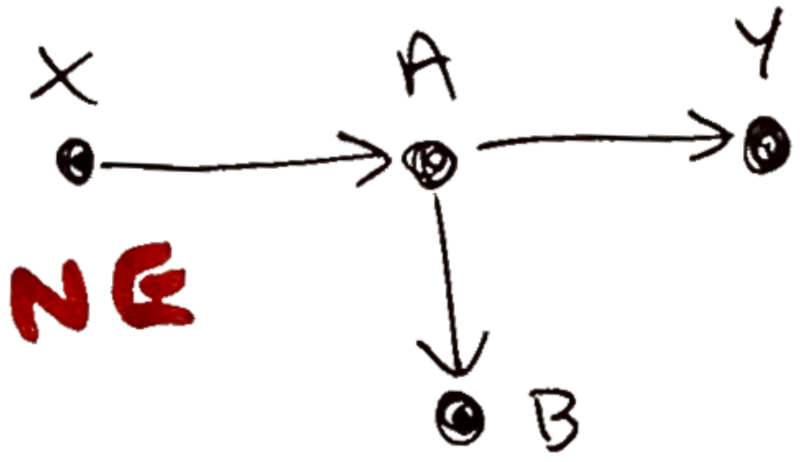




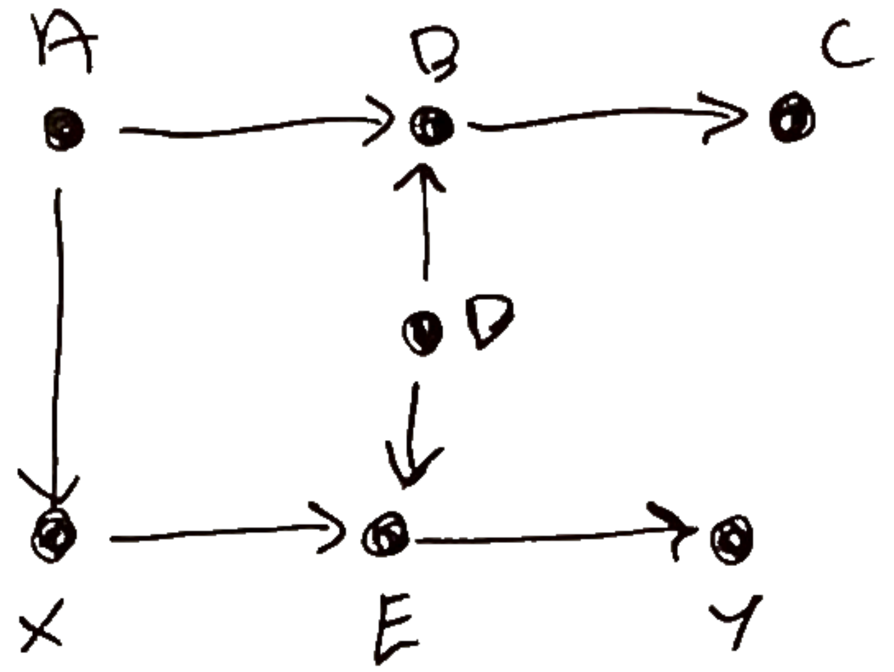
Puzzles: what should we adjust for to close the back-door paths between x and y?

1.

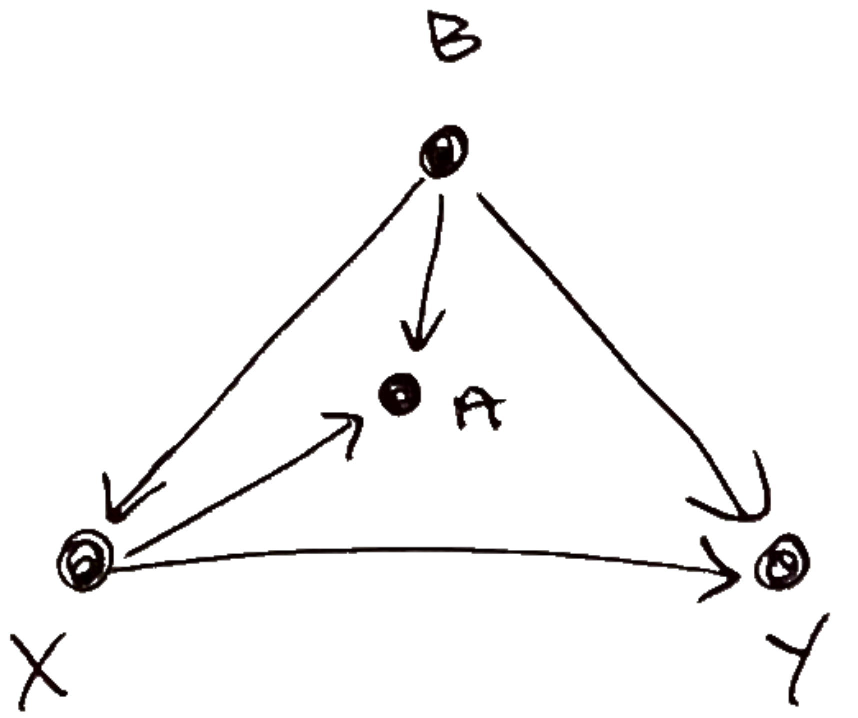
NONE



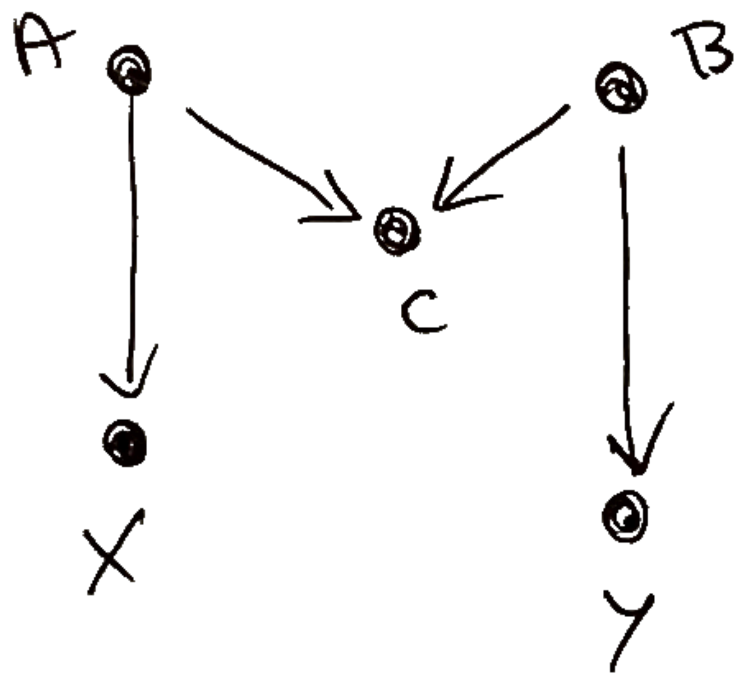
2.



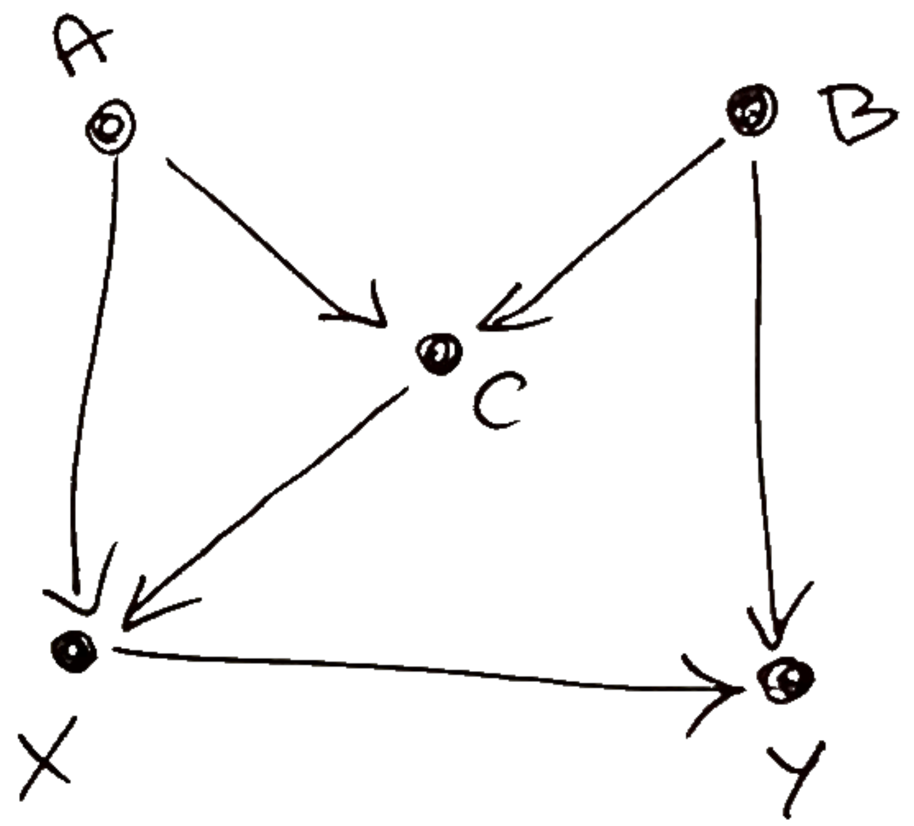
3.



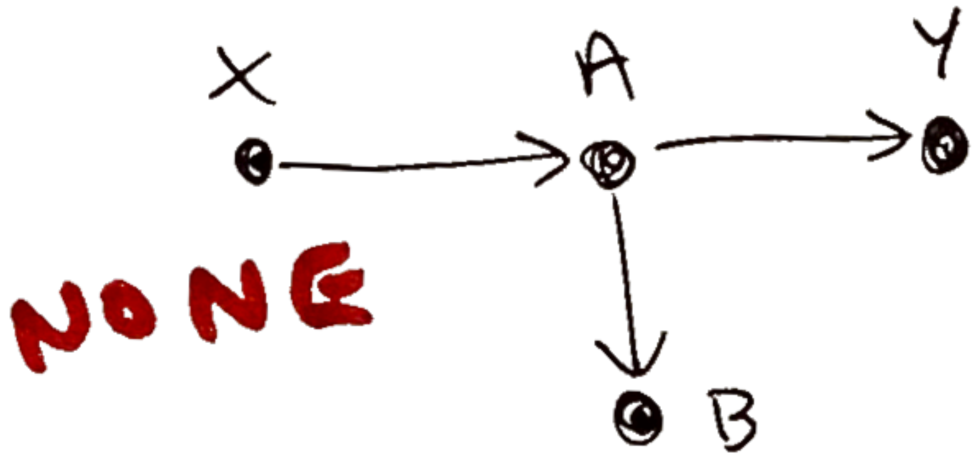
4.



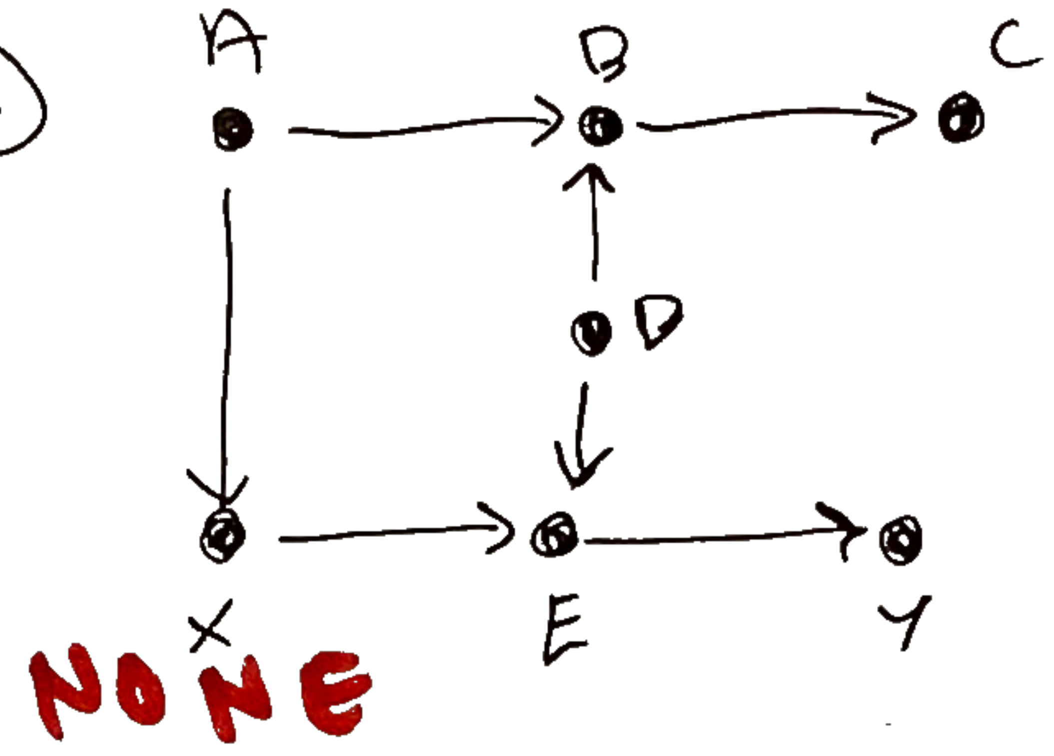
5.



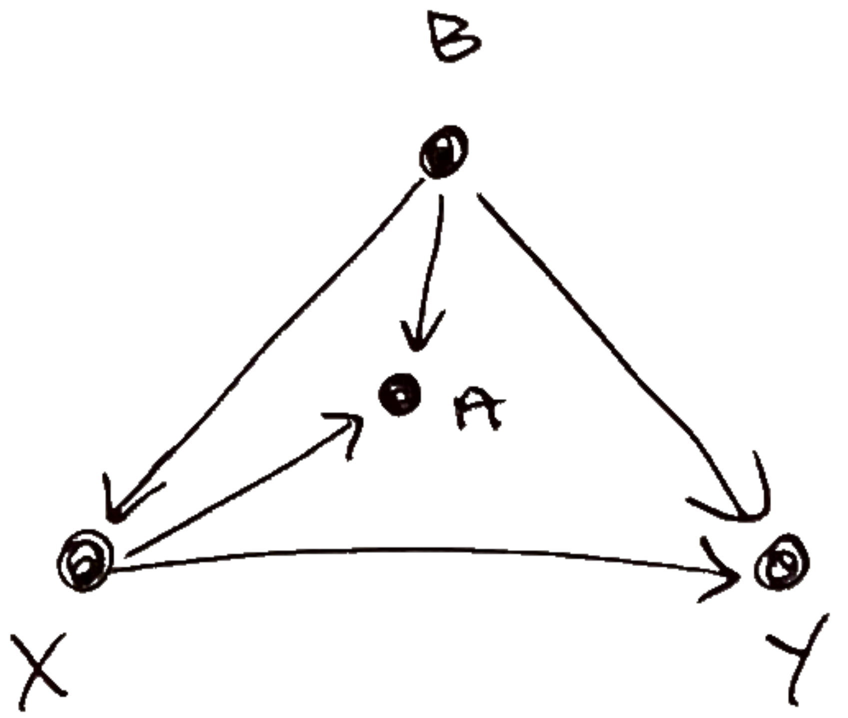
1.



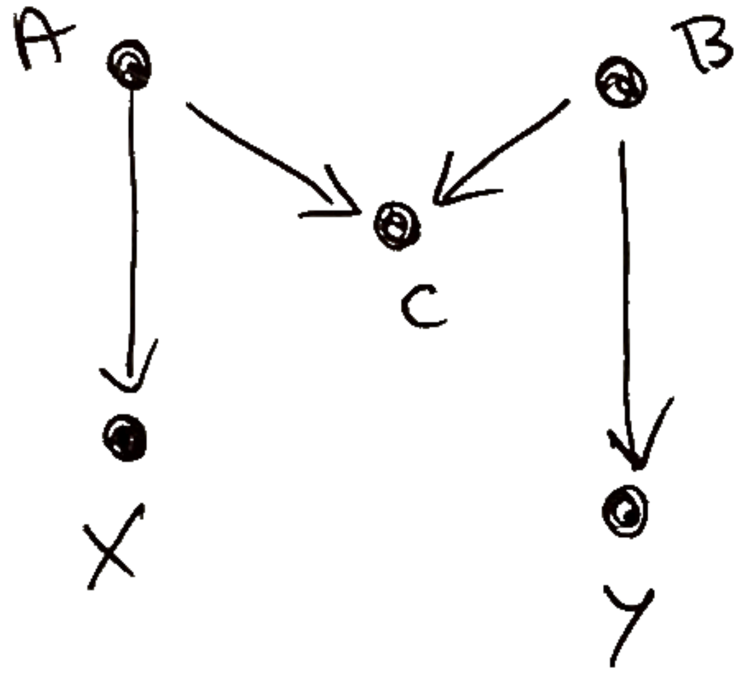
2.



