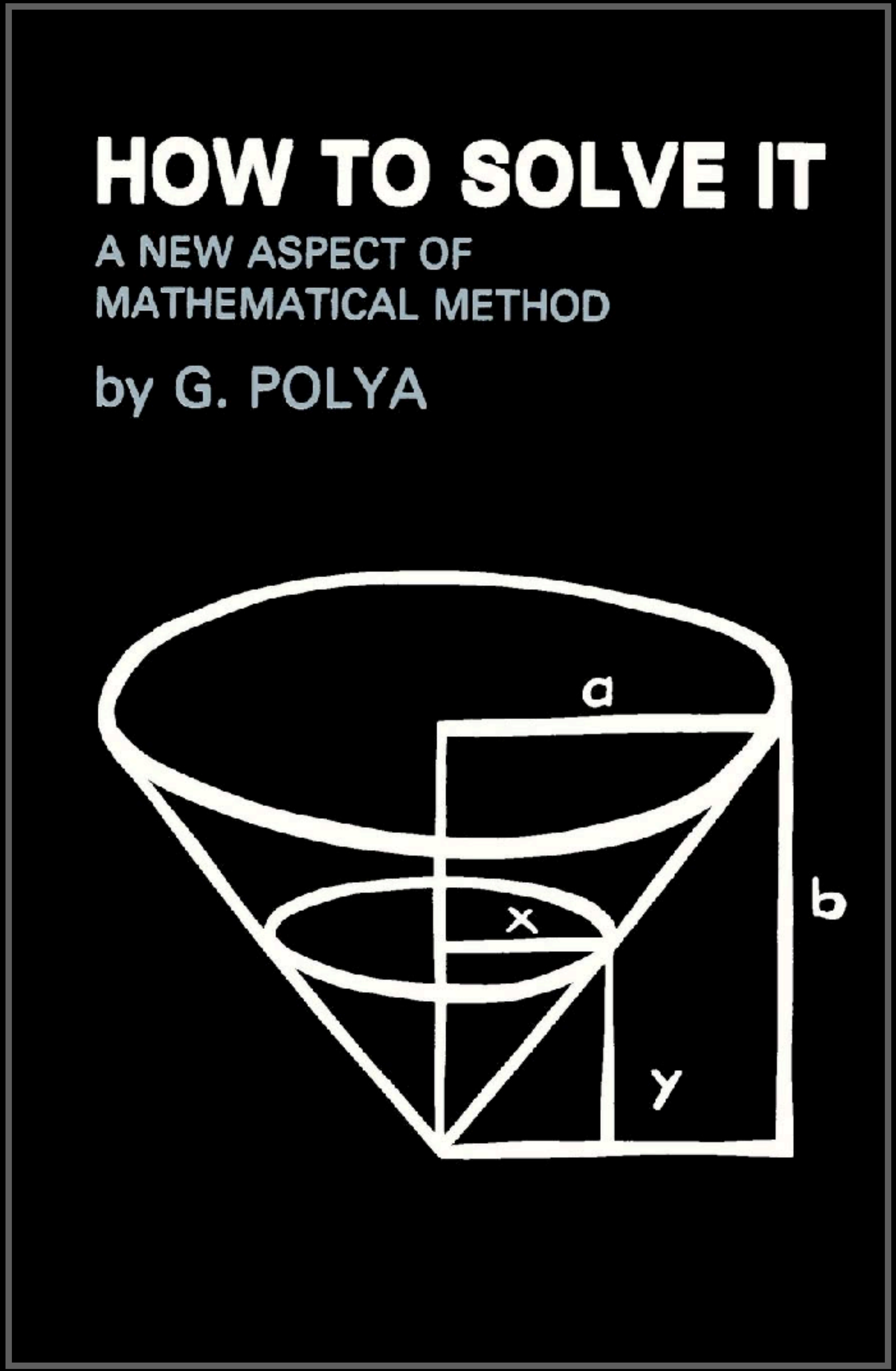
Statistical Rethinking
2023

11. Ordered Categories



George Pólya (1887–1985)

HOW TO SOLVE IT

UNDERSTANDING THE PROBLEM

First.

You have to *understand* the problem.

What is the unknown? What are the data? What is the condition?

Is it possible to satisfy the condition? Is the condition sufficient to determine the unknown? Or is it insufficient? Or redundant? Or contradictory?

Draw a figure. Introduce suitable notation.

Separate the various parts of the condition. Can you write them down?

DEVISING A PLAN

Second.

Find the connection between the data and the unknown. You may be obliged to consider auxiliary problems if an immediate connection cannot be found. You should obtain eventually a *plan* of the solution.

Have you seen it before? Or have you seen the same problem in a slightly different form?

Do you know a related problem? Do you know a theorem that could be useful?

Look at the unknown! And try to think of a familiar problem having the same or a similar unknown.

Here is a problem related to yours and solved before. Could you use it? Could you use its result? Could you use its method? Should you introduce some auxiliary element in order to make its use possible?

Could you restate the problem? Could you restate it still differently? Go back to definitions.

If you cannot solve the proposed problem try to solve first some related problem. Could you imagine a more accessible related problem? A more general problem? A more special problem? An analogous problem? Could you solve a part of the problem? Keep only a part of the condition, drop the other part; how far is the unknown then determined, how can it vary? Could you derive something useful from the data? Could you think of other data appropriate to determine the unknown? Could you change the unknown or the data, or both if necessary, so that the new unknown and the new data are nearer to each other?

Did you use all the data? Did you use the whole condition? Have you taken into account all essential notions involved in the problem?

CARRYING OUT THE PLAN

Third.

Carry out your plan.

Carrying out your plan of the solution, *check each step*. Can you see clearly that the step is correct? Can you prove that it is correct?

LOOKING BACK

Fourth.

Examine the solution obtained.

Can you *check the result?* Can you check the argument?

Can you derive the result differently? Can you see it at a glance?

Can you use the result, or the method, for some other problem?

If you cannot solve the proposed problem try to solve first some related problem. Could you imagine a more accessible related problem? A more general problem? A more special problem? An analogous problem? Could you solve a part of the problem? Keep only a part of the condition, drop the other part; how far is the unknown then determined, how can it vary? Could you derive something useful from the data? Could you think of other data appropriate to determine the unknown? Could you change the unknown or the data, or both if necessary, so that the new unknown and the new data are nearer to each other?

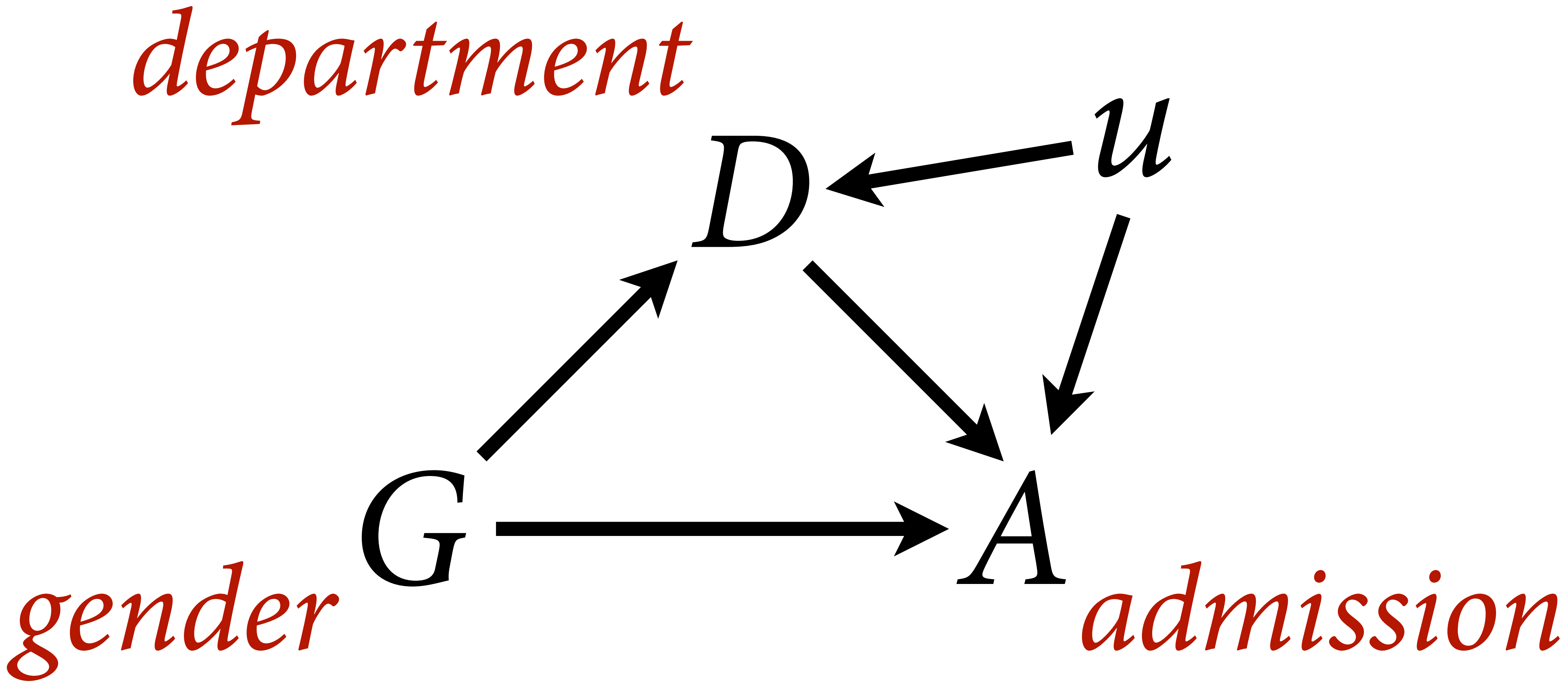
Did you use all the data? Did you use the whole condition? Have you taken into account all essential notions involved in the problem?

CARRYING OUT THE PLAN

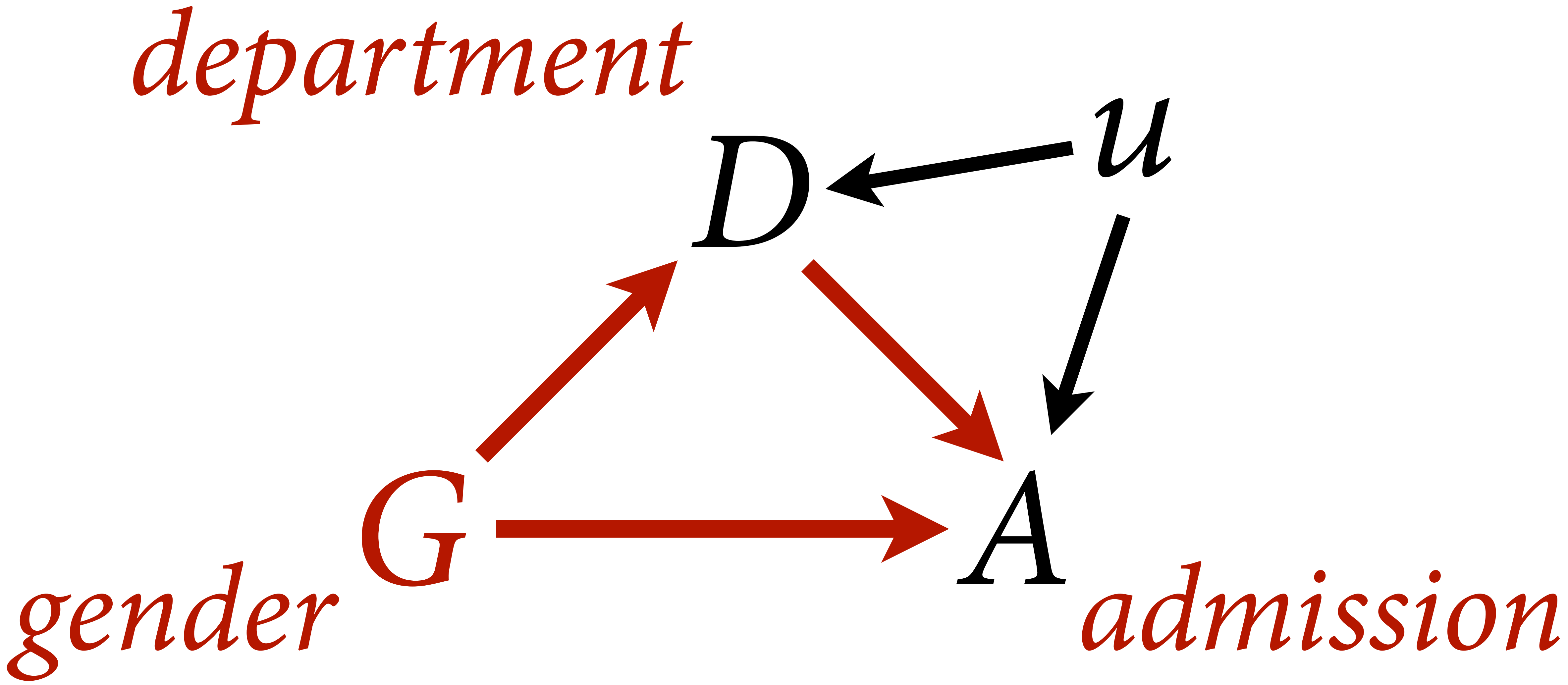
Third.
... your plan.