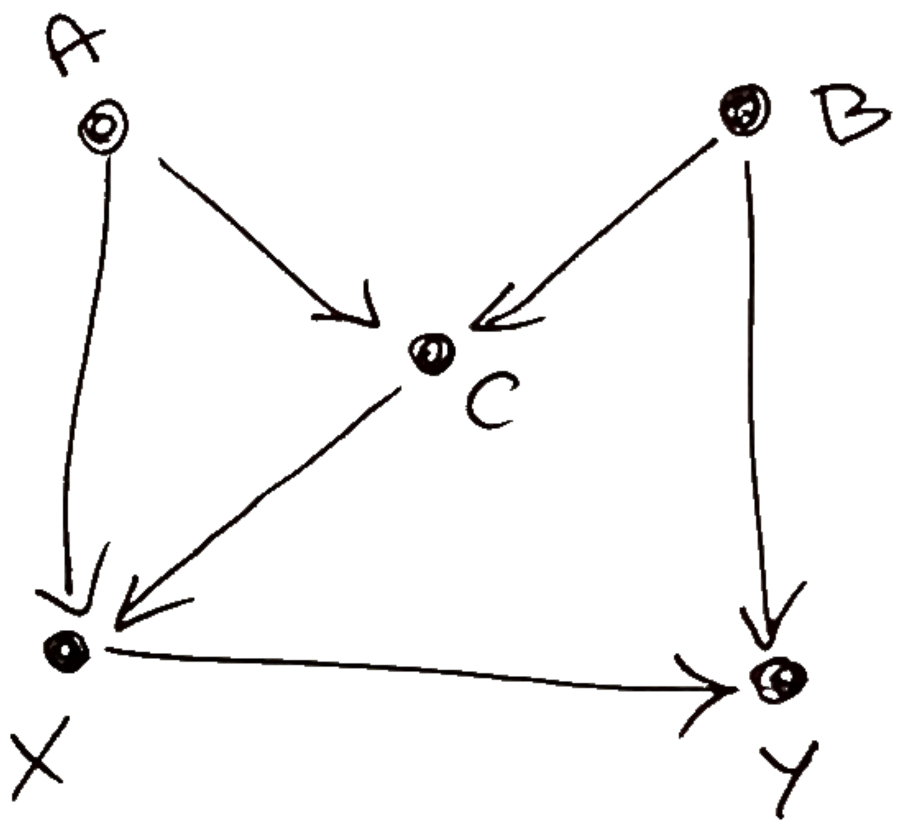
3.



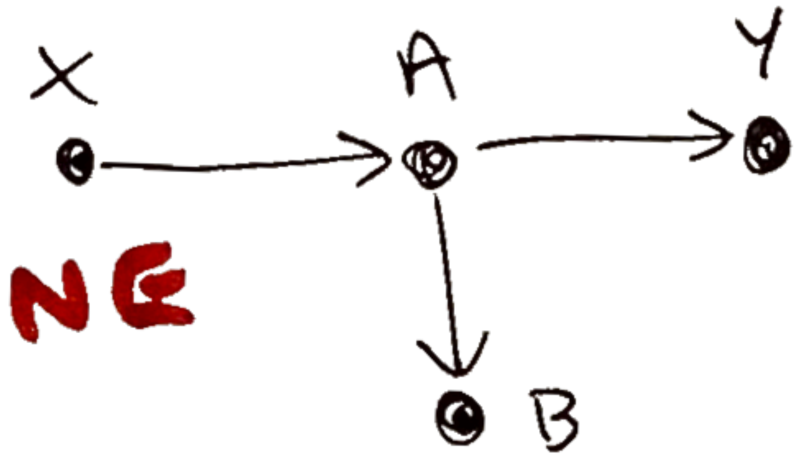
4.



5.

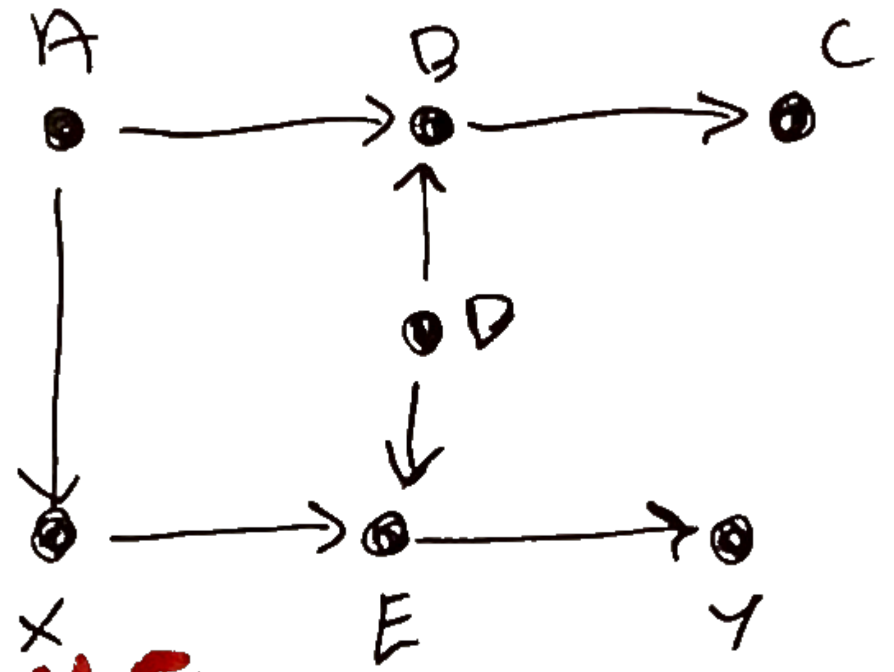


1.



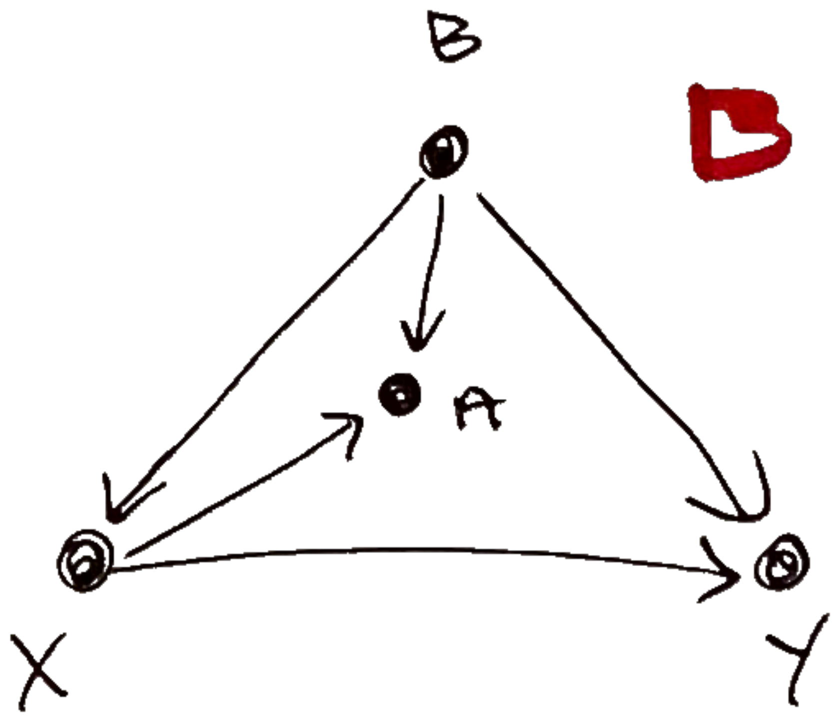
NONE

2.



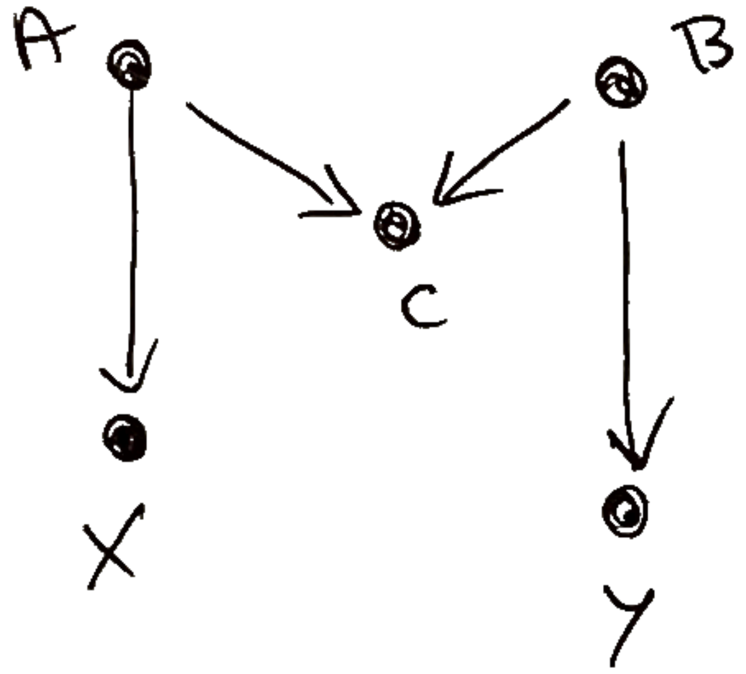
NONE

3.

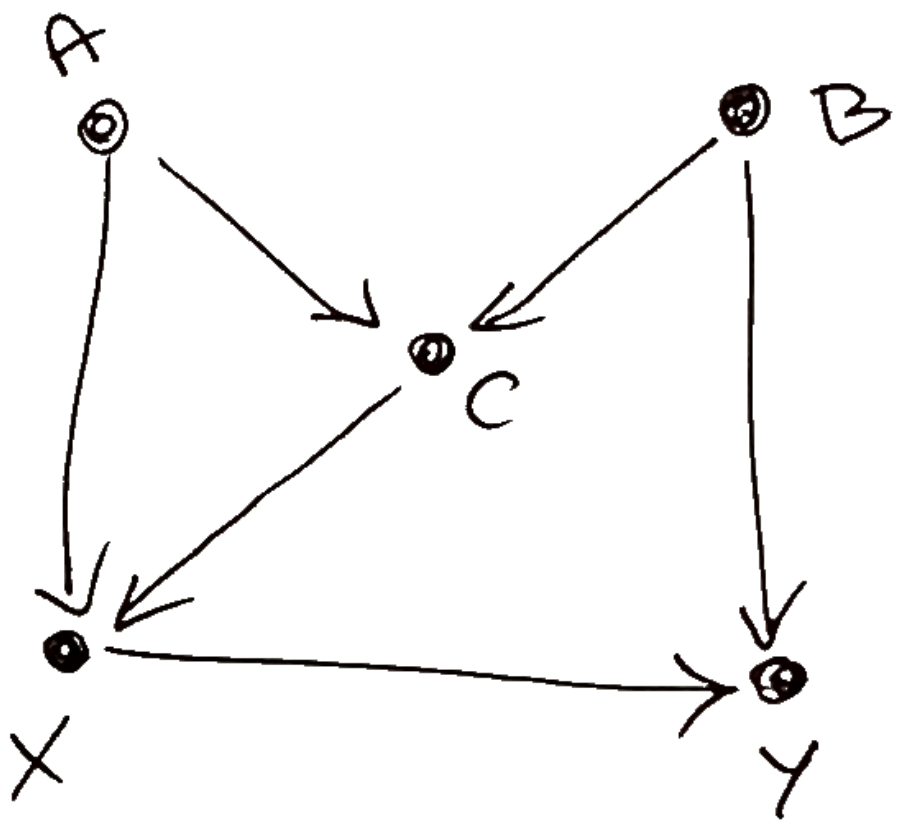


B

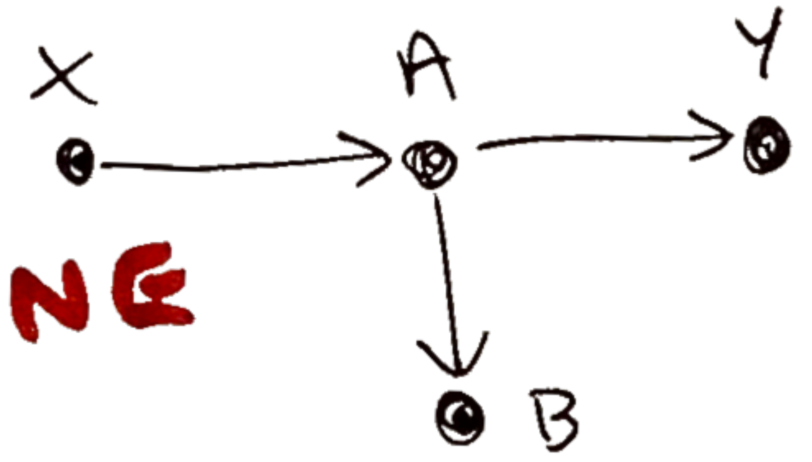
4.



5.

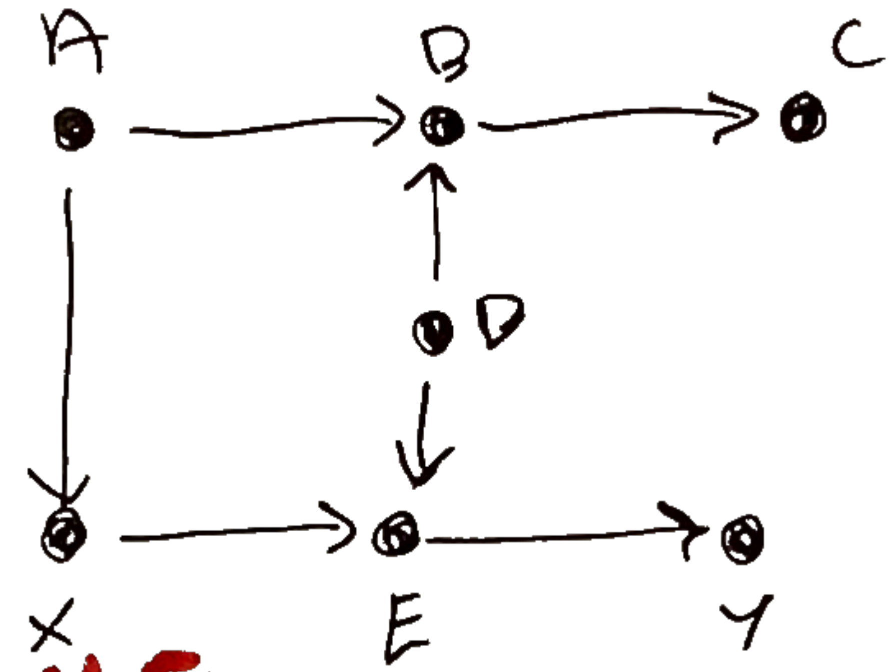


1.



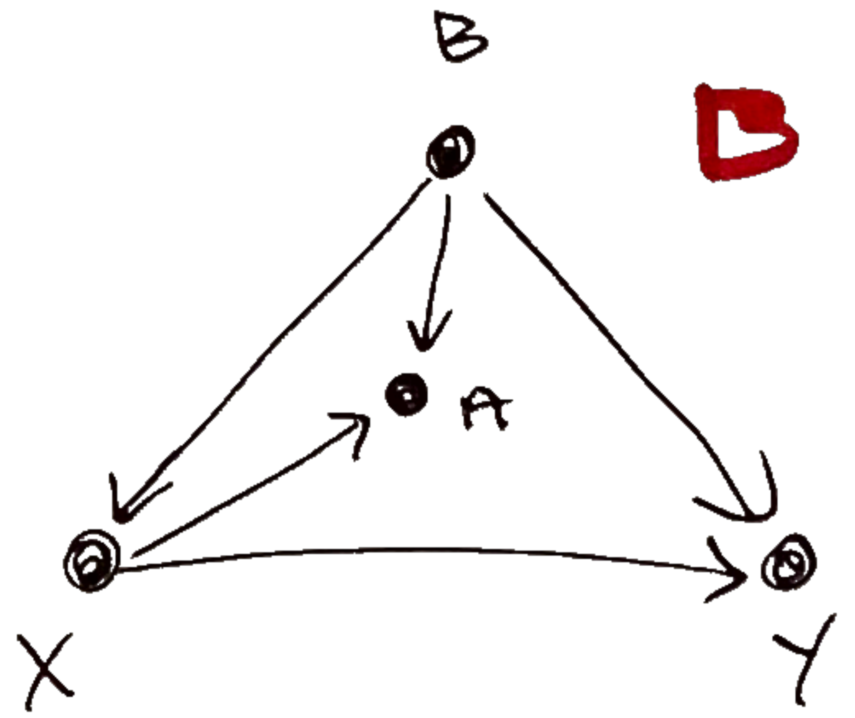
NONE

2.



NONE

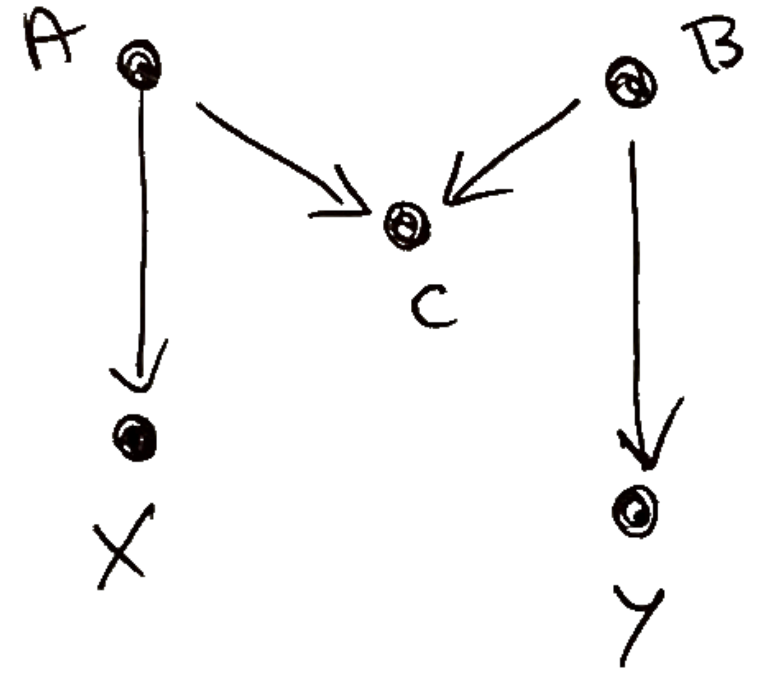
3.



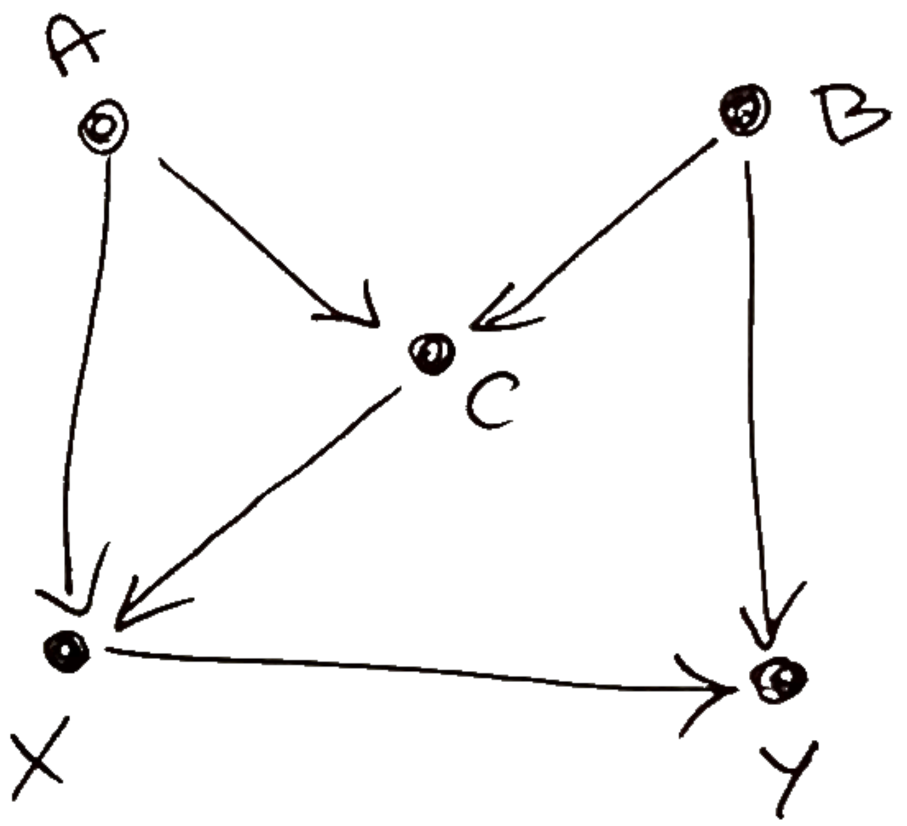
B

4.

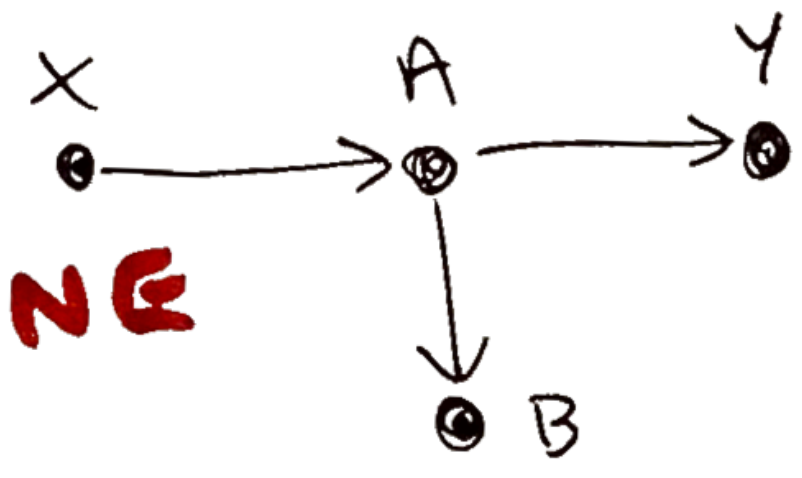
NONE



5.

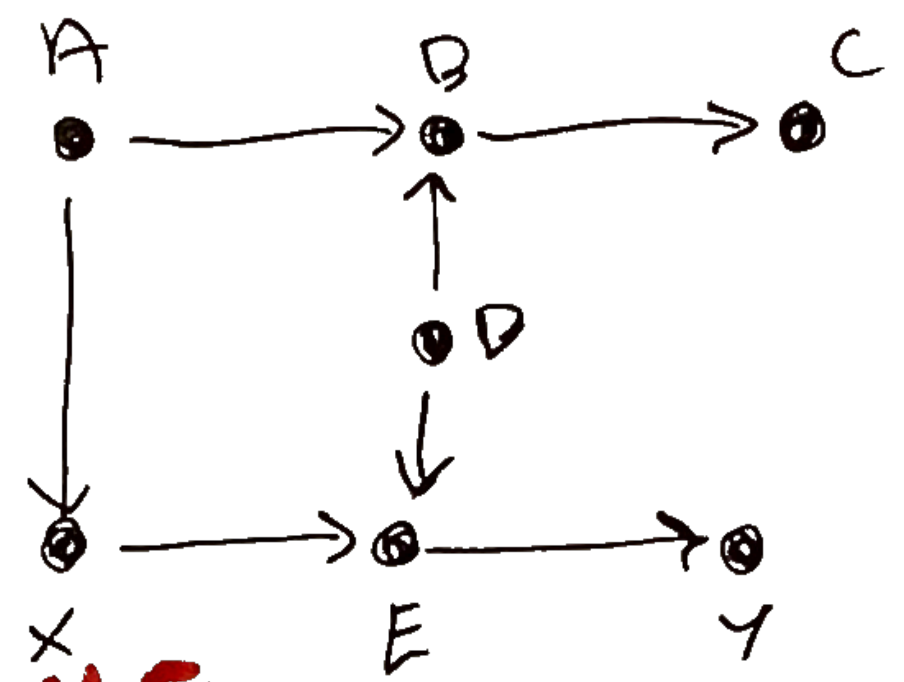


1.



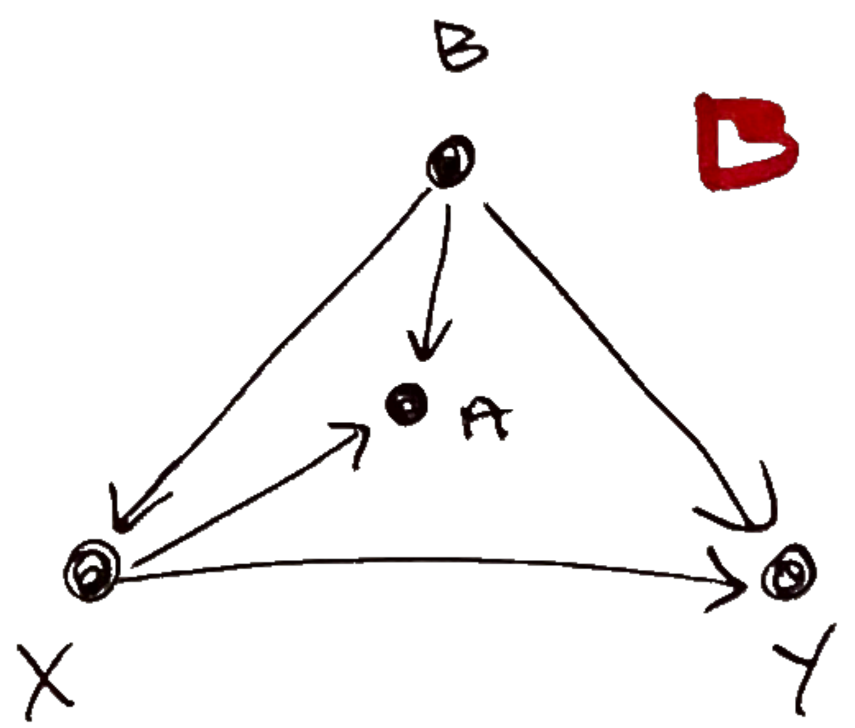
NONE

2.



NONE

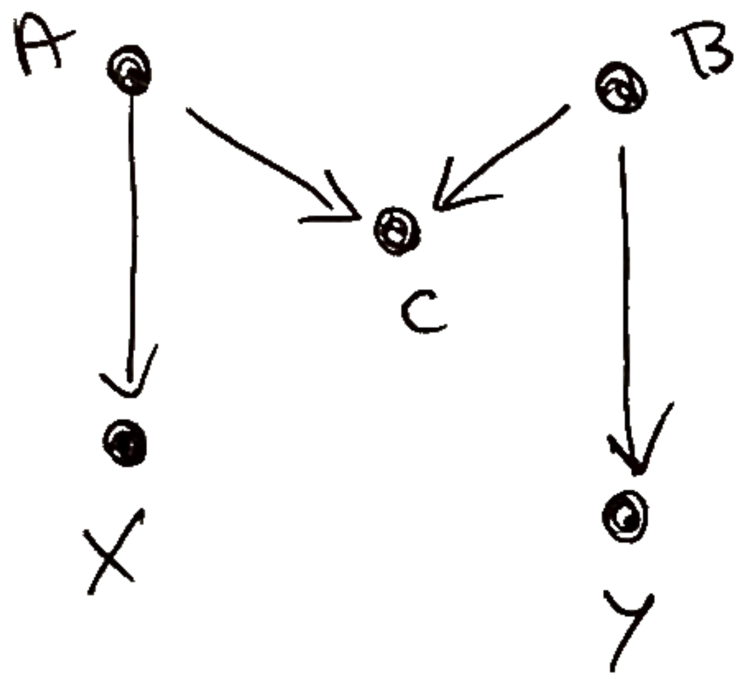
3.



B

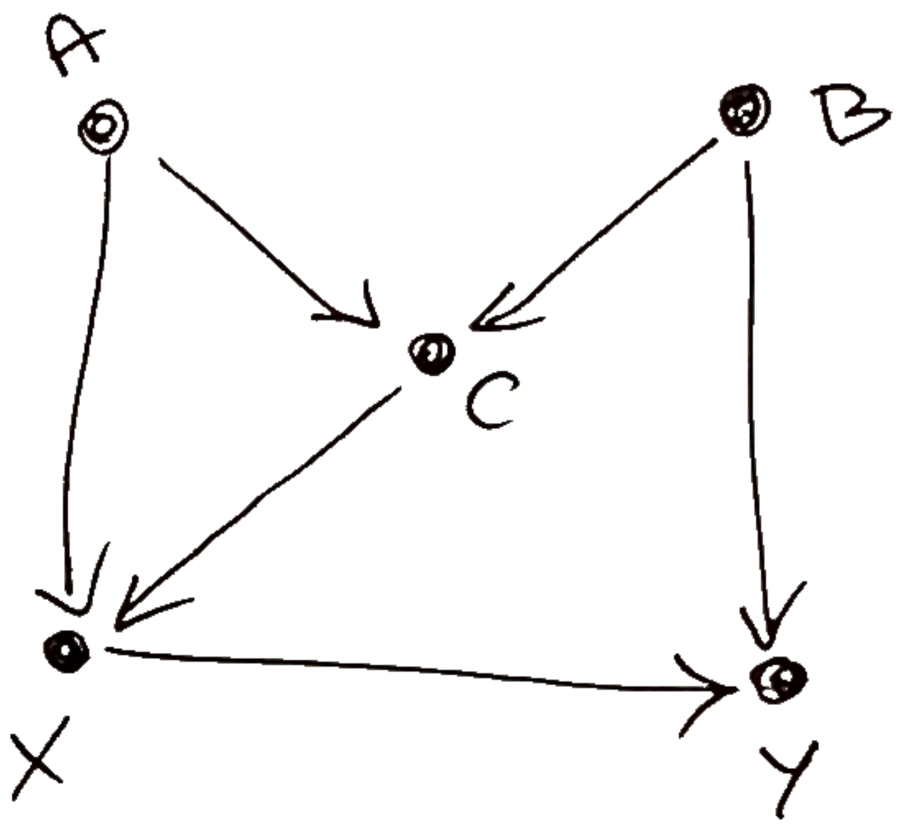
4.

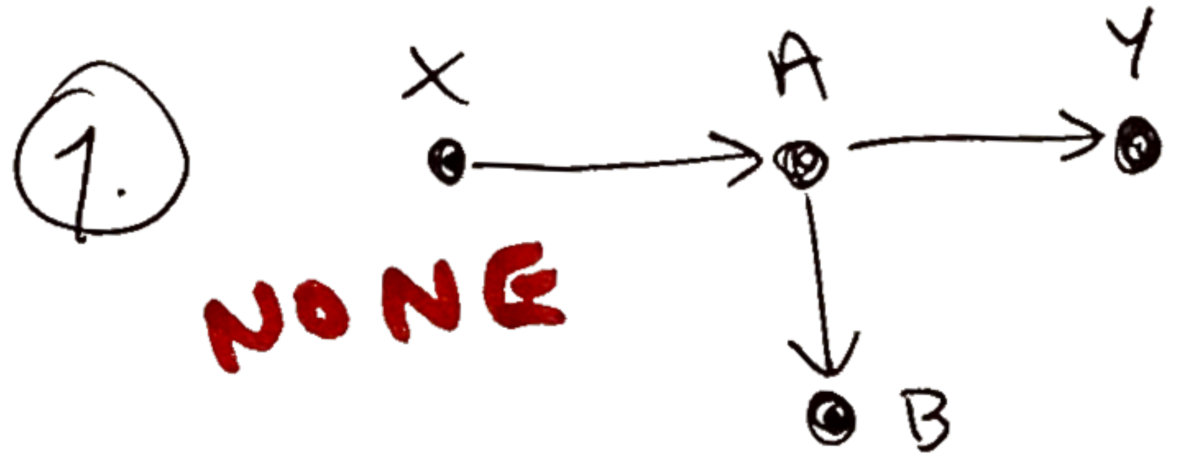
NONE

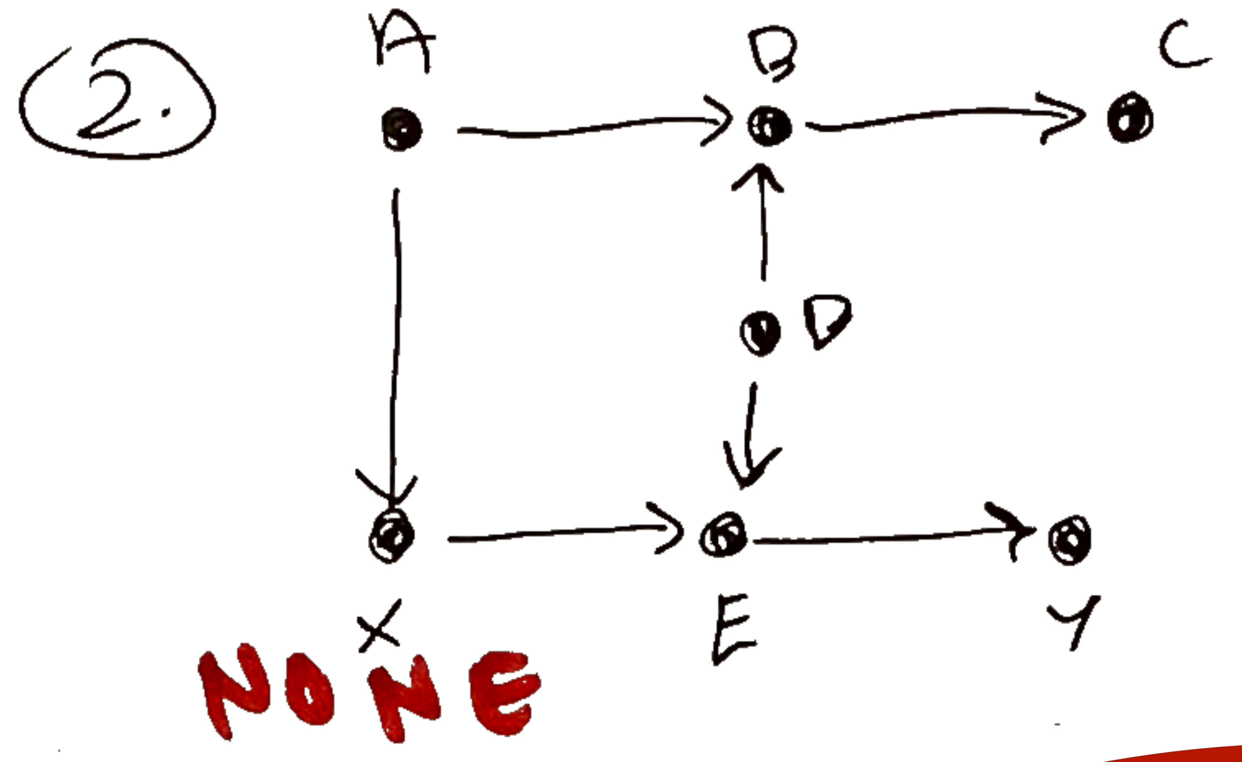


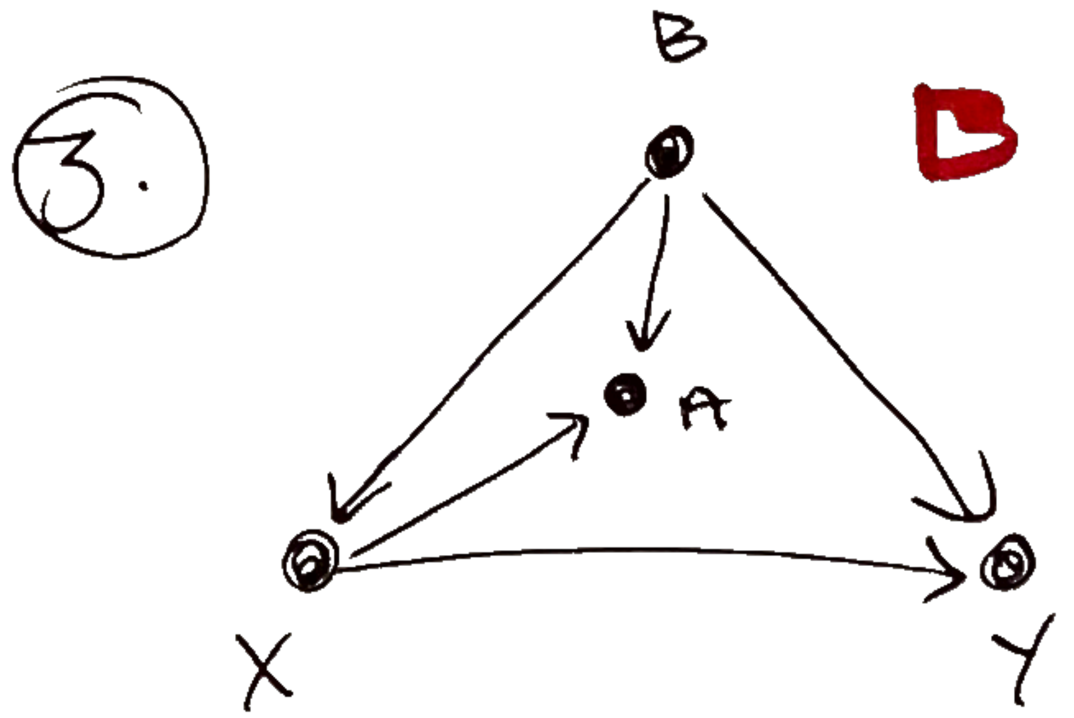
5.