Carrying out your plan of the solution, *check each step*. Can you see clearly that the step is correct? Can you prove that it is correct?

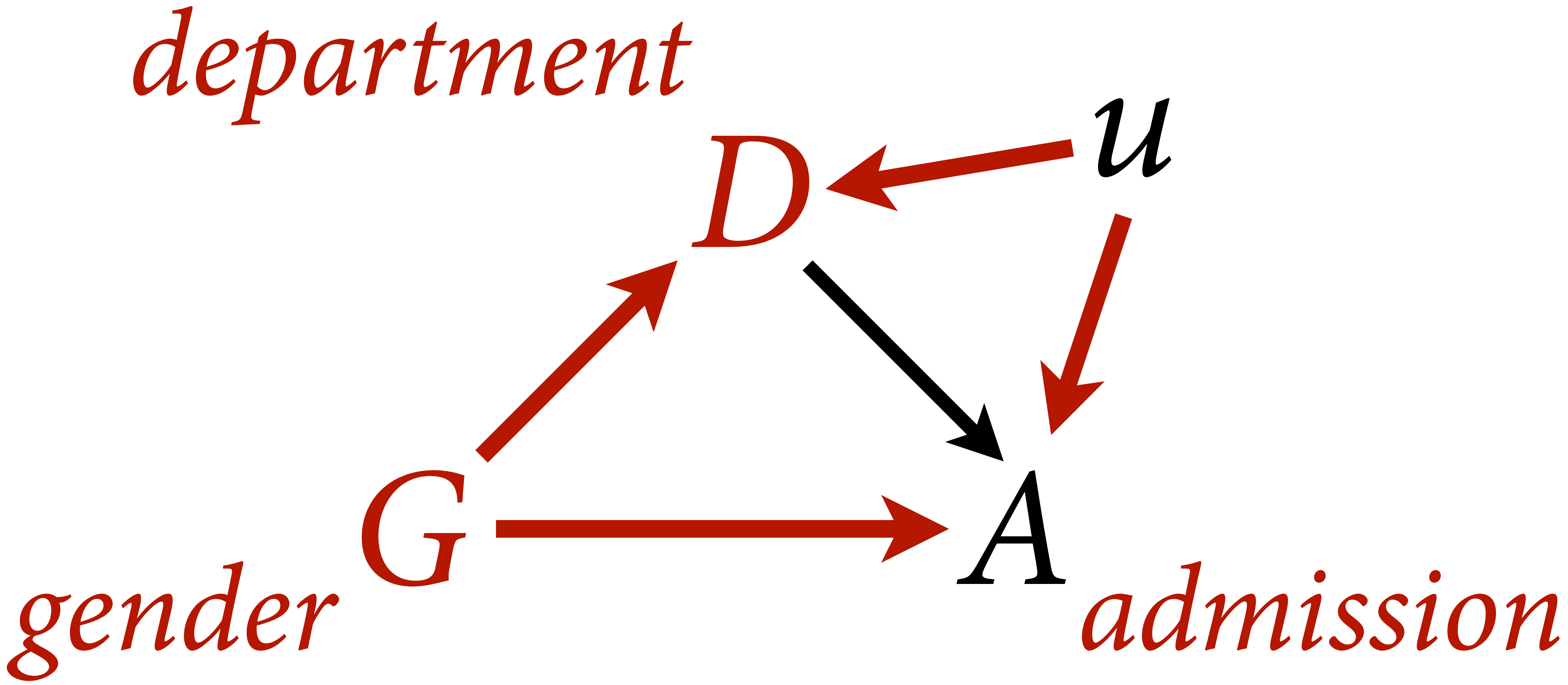
Can't always get what you want

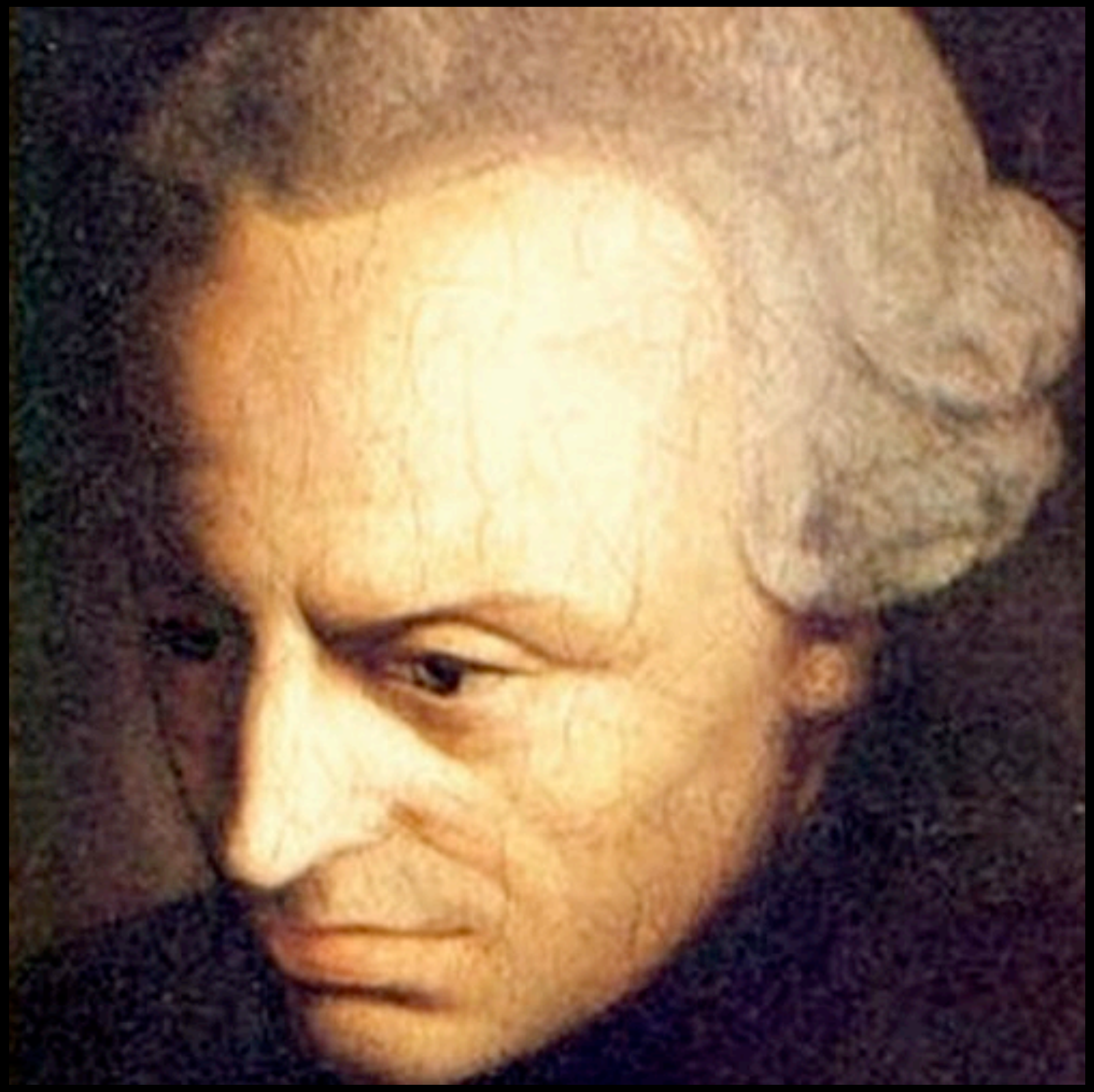


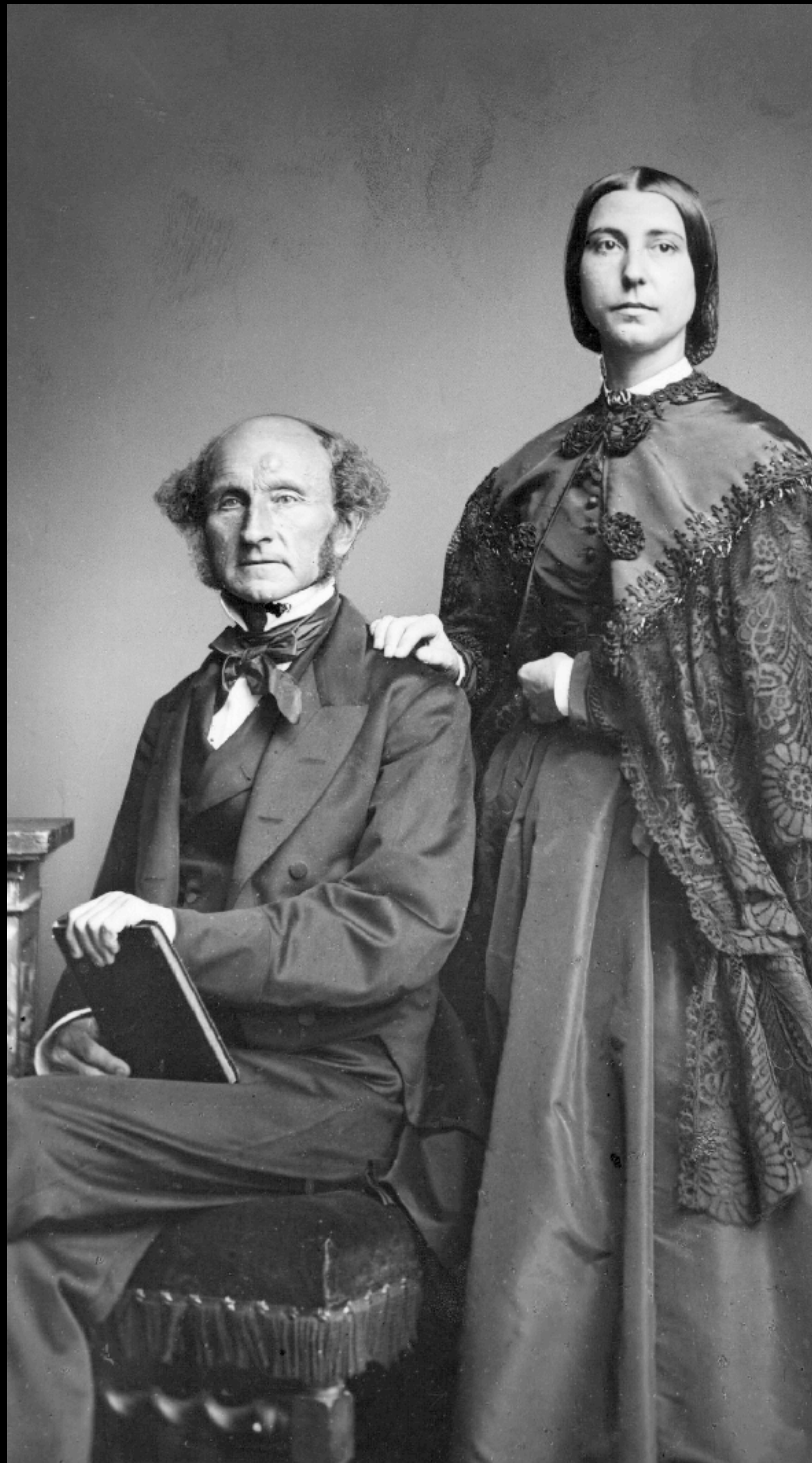
Can't always get what you want

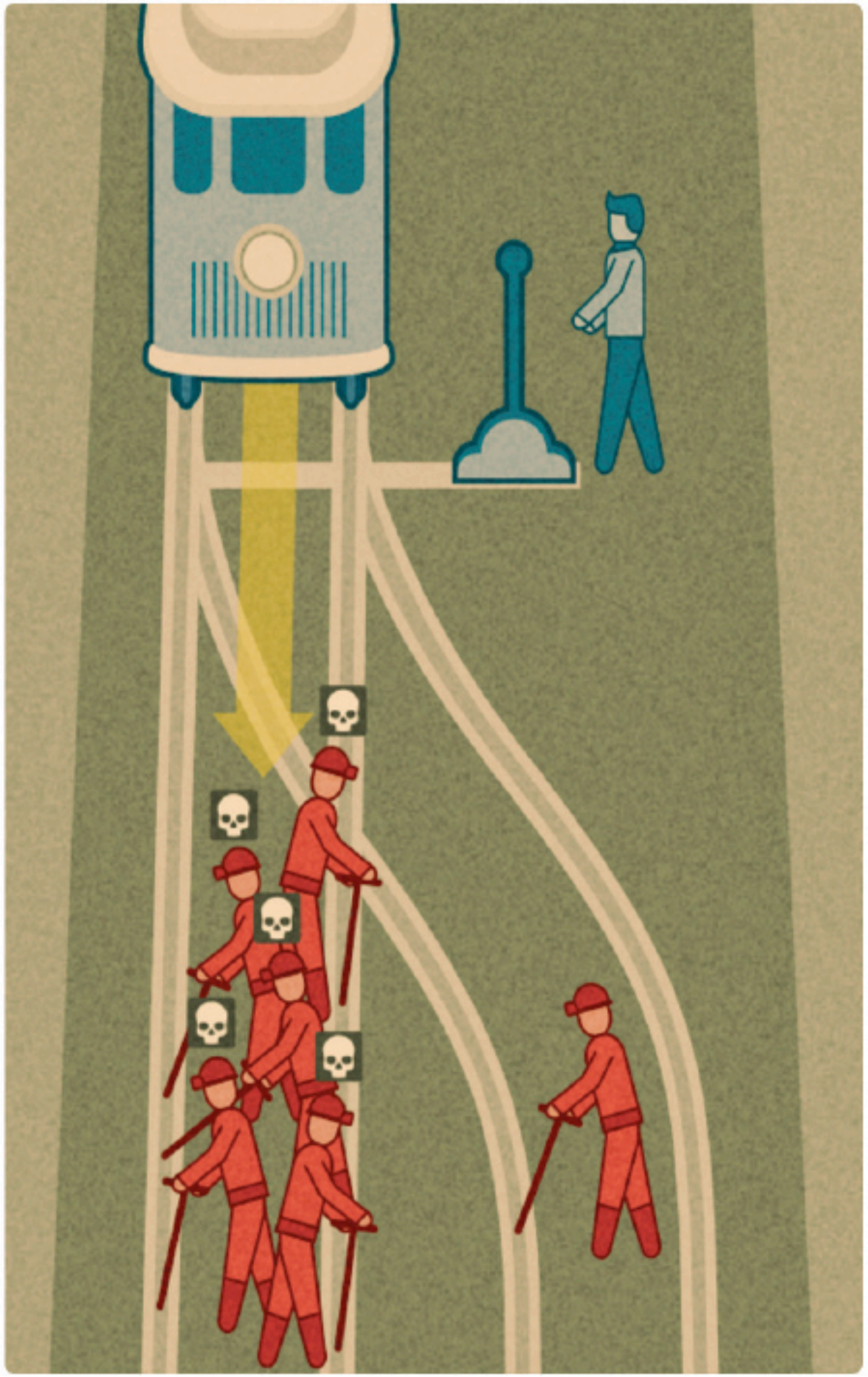


Can't always get what you want

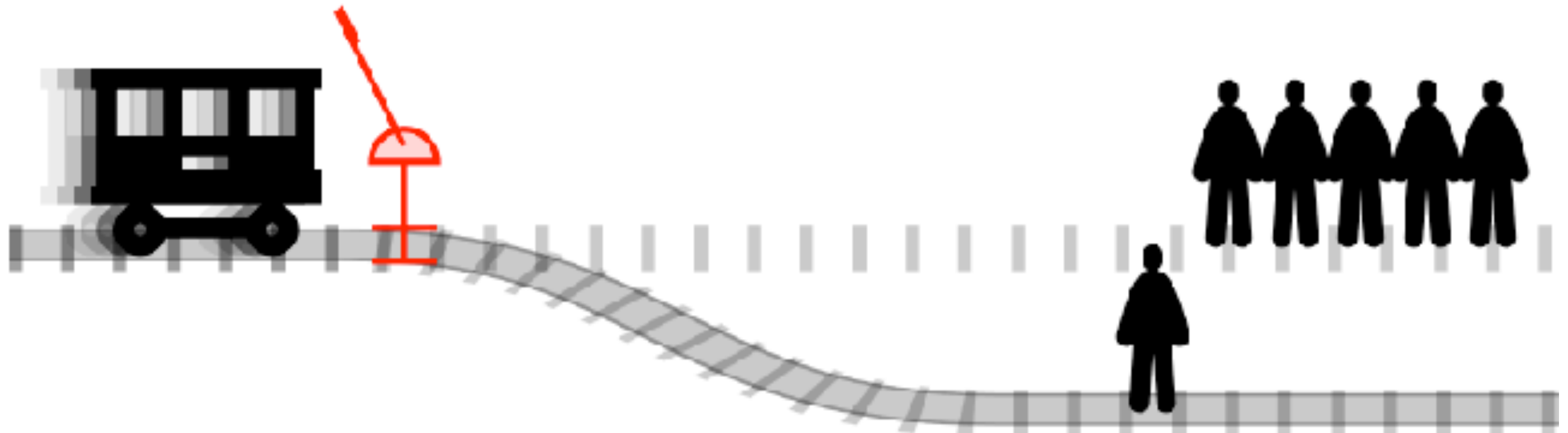








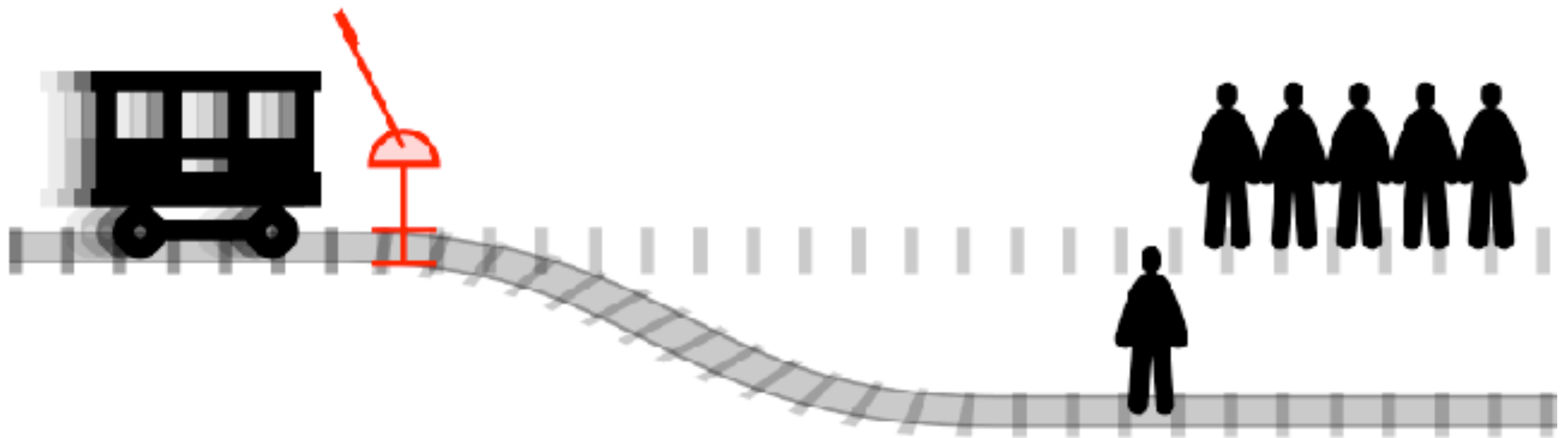
1. The switch
Foot, 1967



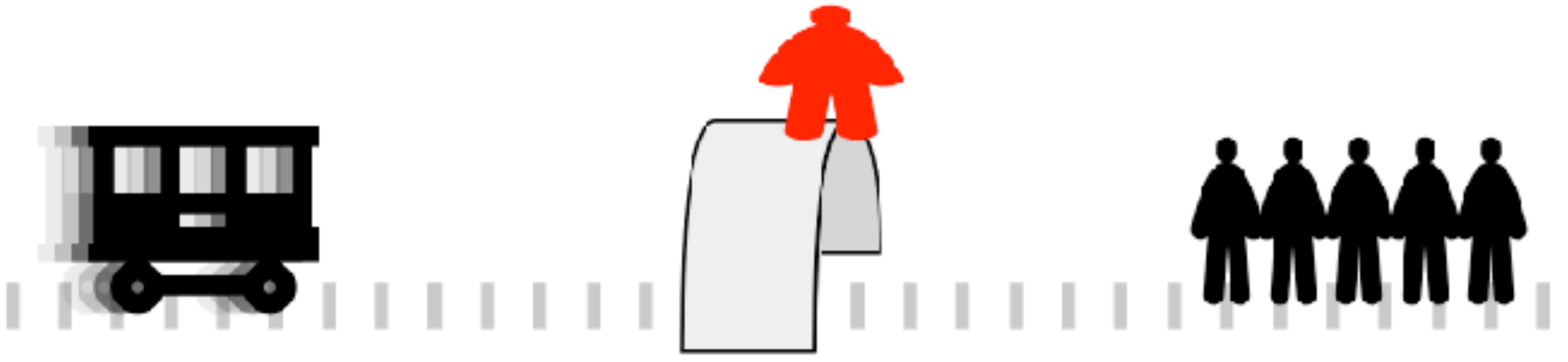
Philippa Foot (1920–2010)