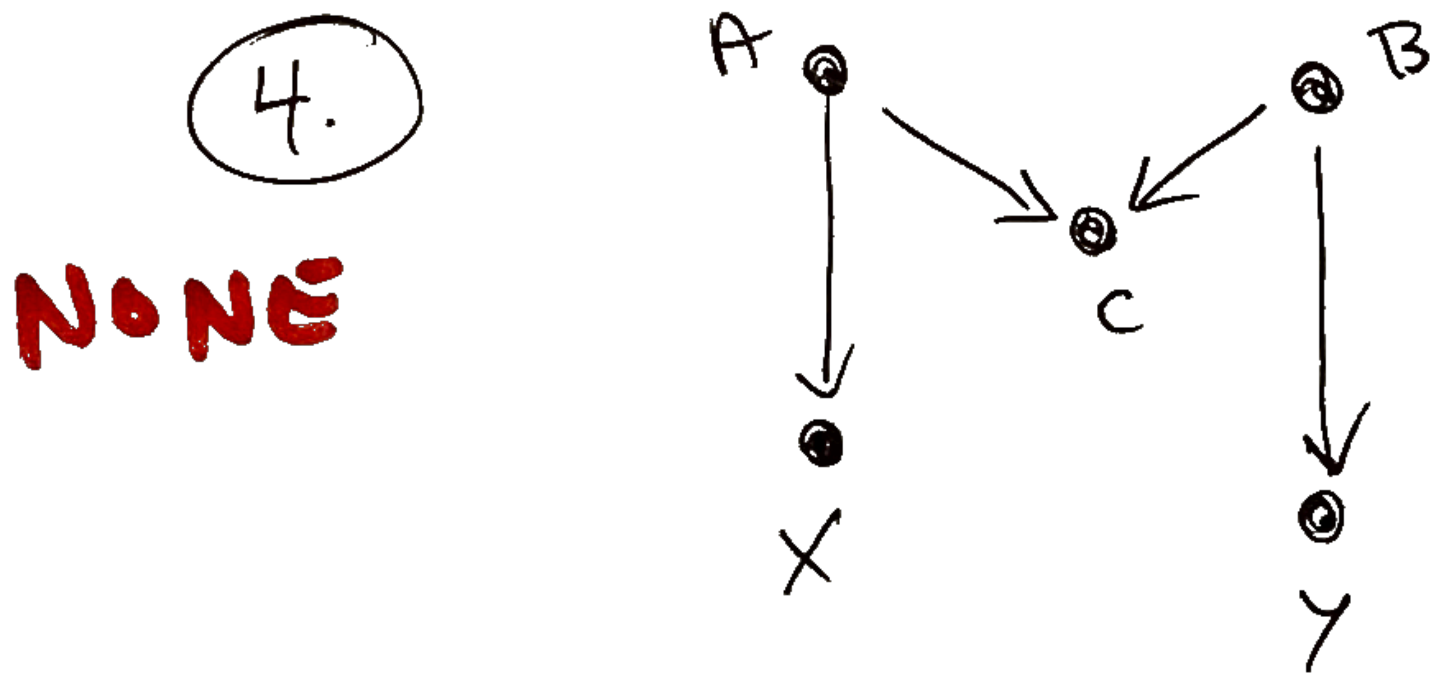
B
- or -
C + A

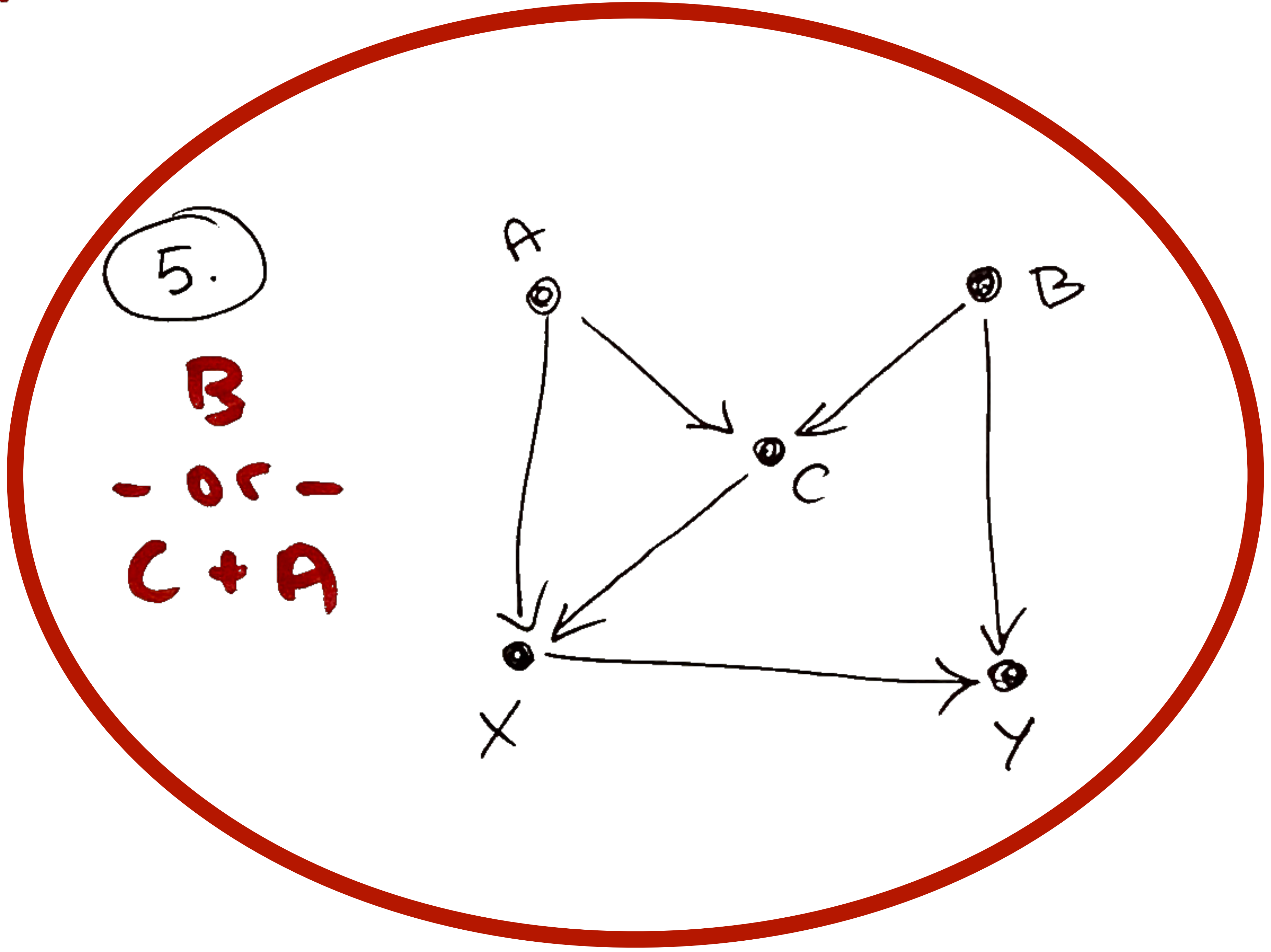








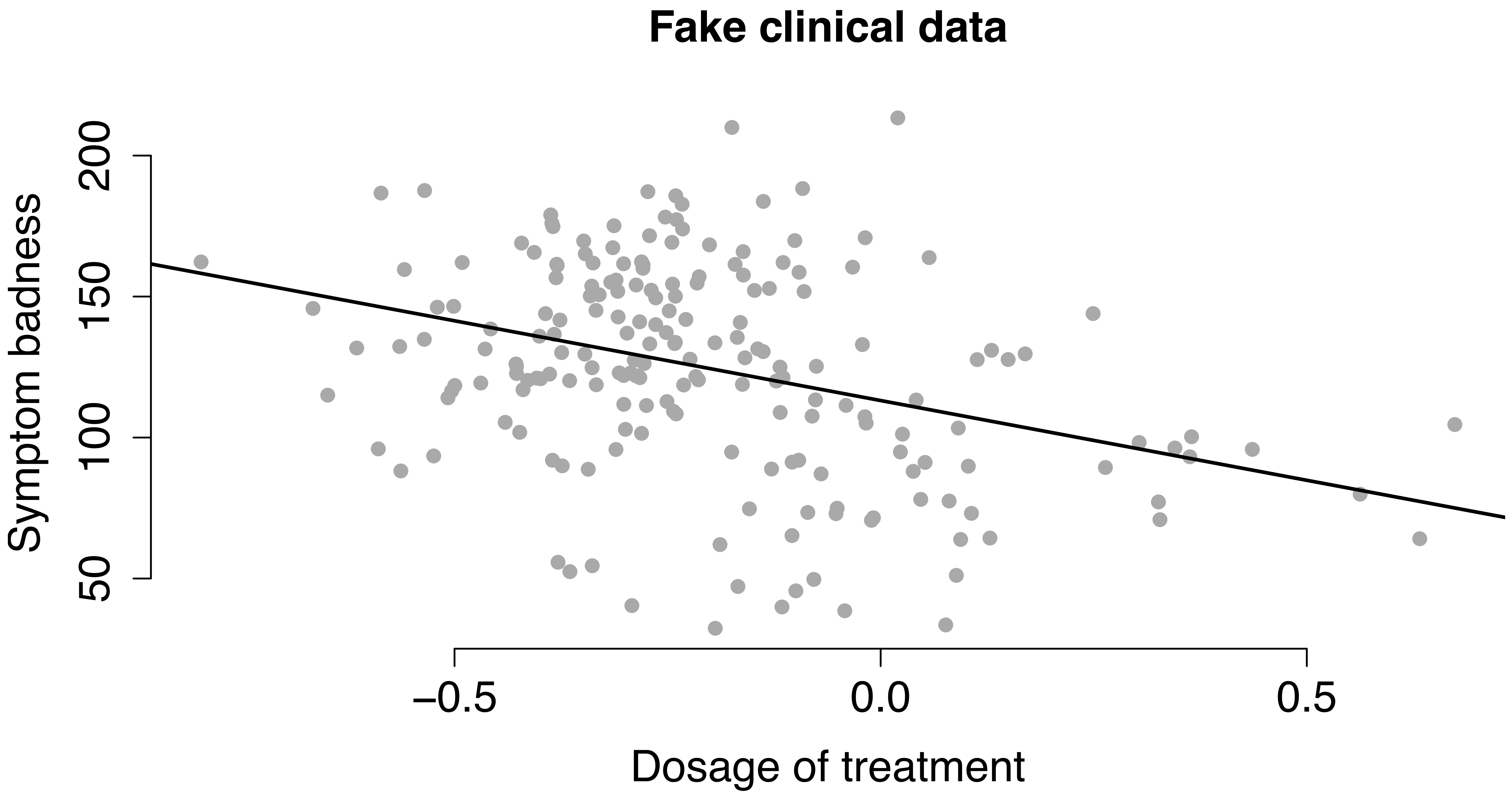




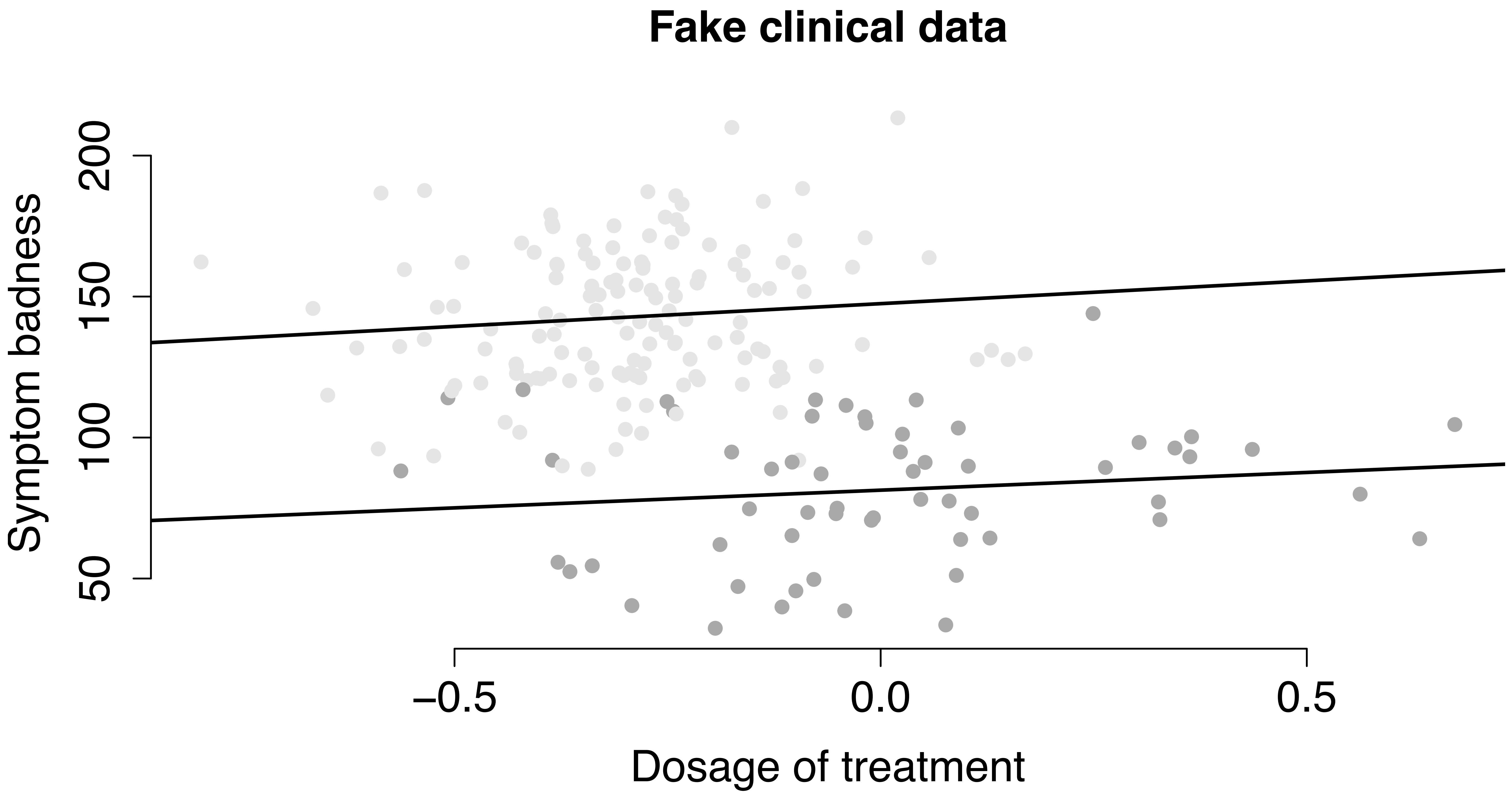
A+B+C also works but this set is not *minimal*, since we can remove something from it and still have a valid adjustment set

A warning

The analysis and interpretation of your data depends on the assumed causal model

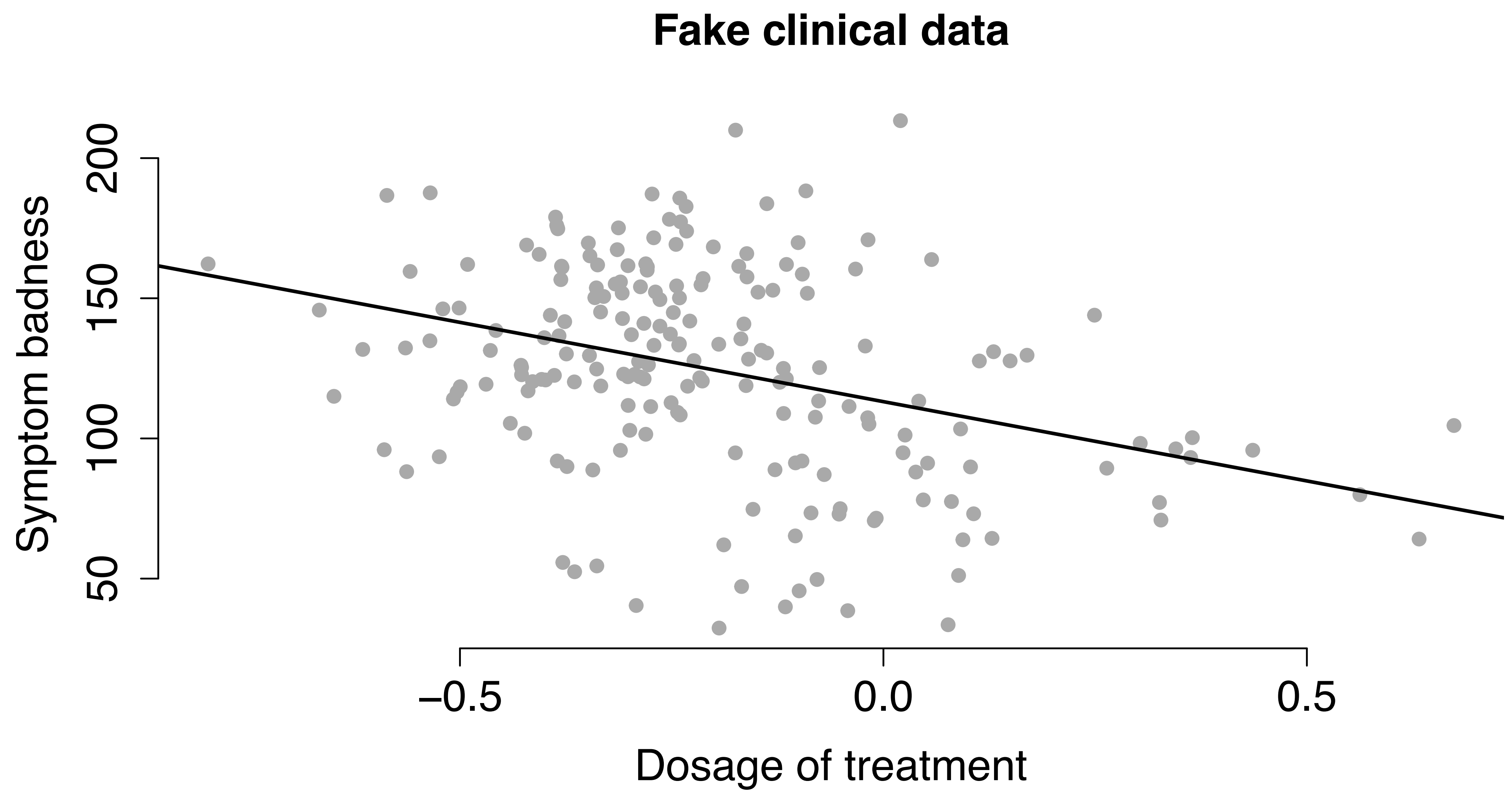


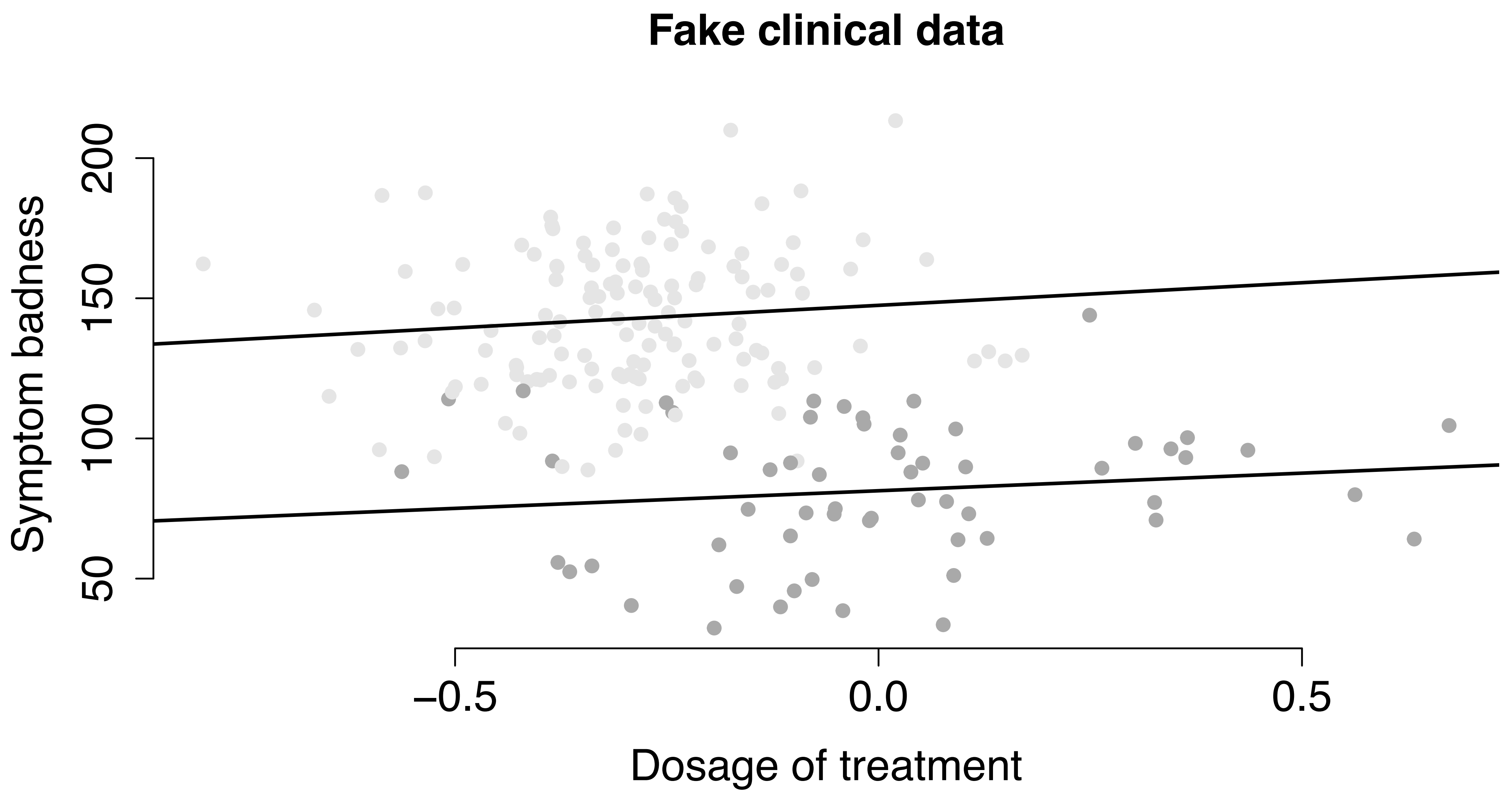
The analysis and interpretation of your data depends on the assumed causal model



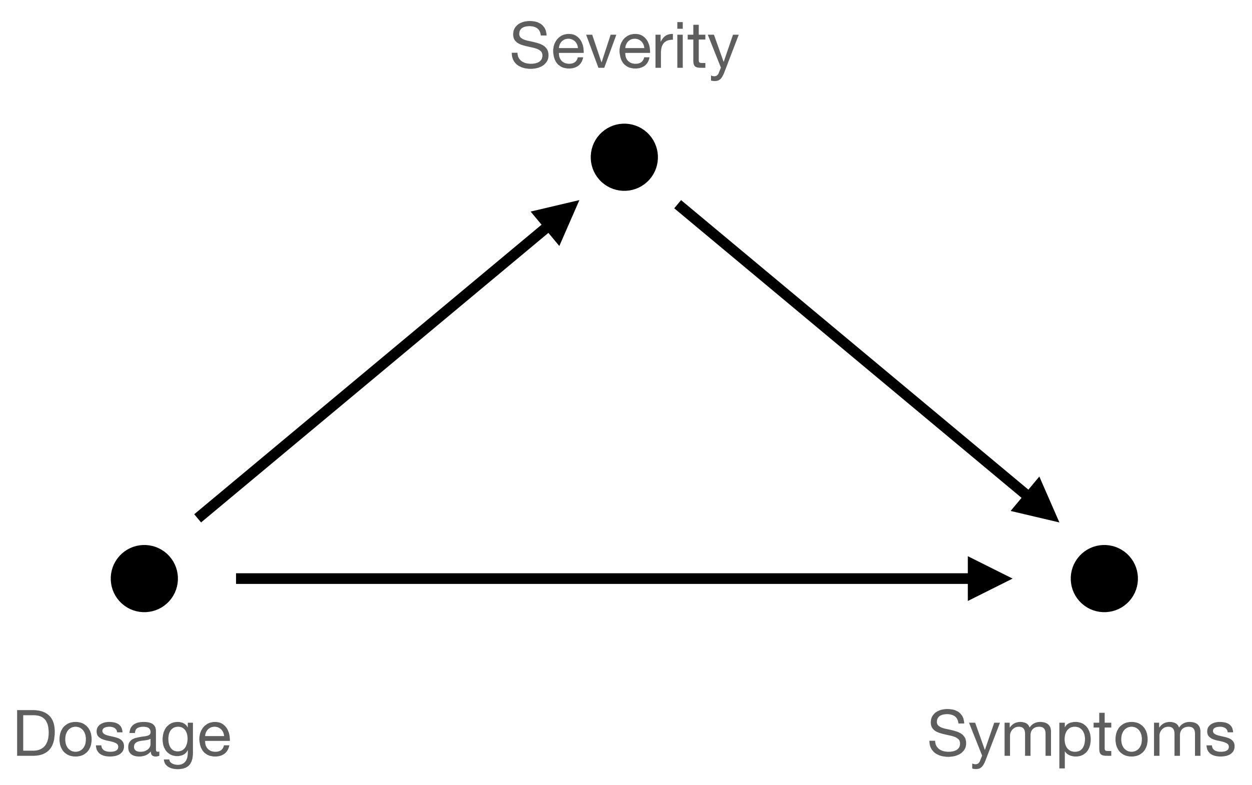
Effect reverses if we condition on disease severity

The analysis and interpretation of your data depends on the assumed causal model

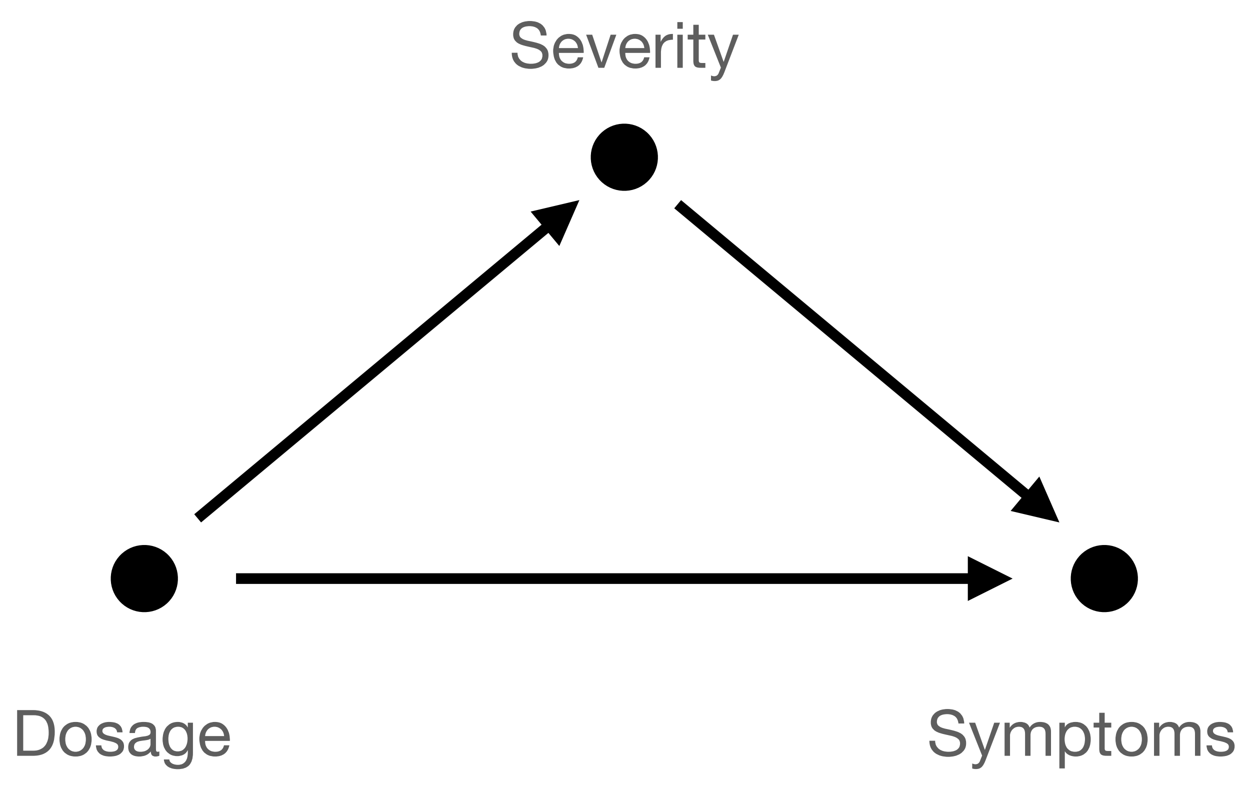




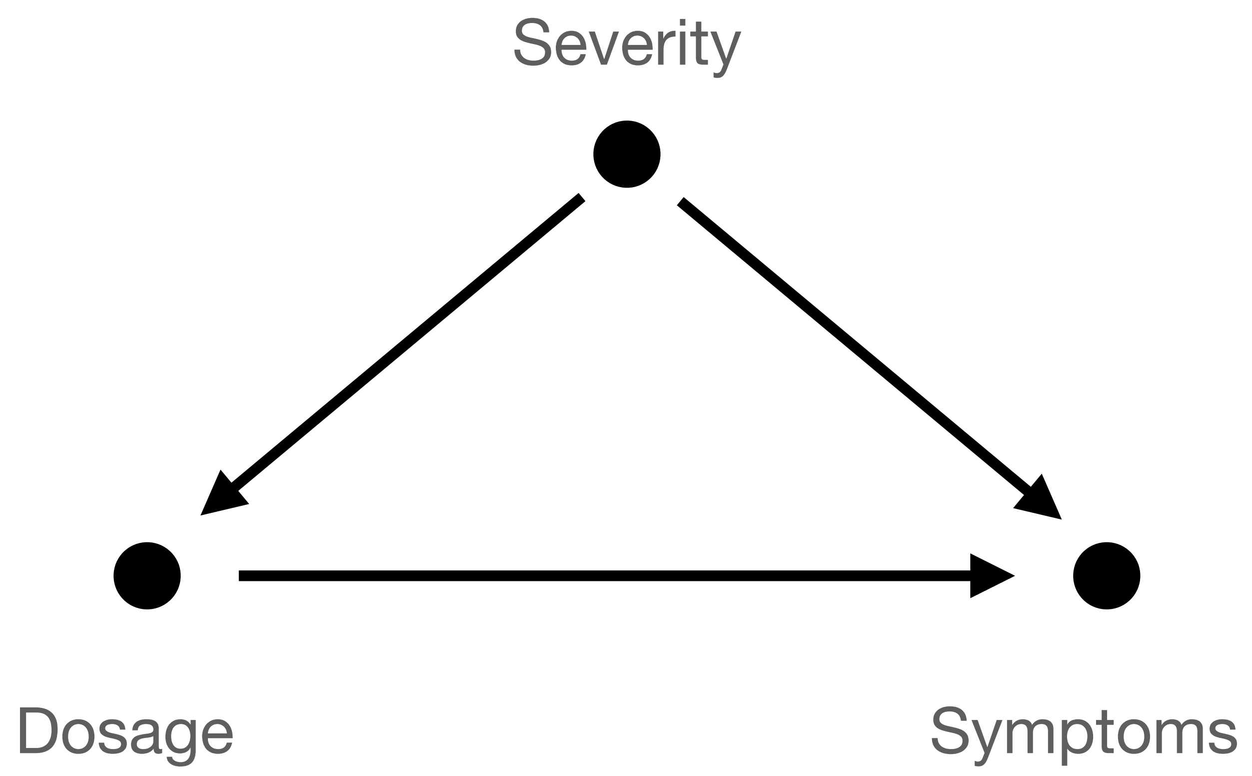
“Simpson’s paradox”: which is it, is the treatment good or bad???



Drug acts by reducing both symptoms and disease severity.
No back-door path: don't adjust.



Drug acts by reducing both symptoms and disease severity.
No back-door path: don’t adjust.



Drug acts on symptoms. Different doses given to severe and non-severe cases. Back-door path present: adjust.

Drug acts by reducing both symptoms and disease severity.
No back-door path: don't adjust.