Foot (1967) The Problem of Abortion and the Doctrine of Double Effect

1. The switch
Foot, 1967



2. The fat man
Thomson, 1976



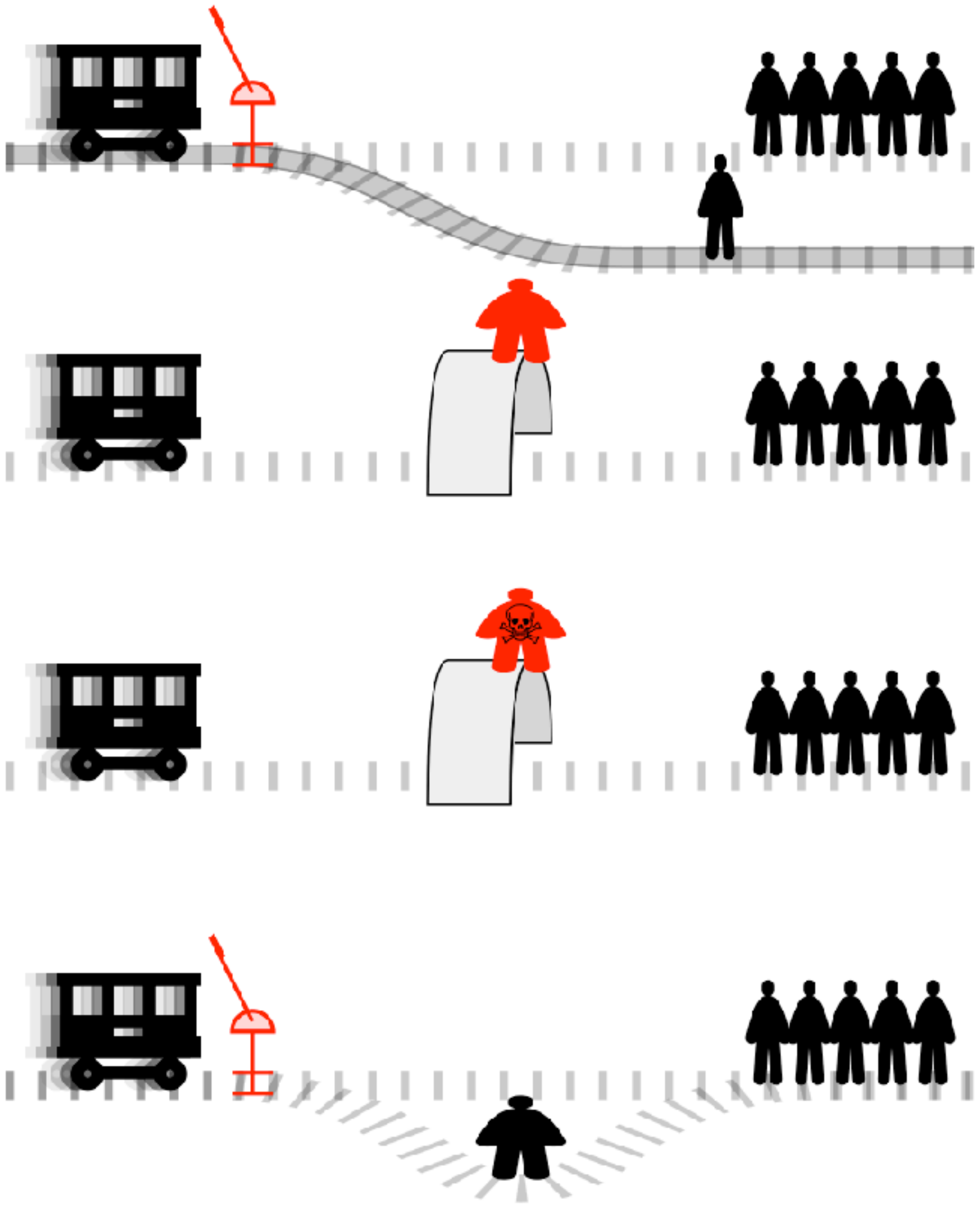
3. The fat villain

1. The switch
Foot, 1967

2. The fat man
Thomson, 1976

3. The fat villain

4. The loop
Costa, 1987



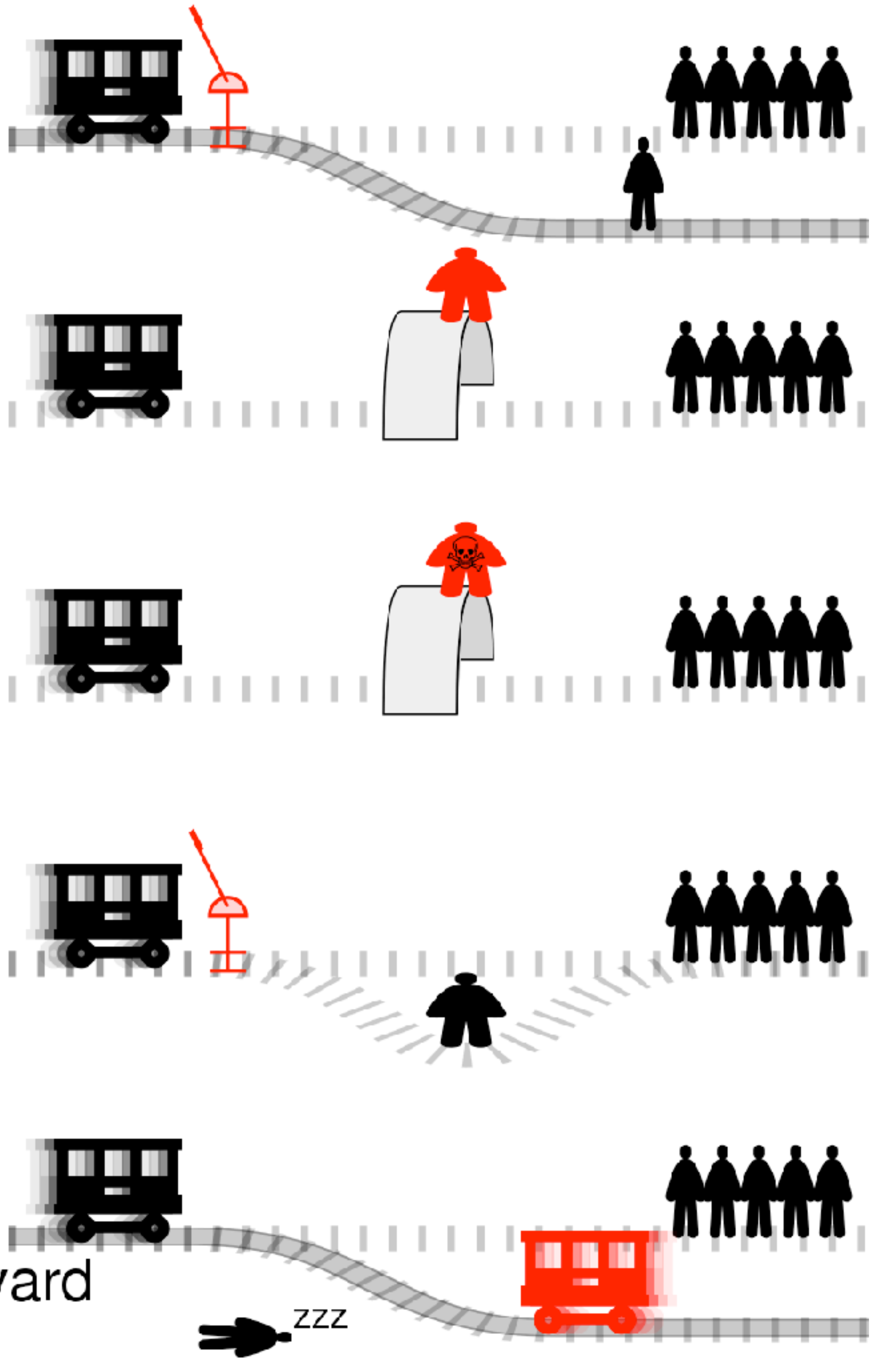
1. The switch
Foot, 1967

2. The fat man
Thomson, 1976

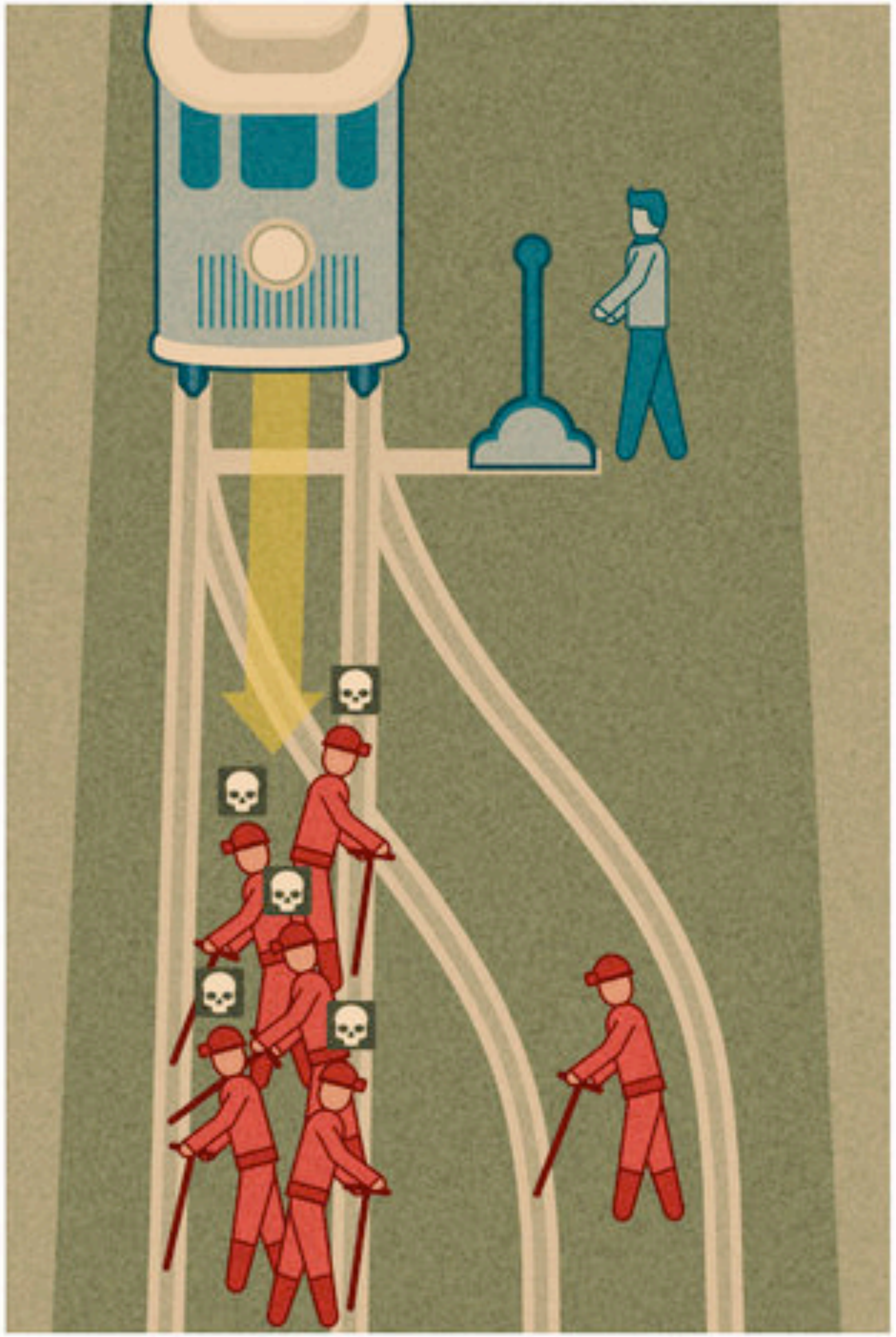
3. The fat villain

4. The loop
Costa, 1987

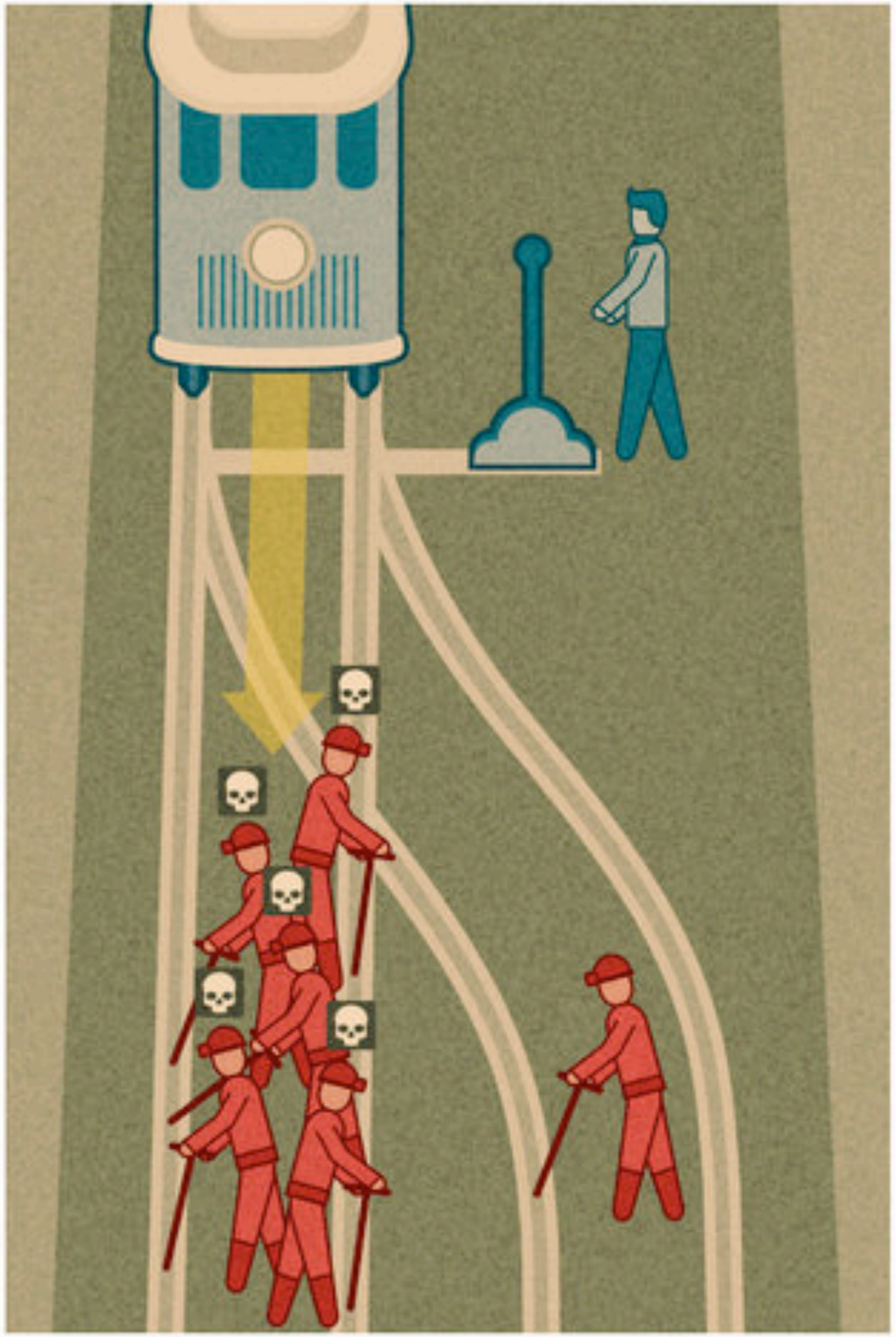
5. The man in the yard
Unger, 1992



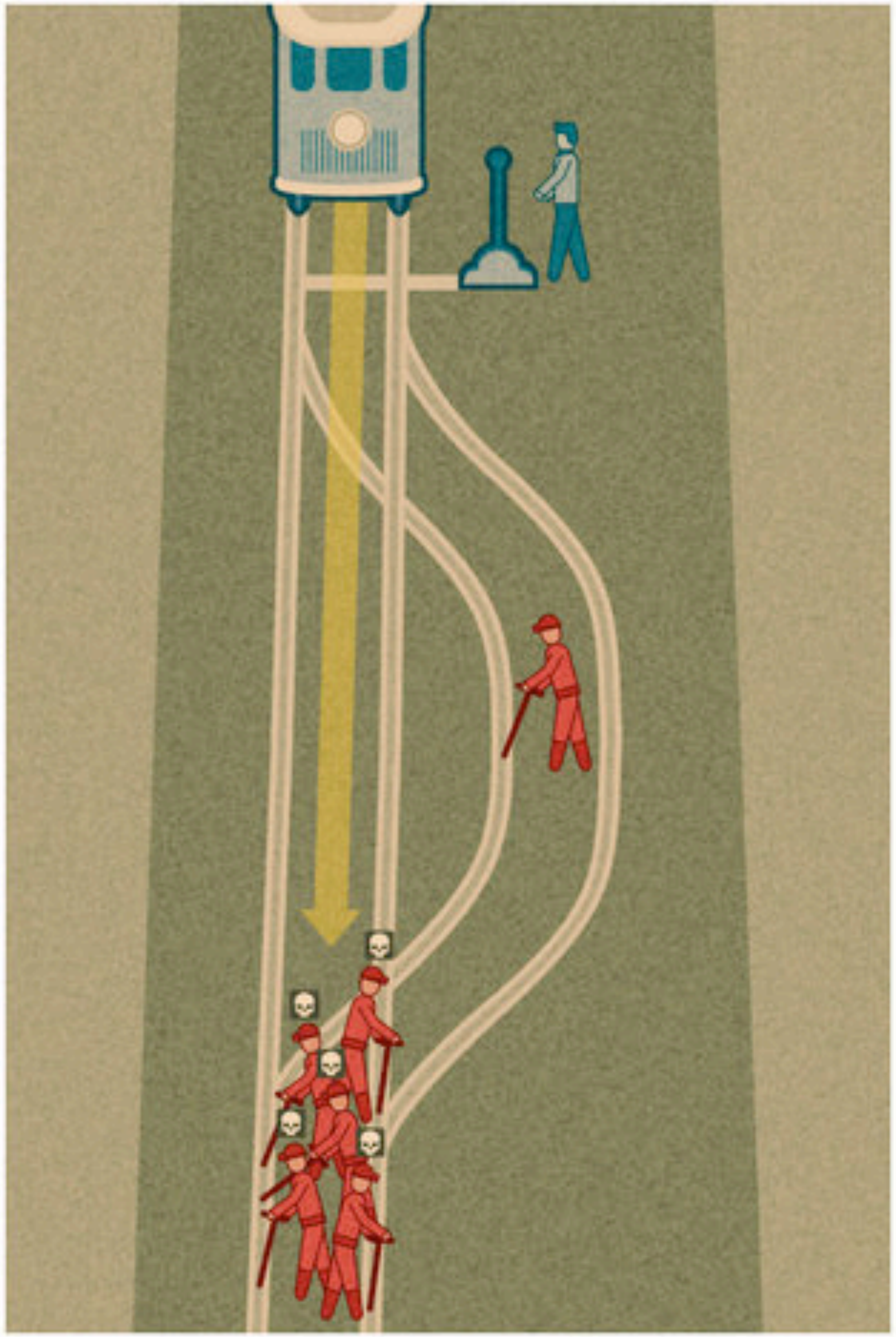
Action



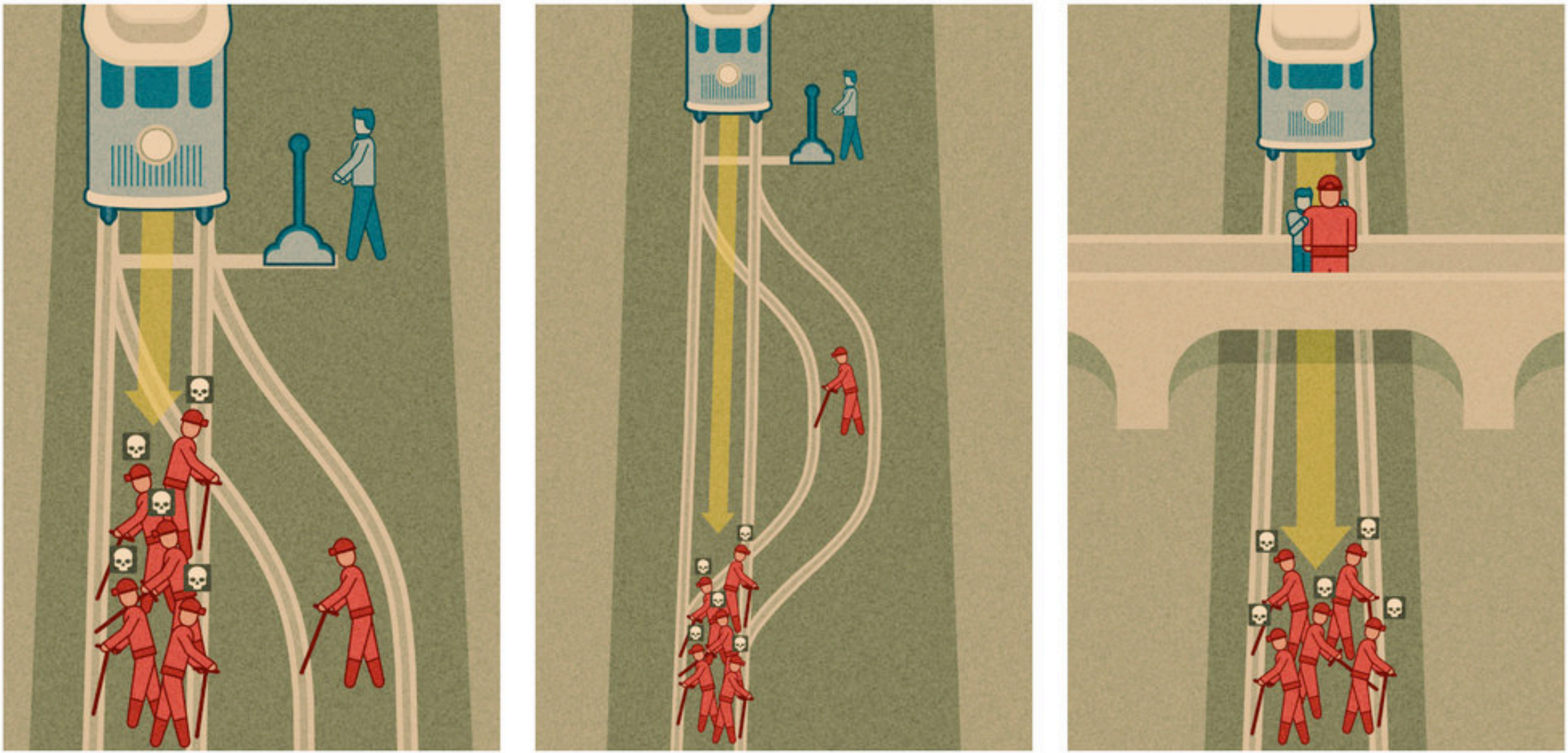
Action



Intention



Action **Intention** **Contact**



Trolley Problems

data(Trolley)

331 individuals (age, gender, edu)

Voluntary participation (online)

30 different trolley problems

action / intention / contact

9930 responses:
How appropriate (from 1 to 7)?

Cushman et al. 2006. The role of conscious reasoning and intuition in moral judgment

Trolley Problems

data(Trolley)

331 individuals (age, gender, edu)

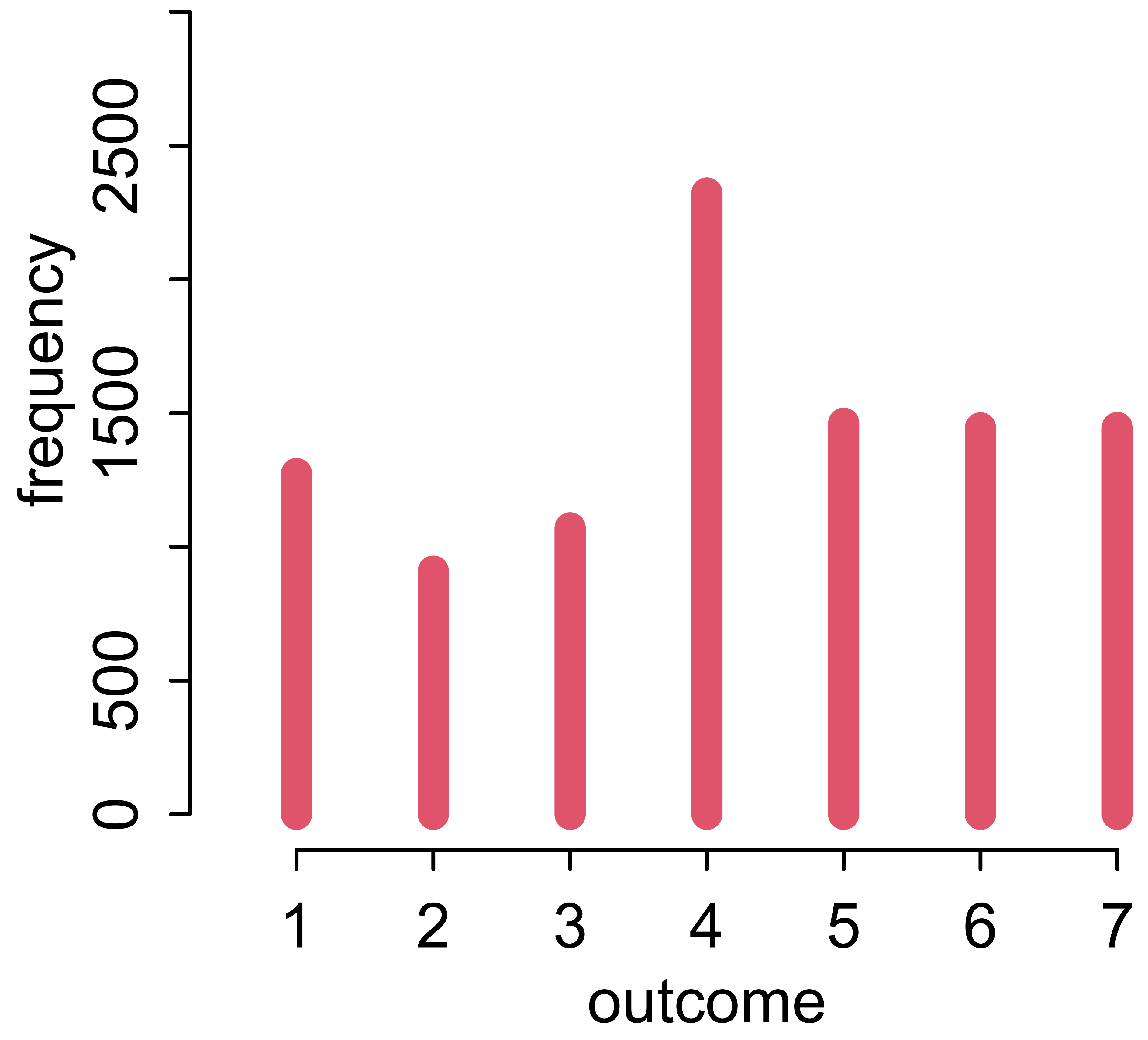
Voluntary participation (online)