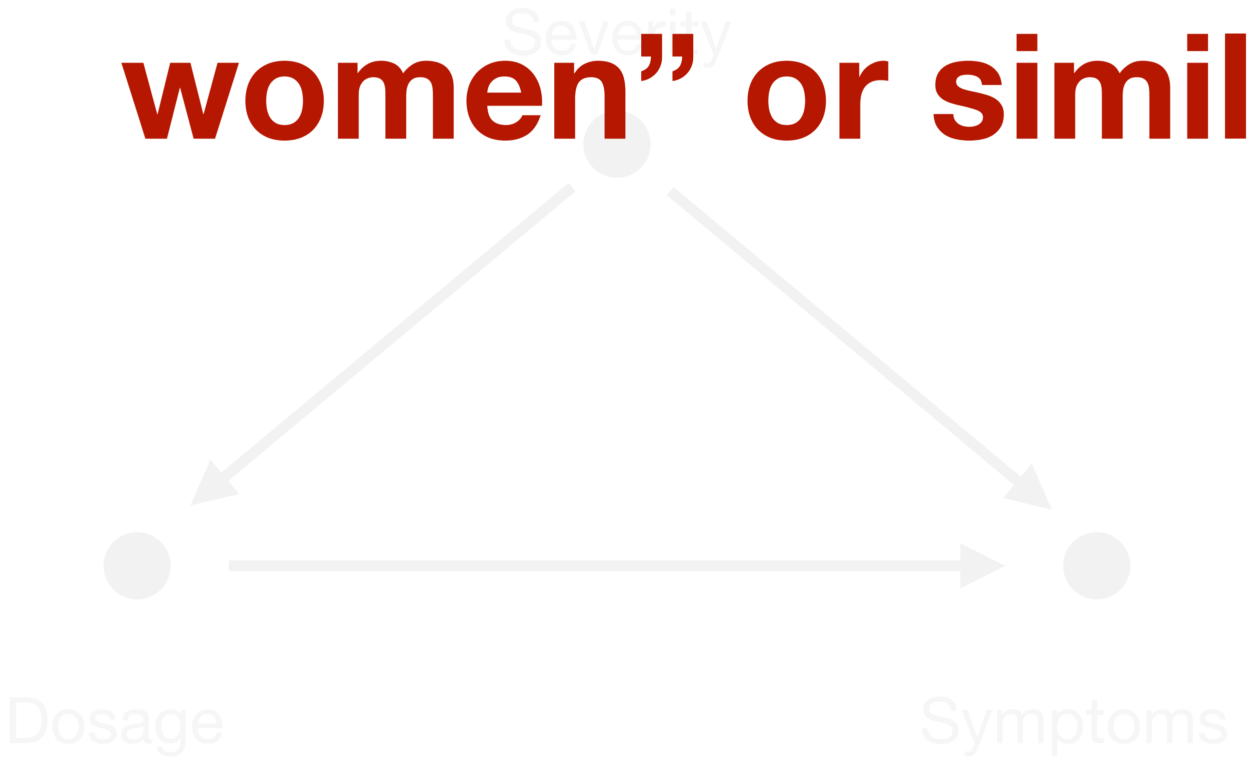
The data can NEVER tell you which is right.

Drug acts on symptoms. Different doses given to severe and non-severe cases. Back-door path present: adjust.

Drug acts by reducing both symptoms and disease severity. No back-door path: don't adjust.

The data can NEVER tell you which is right.

Be very skeptical of claims like "we found an effect but only for women" or similar.



Drug acts on symptoms. Different doses given to severe and non-severe cases. Back-door path present: adjust.

The data can NEVER tell you which is right.

Be very skeptical of claims like “we found an effect but only for women” or similar.

Could be a collider, a mediator, or a confounding factor.

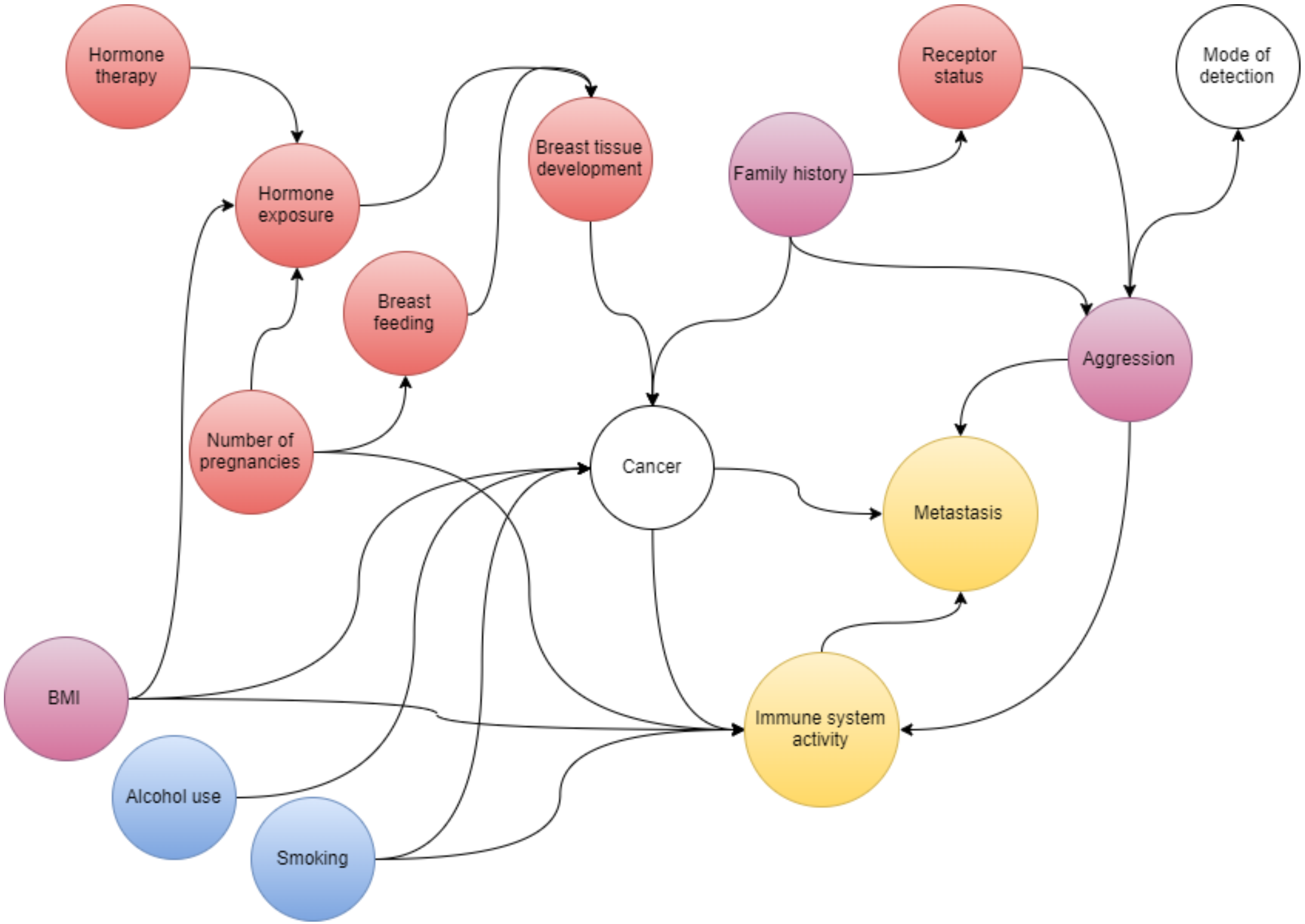
The data can NEVER tell you which is right.

Be very skeptical of claims like "we found an effect but only for women" or similar.

Could be a collider, a mediator, or a confounding factor.

Adjustment without some kind of domain knowledge (which you can encode as a DAG) practically impossible to interpret

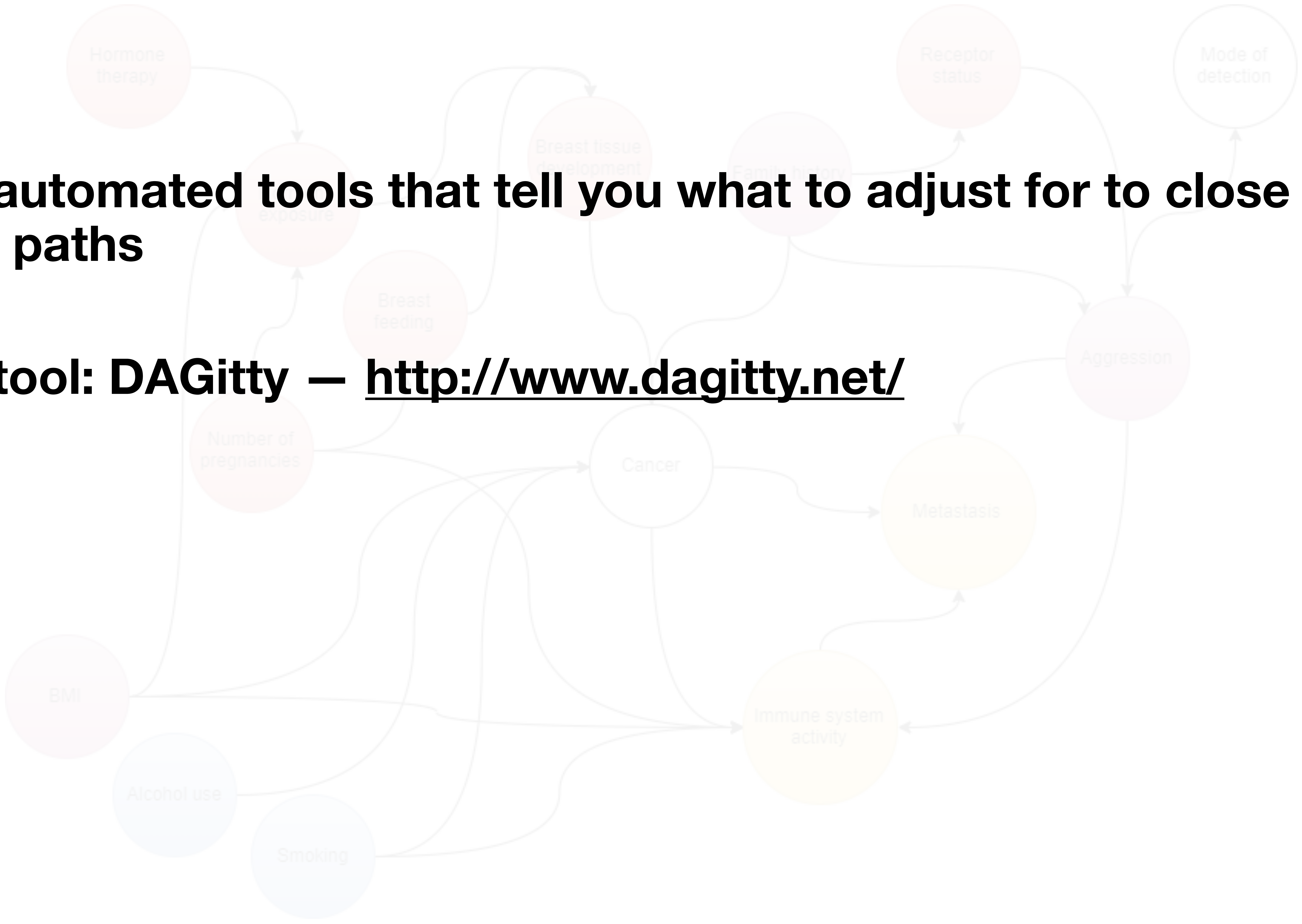
Real DAGs can be complicated



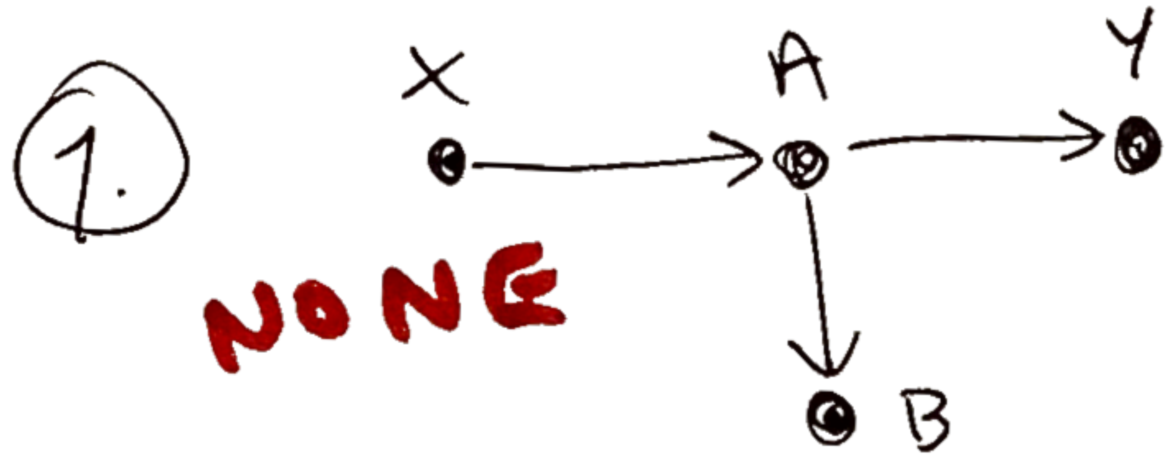
E. Holsbø and K. S. Olsen. Metastatic breast cancer and pre-diagnostic blood gene expression profiles—the norwegian women and cancer (nowac) post-genome cohort. *Frontiers in Oncology*, 10, 2020.

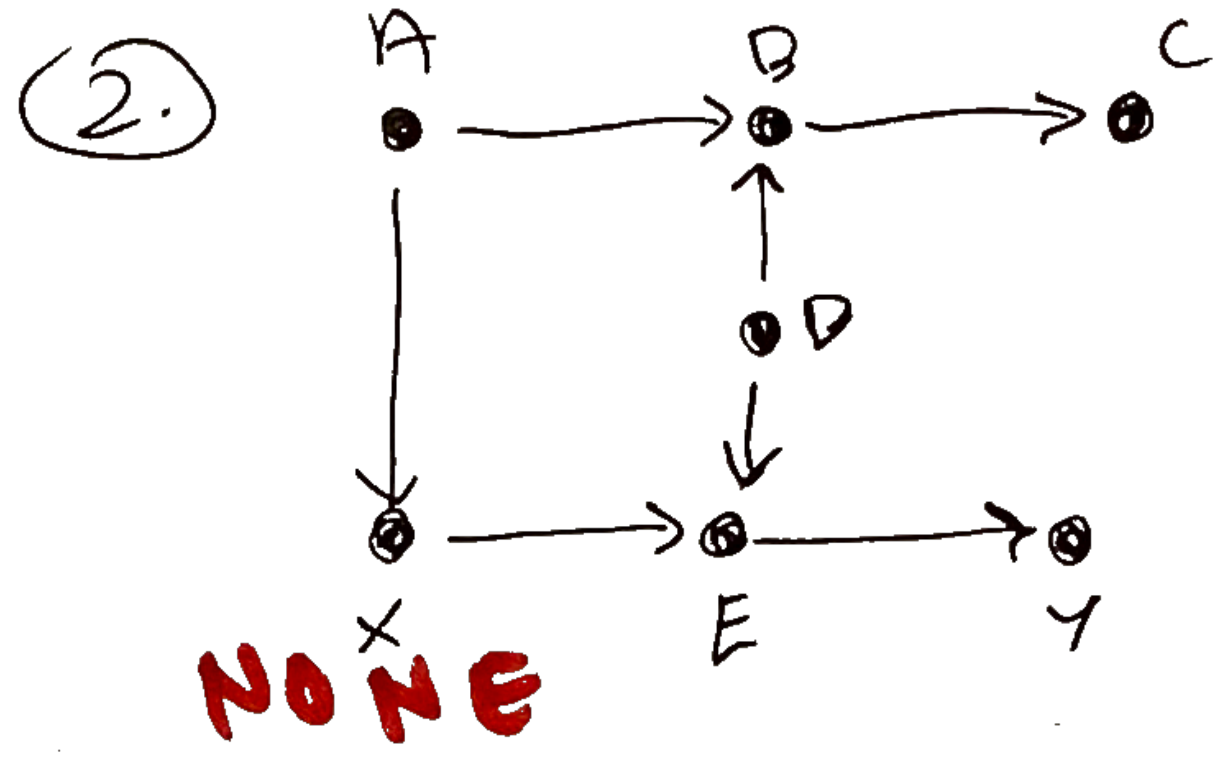
There are automated tools that tell you what to adjust for to close back-door paths