30 different trolley problems

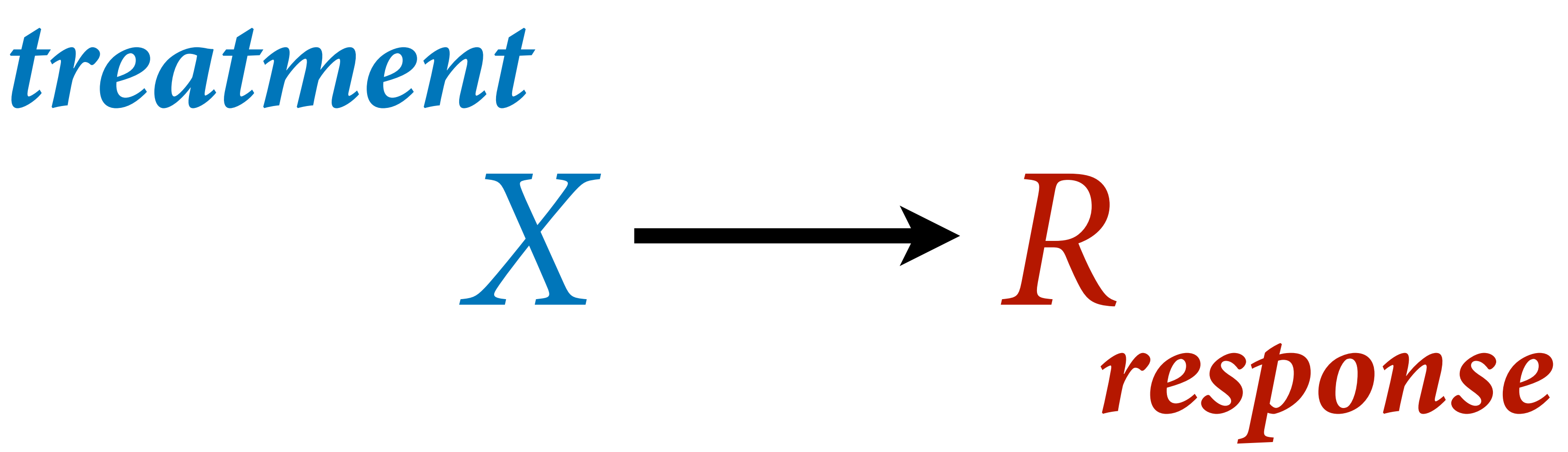
action / intention / contact

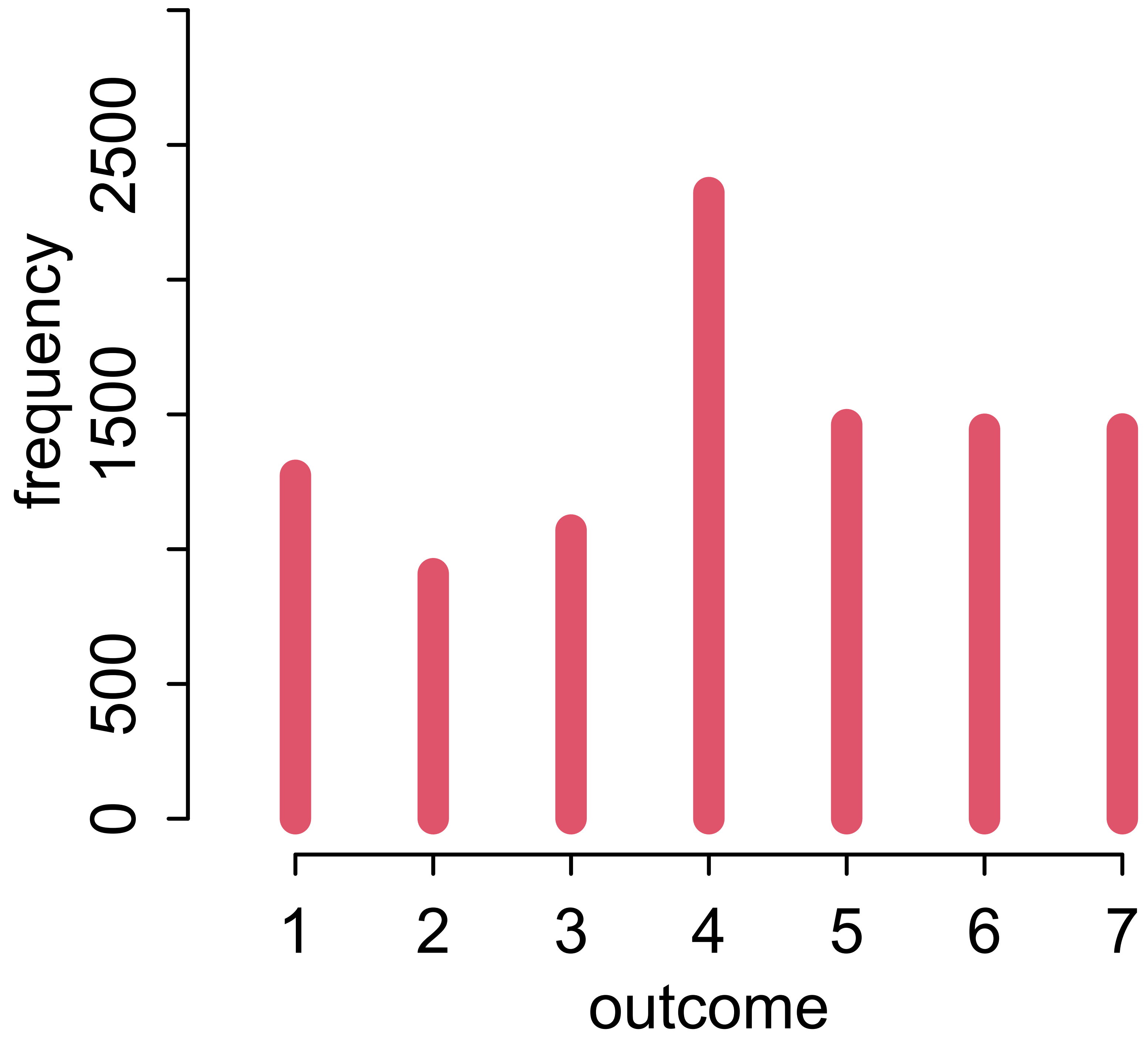
9930 responses:
How appropriate (from 1 to 7)?

Ordered categorical



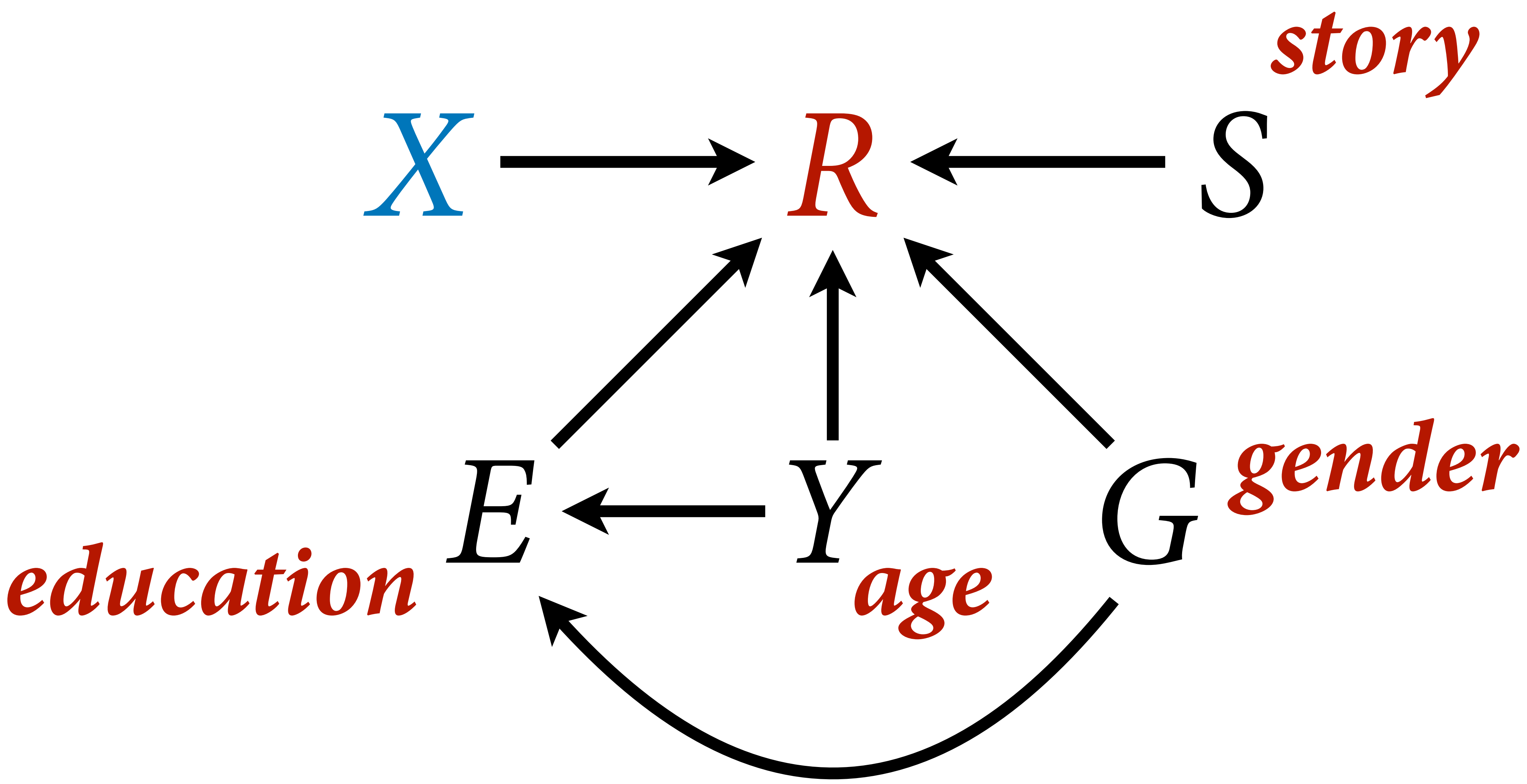
Estimand: How do **action**, **intention**, **contact** influence **response** to a trolley story?

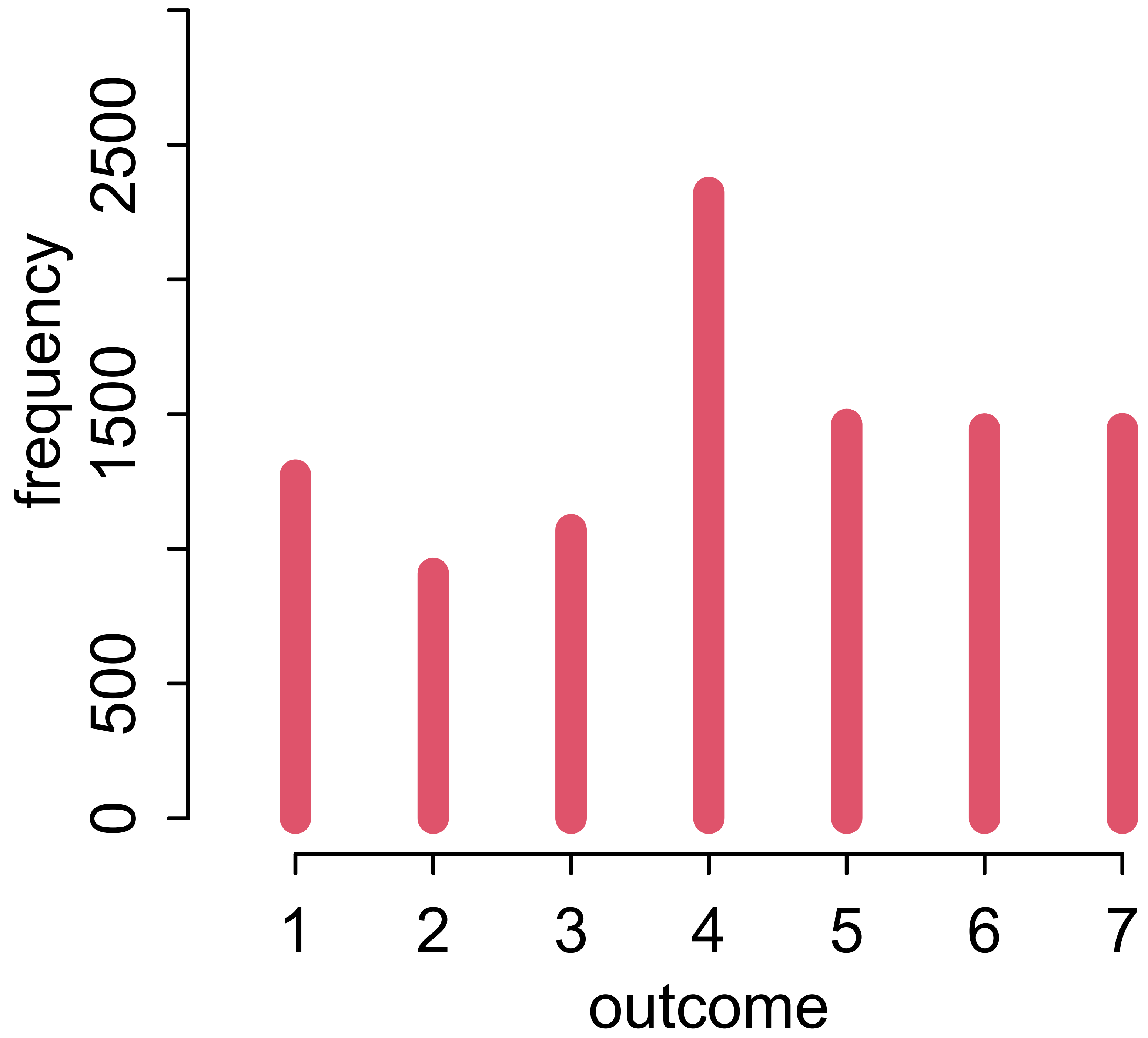




Estimand: How do **action**, **intention**, **contact** influence **response** to a trolley story?

How are influences of A/I/C associated with other variables?





Ordered categories

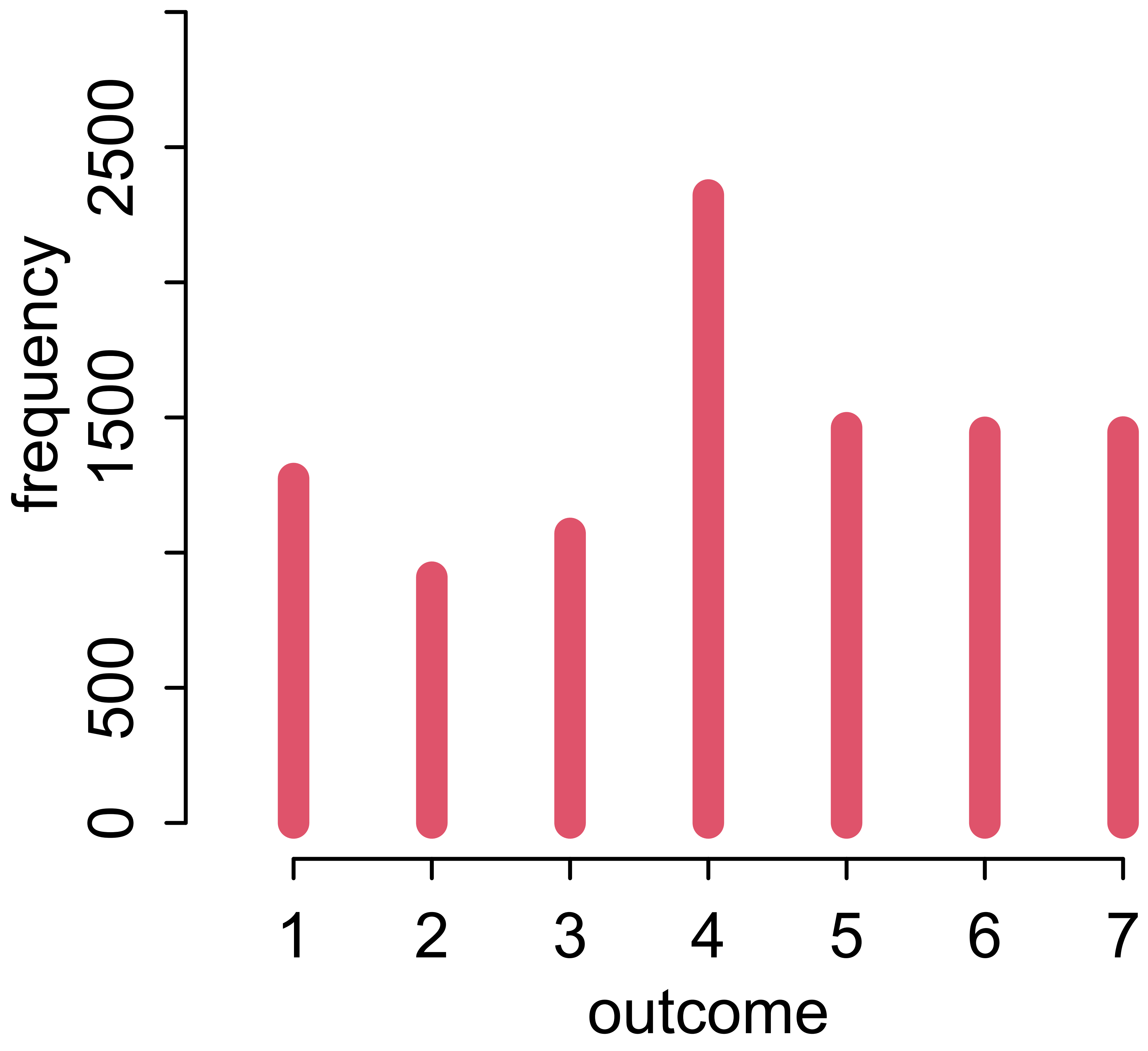
Categories: Discrete types

cat, dog, chicken

Ordered categories: Discrete types with ordered relationships

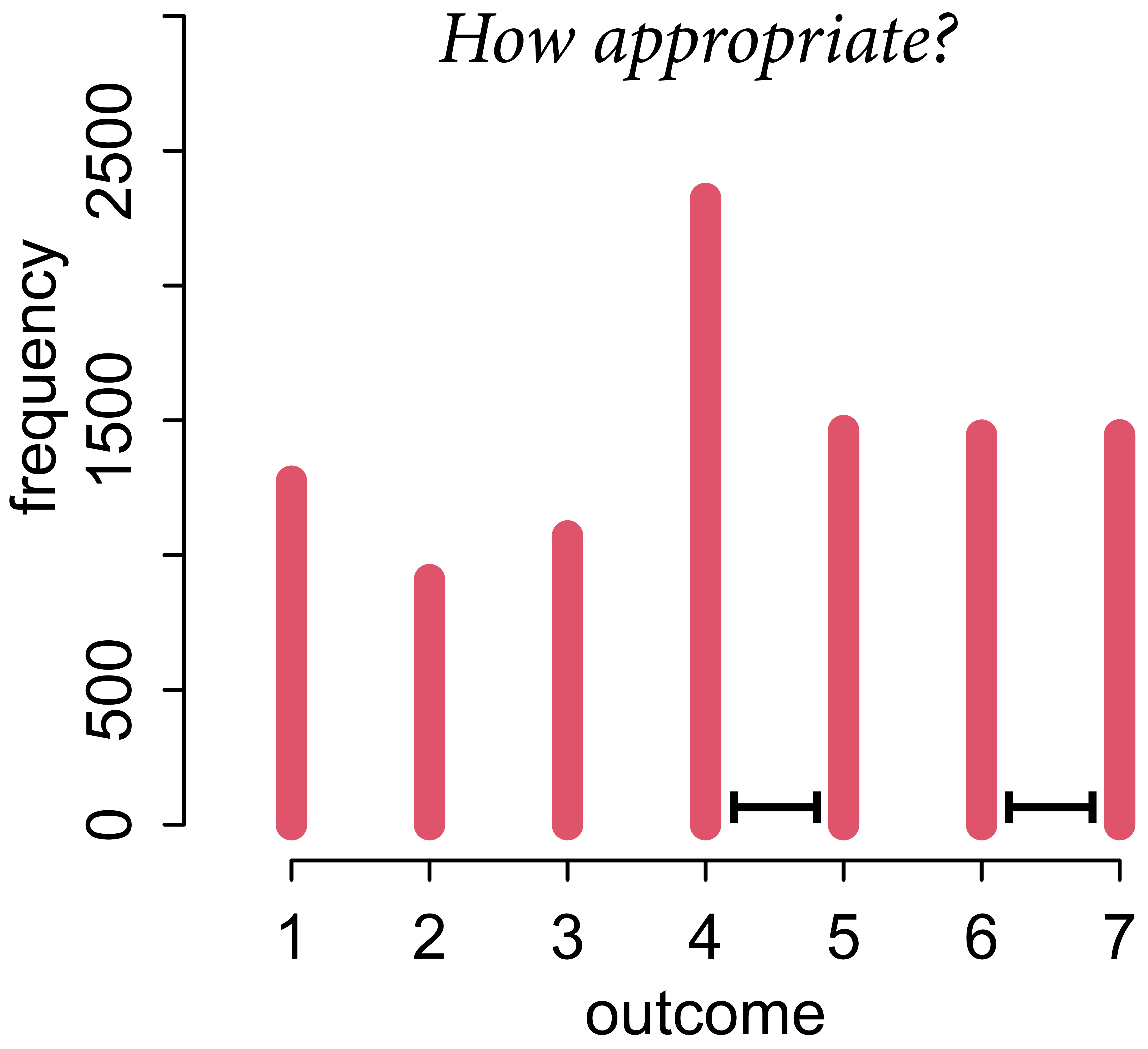
bad, good, excellent

Ordered categorical



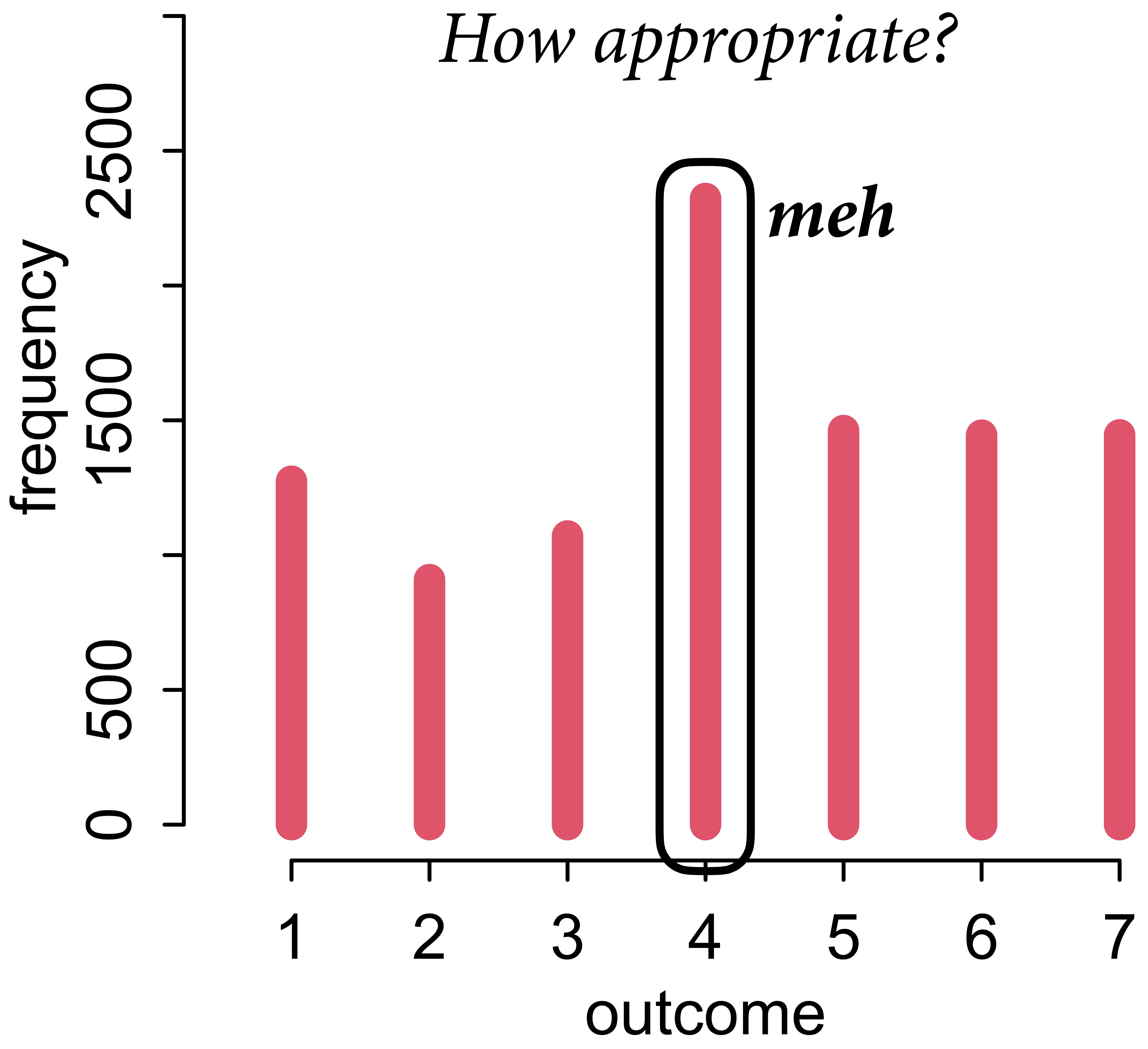