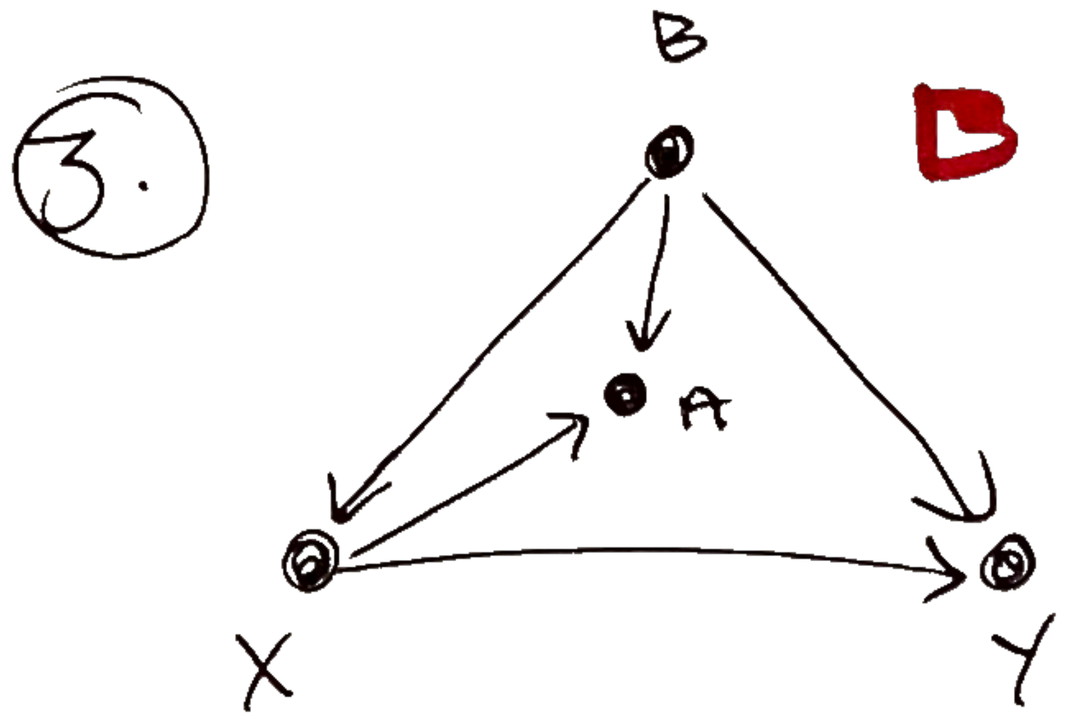
One such tool: DAGitty — http://www.dagitty.net/

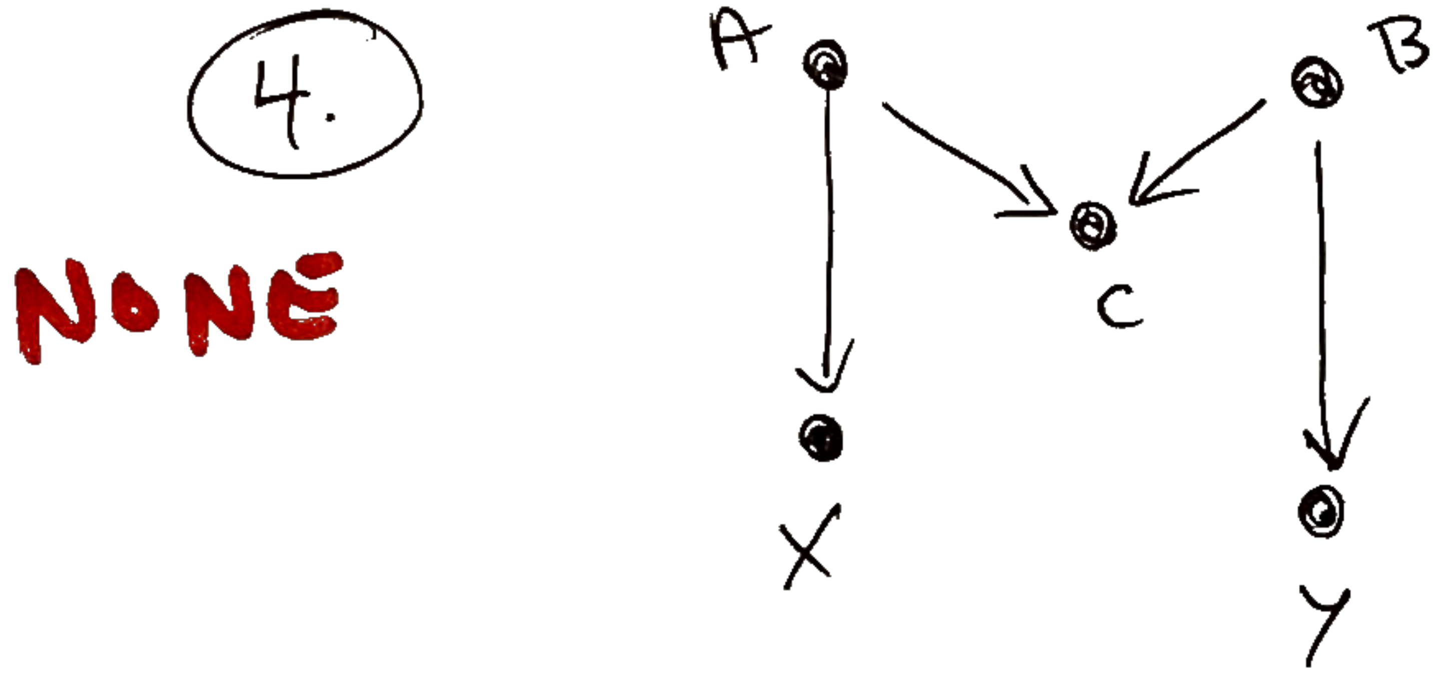


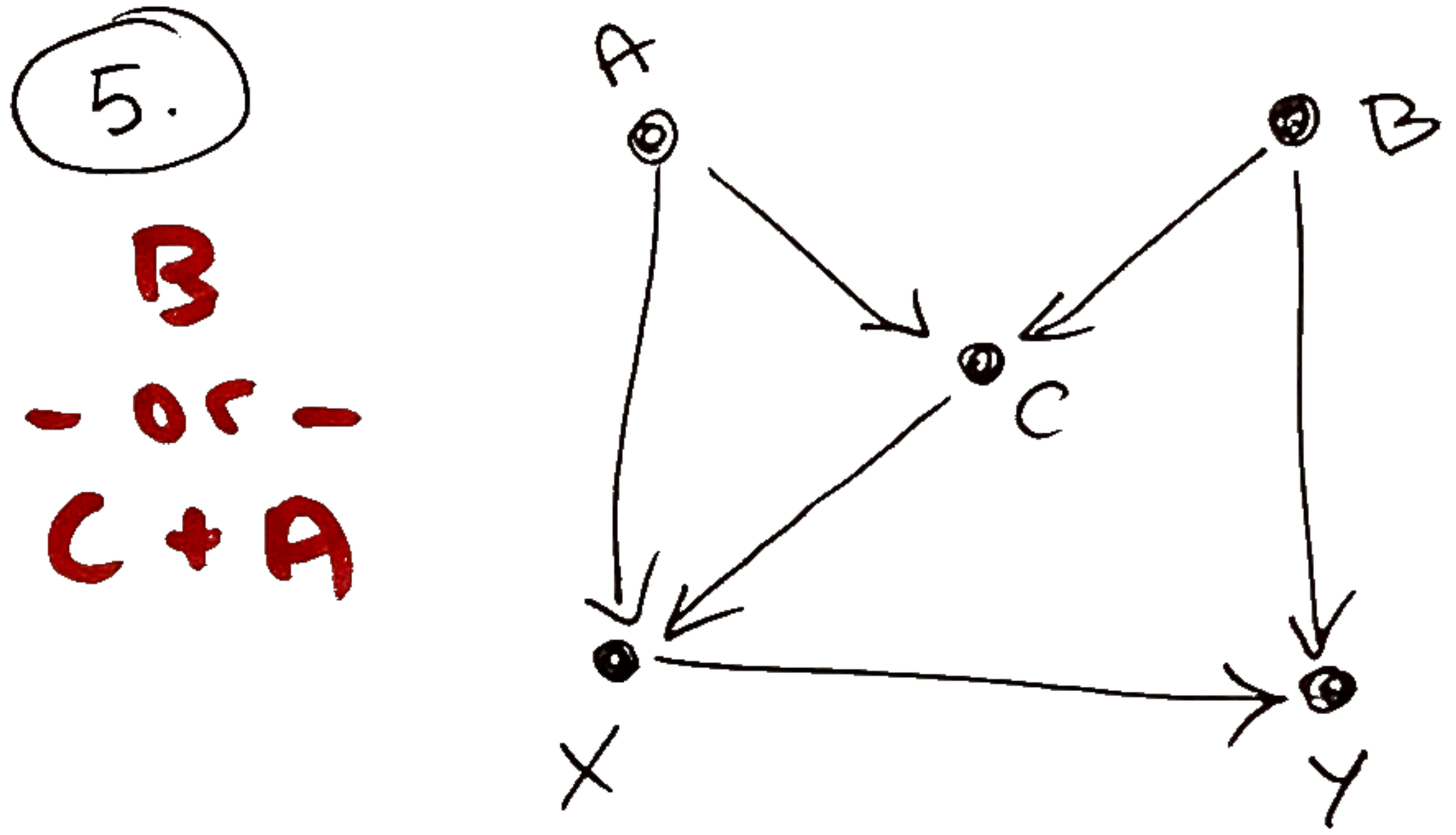
E. Holsbø and K. S. Olsen. Metastatic breast cancer and pre-diagnostic blood gene expression profiles—the norwegian women and cancer (nowac) post-genome cohort. *Frontiers in Oncology*, 10, 2020.











Go to dagitty.net, start the online browser mode. Input these DAGs and see that you get the adjustment sets you expect

Model code

```
dag {
A [pos="-2.200,-1.520"]
B [pos="1.400,-1.460"]
D [outcome,pos="1.400,1.621"]
E [exposure,pos="-
2.200,1.597"]
Z [pos="-0.300,-0.082"]
A -> E
A -> Z [pos="-0.791,-1.045"]
B -> D
B -> Z [pos="0.680,-0.496"]
E -> D
}
```

Save your DAG by copying this text. Can be pasted into the same box when you want to continue

Once you have a diagram, computers will help you do a lot of the analysis. Frees up time to think about the science

Adjustment in terms of regression models

Adjustment in terms of regression

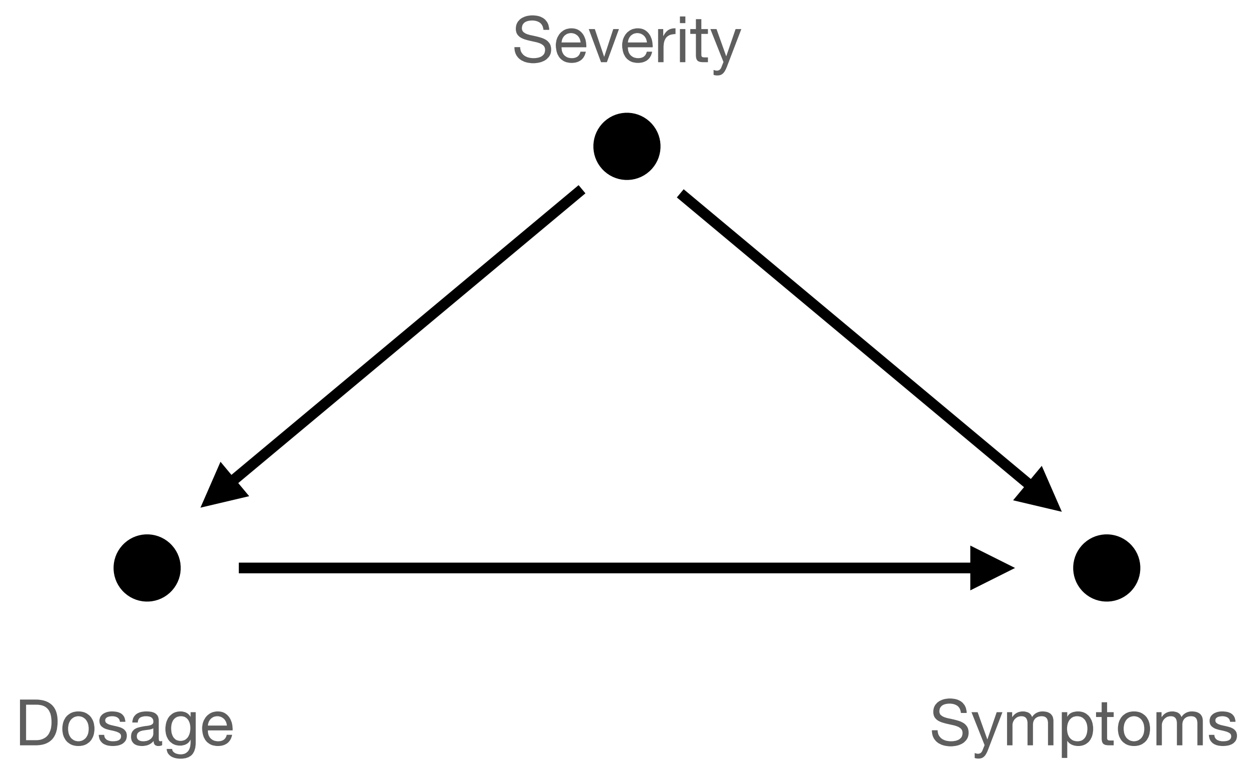
Broadly speaking: adjustment for Z to close back door paths is simply to add it as a predictor in the regression model

Big assumption: the mathematical form of the regression model is correct

Adjustment in terms of regression

Basic linear model (symptom is continuous: blood pressure maybe)

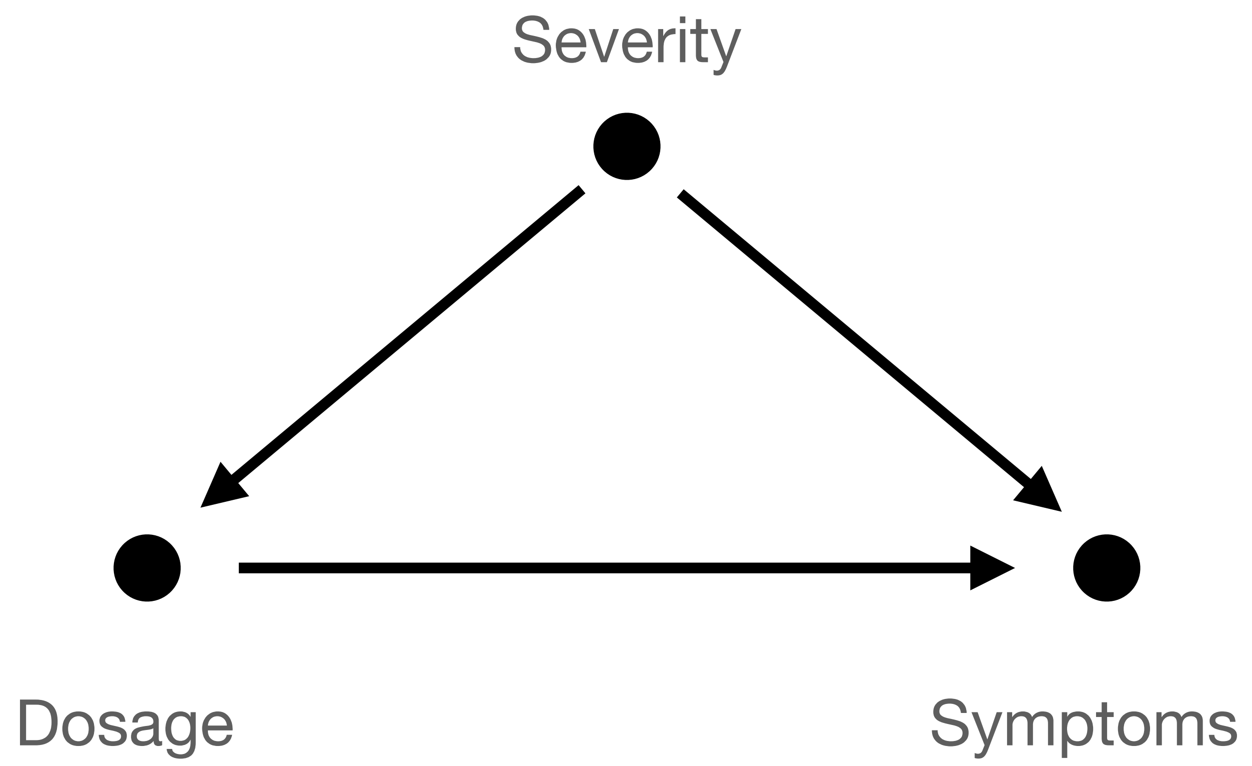
Enough to include the adjustment variable as a predictor:



Adjustment in terms of regression

Basic linear model (symptom is continuous: blood pressure maybe)

Enough to include the adjustment variable as a predictor:

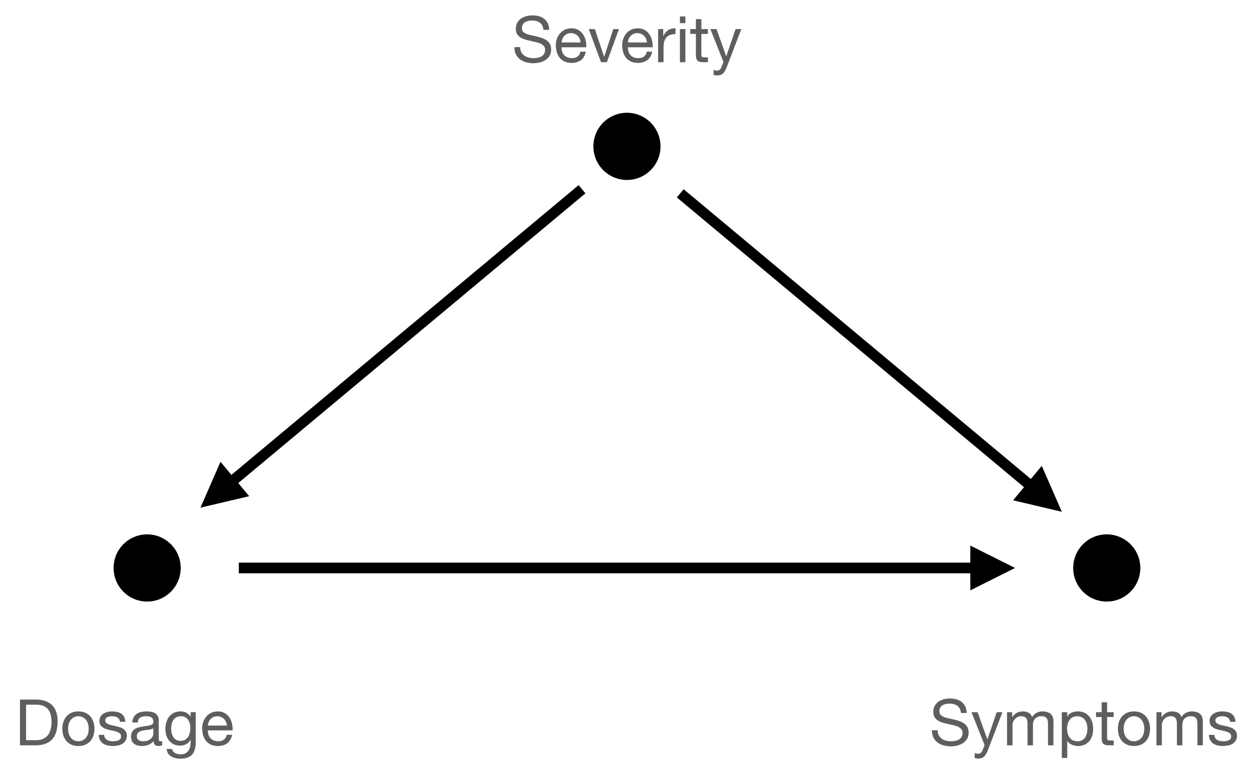


$$\text{symptom} = \beta_0 + \beta_1 \text{dosage} + \beta_2 \text{severity}$$

Adjustment in terms of regression

Basic linear model (symptom is continuous: blood pressure maybe)

Enough to include the adjustment variable as a predictor:



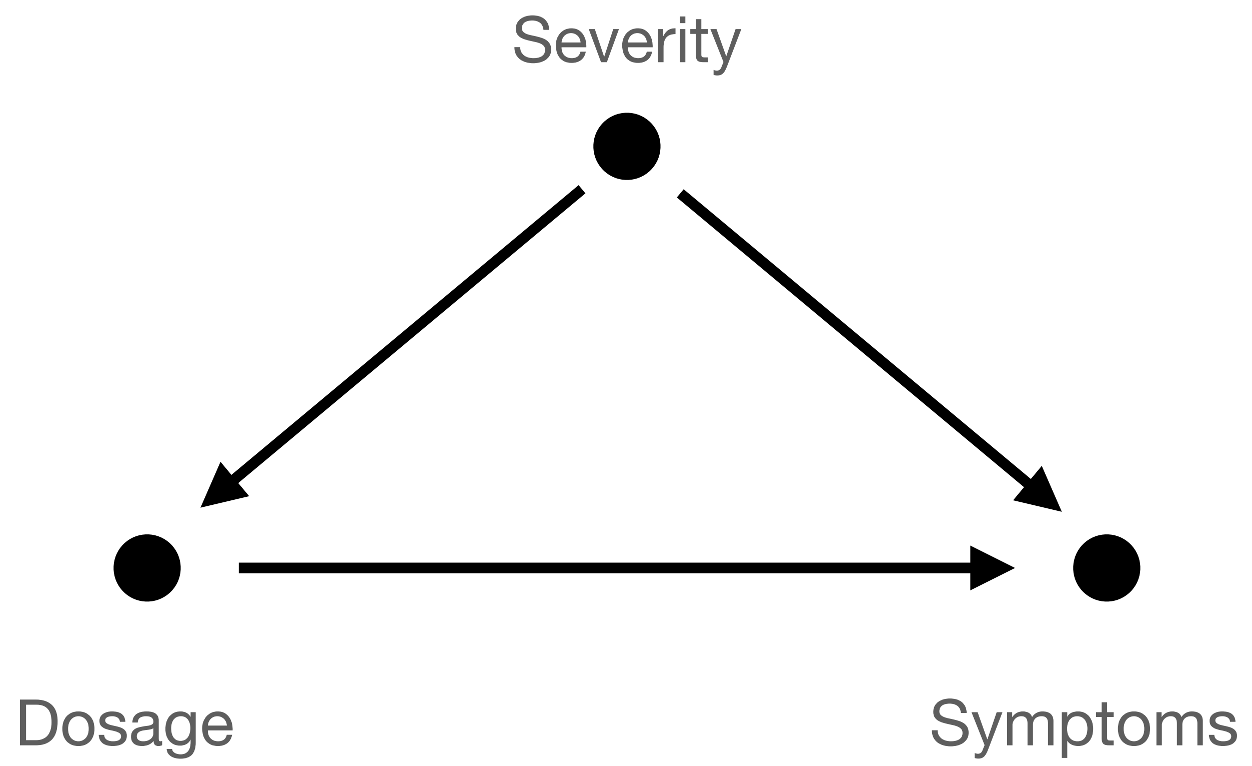
$$\text{symptom} = \beta_0 + \beta_1 \text{dosage} + \beta_2 \text{severity}$$

β_1 is an estimate of the causal effect of dosage on symptom

Adjustment in terms of regression

Basic linear model (symptom is continuous: blood pressure maybe)

Enough to include the adjustment variable as a predictor:



$$\text{symptom} = \beta_0 + \beta_1 \text{dosage} + \beta_2 \text{severity}$$

β_1 is an estimate of the causal effect of dosage on symptom

β_2 **does not** have a straight-forward interpretation: Simply there to "deconfound"

Daniel Westreich, Sander Greenland, The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients, American Journal of Epidemiology, Volume 177, Issue 4, 15 February 2013, Pages 292–298, https://doi.org/10.1093/aje/kws412

Adjustment in terms of regression

Average treatment effect

The β_1 of linear regression is an estimate of the **average treatment effect:**

$$\mathbb{E}[Y_1 - Y_0]$$

Adjustment in terms of regression

Average treatment effect

The β_1 of linear regression is an estimate of the **average treatment effect:**

$$\mathbb{E}[Y_1 - Y_0]$$

Which reads "average difference between giving someone the treatment and not giving them the treatment"

Adjustment in terms of regression

Logistic regression and similar (eg. binary outcome)

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x + \beta_2 z \qquad p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x + \beta_2 z)}}$$

Adjustment is still to add Z as predictor, BUT interpretation more tricky.

Adjustment in terms of regression

Logistic regression and similar (eg. binary outcome)

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x + \beta_2 z \qquad p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x + \beta_2 z)}}$$

Adjustment is still to add Z as predictor, BUT interpretation more tricky.

We know that e^{β_1} is an estimate of odds ratio, but it is not the “average causal” odds ratio. We say that the odds ratio is *noncollapsible*.

Adjustment in terms of regression

Logistic regression and similar (eg. binary outcome)

It's always possible to "stratify-and-average:"

Adjustment in terms of regression

Logistic regression and similar (eg. binary outcome)

It’s always possible to “stratify-and-average:”

$$P(Y_x) = \sum_z P(Y \mid X = x, Z = z) P(Z = z)$$

Adjustment in terms of regression

Logistic regression and similar (eg. binary outcome)

It's always possible to "stratify-and-average:"

$$P(Y_x) = \sum_z P(Y \mid X = x, Z = z) P(Z = z)$$

i) Estimate effect in each stratum (ie. Condition on Z = z)

Adjustment in terms of regression

Logistic regression and similar (eg. binary outcome)

It's always possible to "stratify-and-average:"

$$P(Y_x) = \sum_z P(Y \mid X = x, Z = z) P(Z = z)$$

i) Estimate effect in each stratum (ie. Condition on Z = z)
ii) Weight by fraction of observations in stratum z

Adjustment in terms of regression

Logistic regression and similar (eg. binary outcome)

It's always possible to "stratify-and-average:"

$$P(Y_x) = \sum_z P(Y \mid X = x, Z = z) P(Z = z) = \mathbb{E}_Z P(Y \mid X = x, Z = z)$$

i) Estimate effect in each stratum (ie. Condition on Z = z)

ii) Weight by fraction of observations in stratum z

Adjustment in terms of regression

Logistic regression and similar (eg. binary outcome)

It's always possible to "stratify-and-average:"

$$P(Y_x) = \sum_z P(Y \mid X = x, Z = z) P(Z = z) = \mathbb{E}_Z P(Y \mid X = x, Z = z)$$

i) Estimate effect in each stratum (ie. Condition on Z = z)
ii) Weight by fraction of observations in stratum z

Adjustment in terms of regression

Logistic regression and similar (eg. binary outcome)

It’s always possible to “stratify-and-average:”

$$P(Y_x) = \sum_z P(Y \mid X = x, Z = z) P(Z = z) = \mathbb{E}_Z P(Y \mid X = x, Z = z)$$

i) Estimate effect in each stratum (ie. Condition on Z = z)
ii) Weight by fraction of observations in stratum z

Slightly annoying to do by hand, but there is software

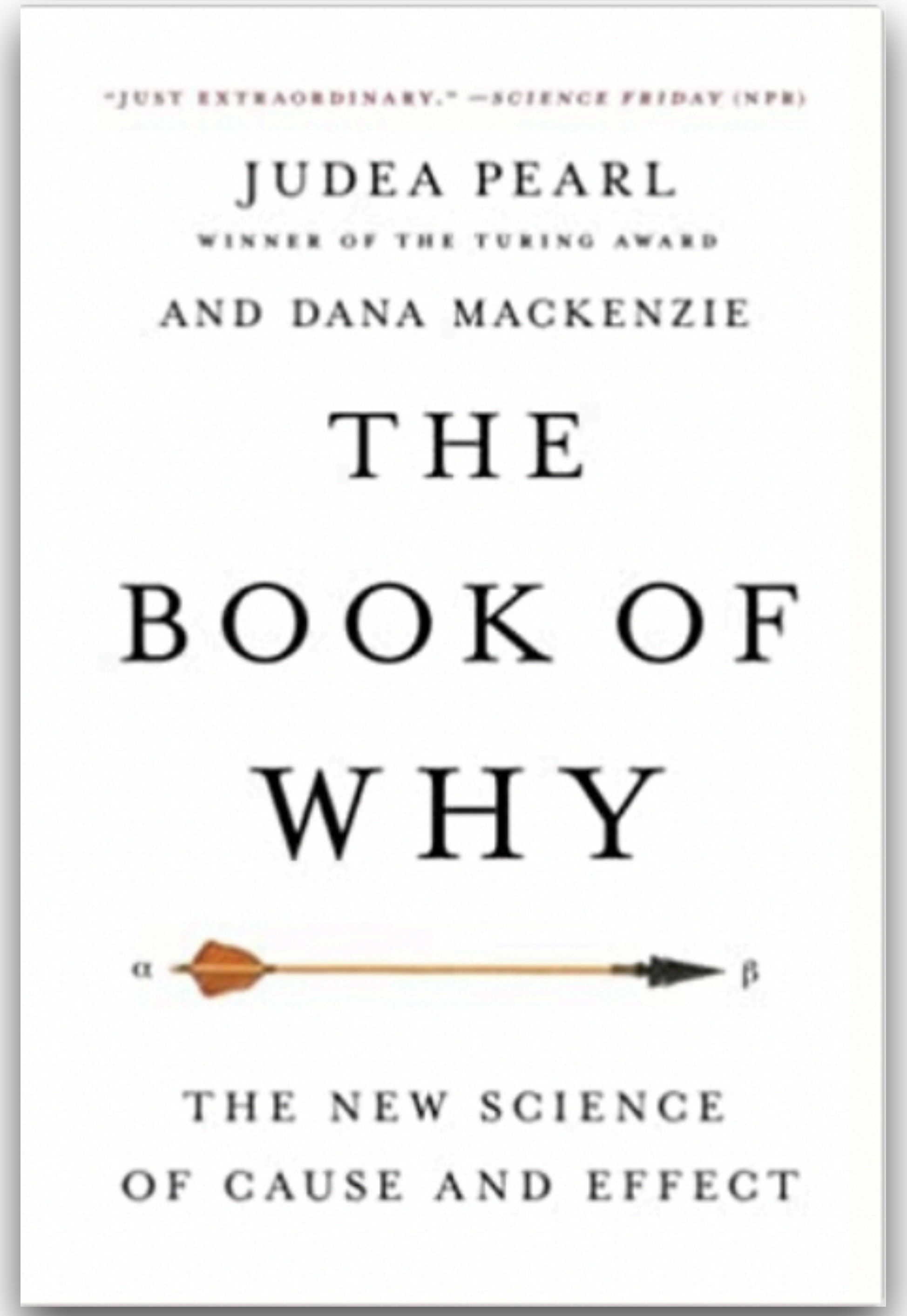
Eur J Epidemiol. 2016 Jun;31(6):563-74. doi: 10.1007/s10654-016-0157-3. Epub 2016 May 14.

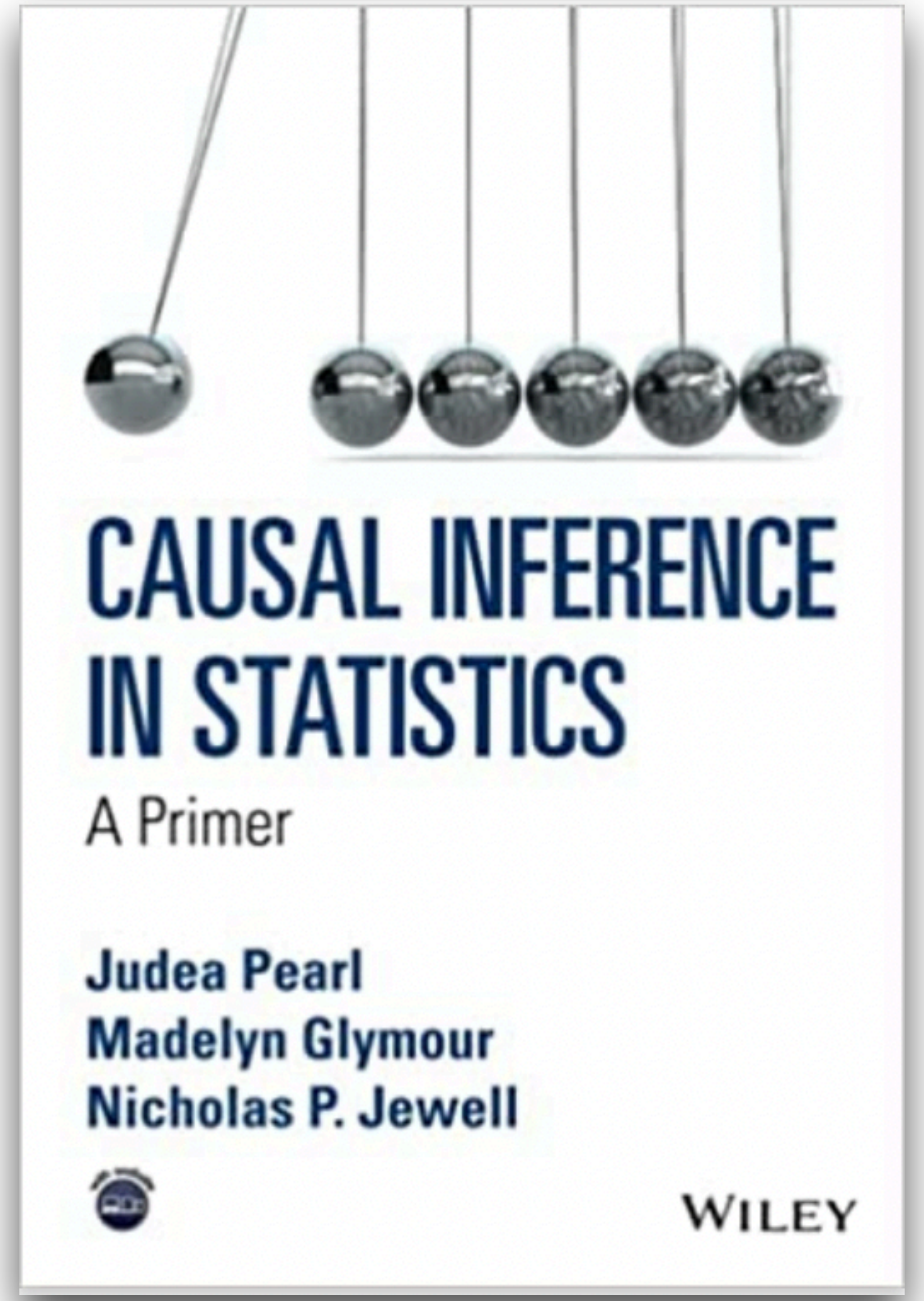
Regression standardization with the R package stdReg

Arvid Sjölander [1]

Last slides coming up!


Humans are great at drawings,
terrible at text and numbers


"JUST EXTRAORDINARY." —SCIENCE FRIDAY (NPR)
JUDEA PEARL
WINNER OF THE TURING AWARD
AND DANA MACKENZIE
THE BOOK OF WHY
THE NEW SCIENCE OF CAUSE AND EFFECT


CAUSAL INFERENCE IN STATISTICS
A Primer
Judea Pearl
Madelyn Glymour
Nicholas P. Jewell
WILEY

Thank you.