Distance between values not constant

Probably much easier to go from 4 to 5 than from 6 to 7

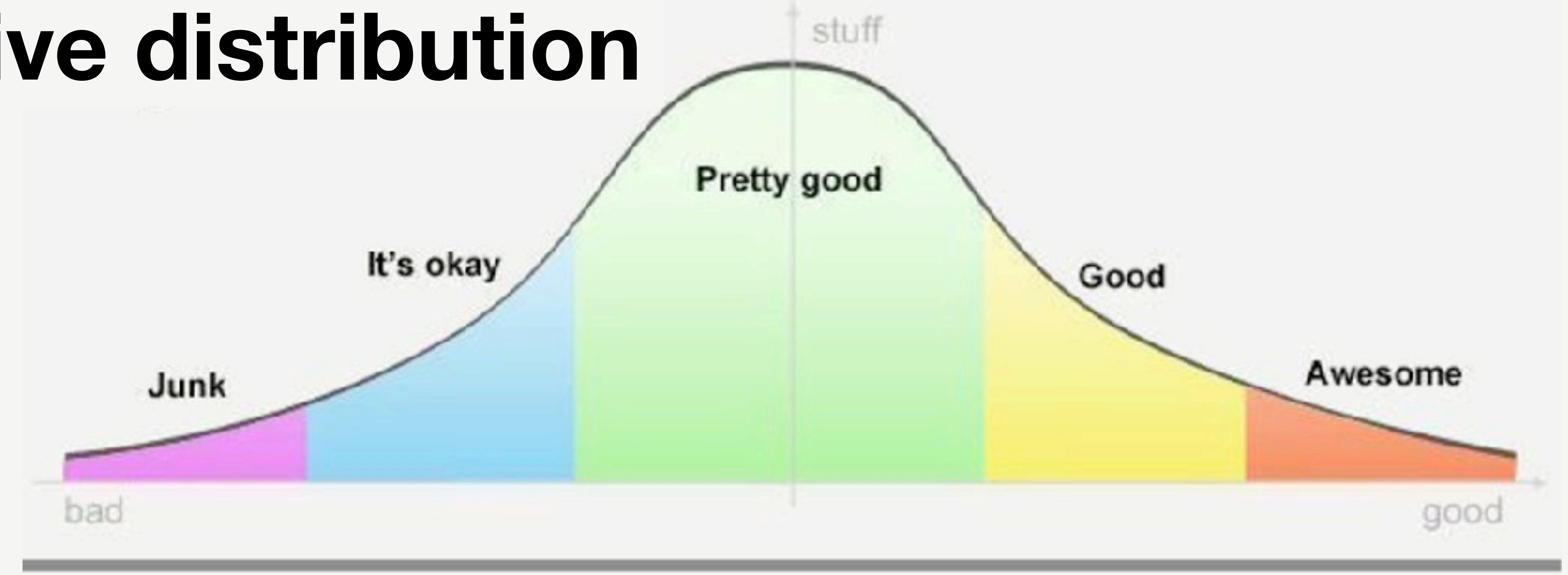


Anchor points common

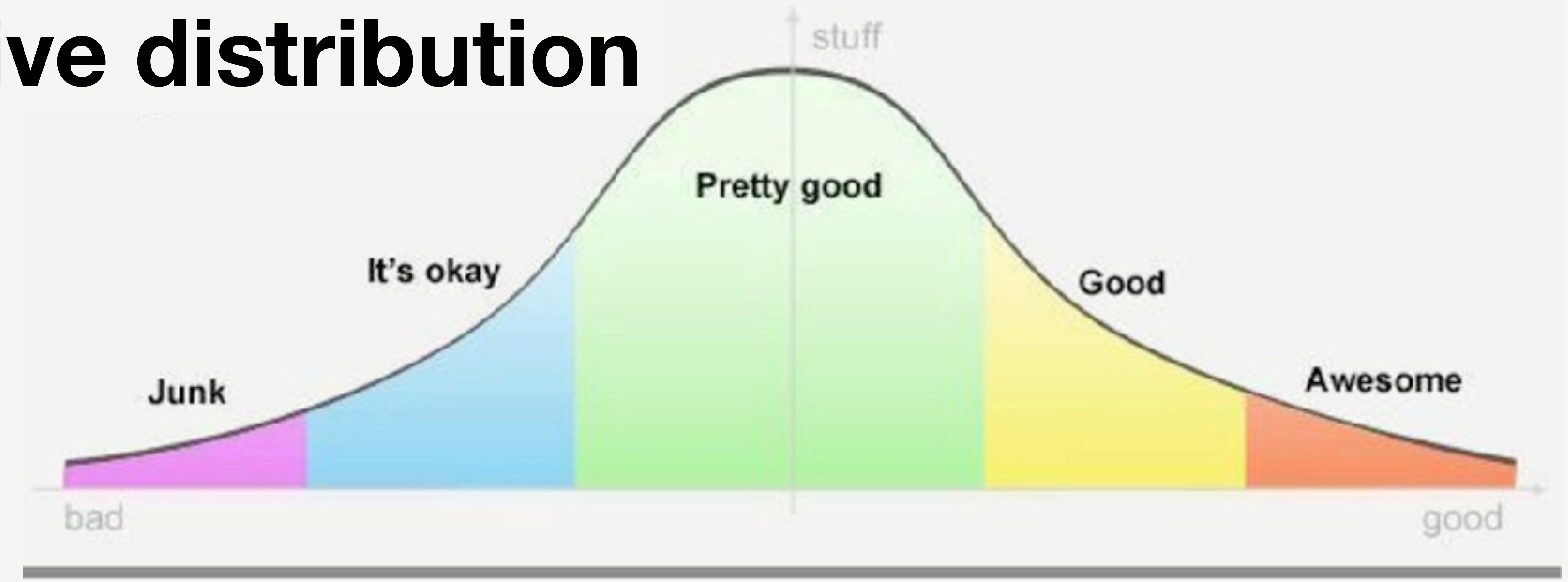
Not everyone shares the same anchor points



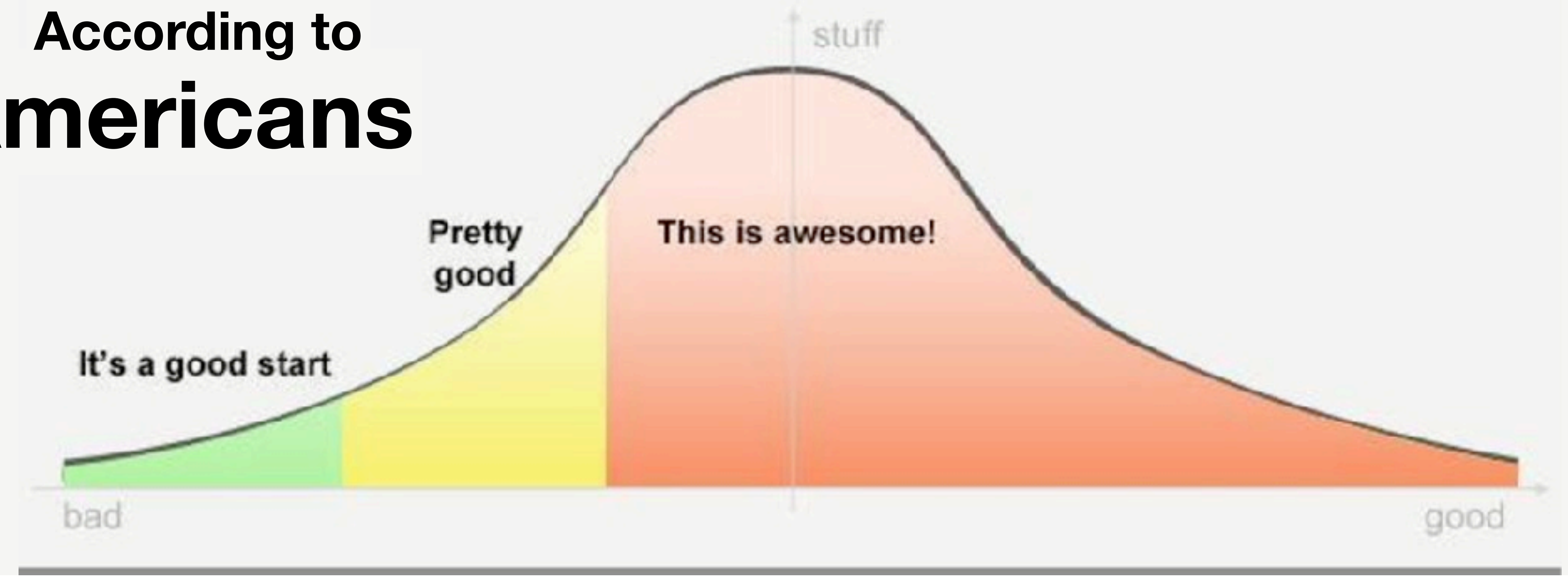
Objective distribution

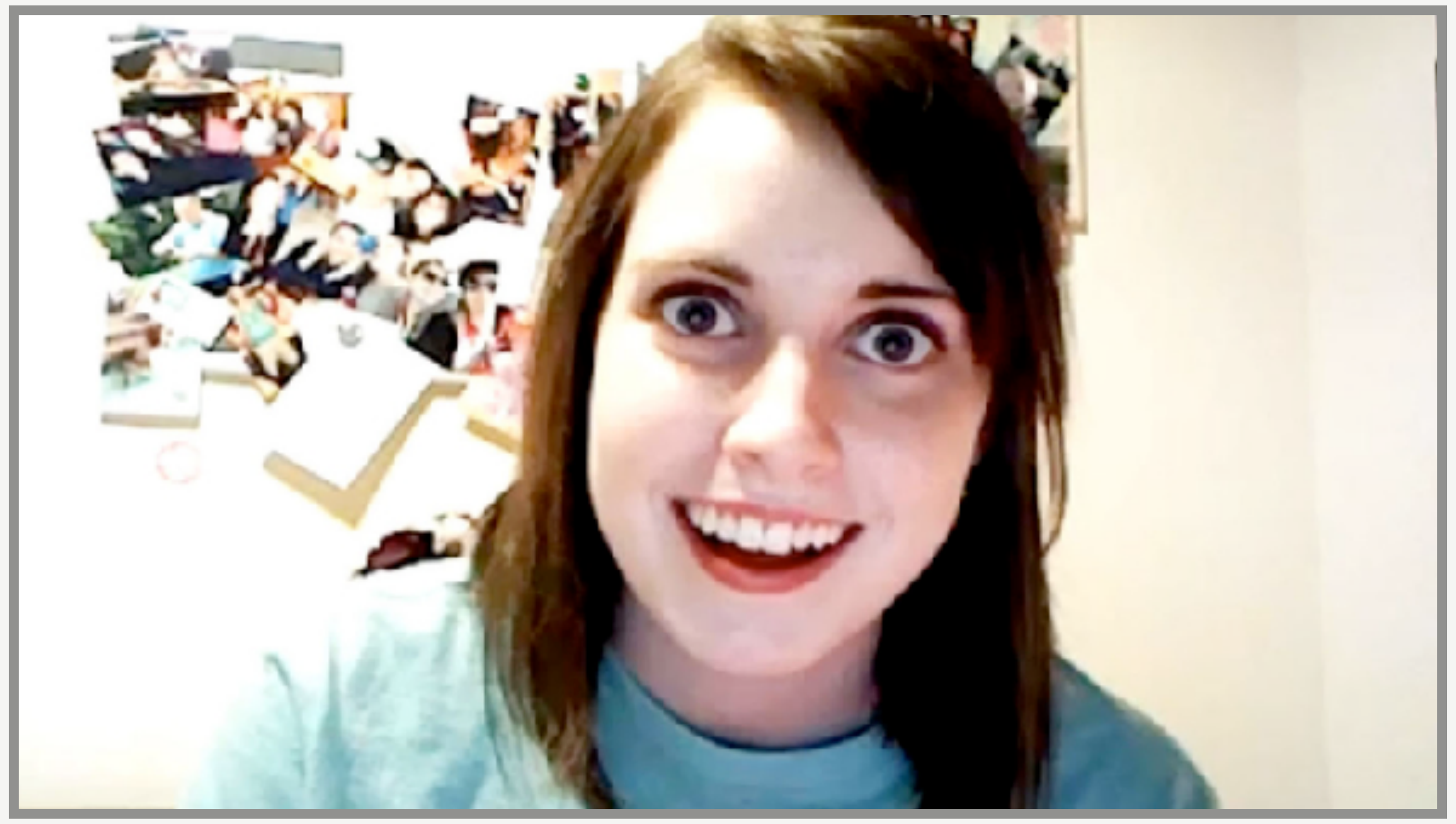


Objective distribution

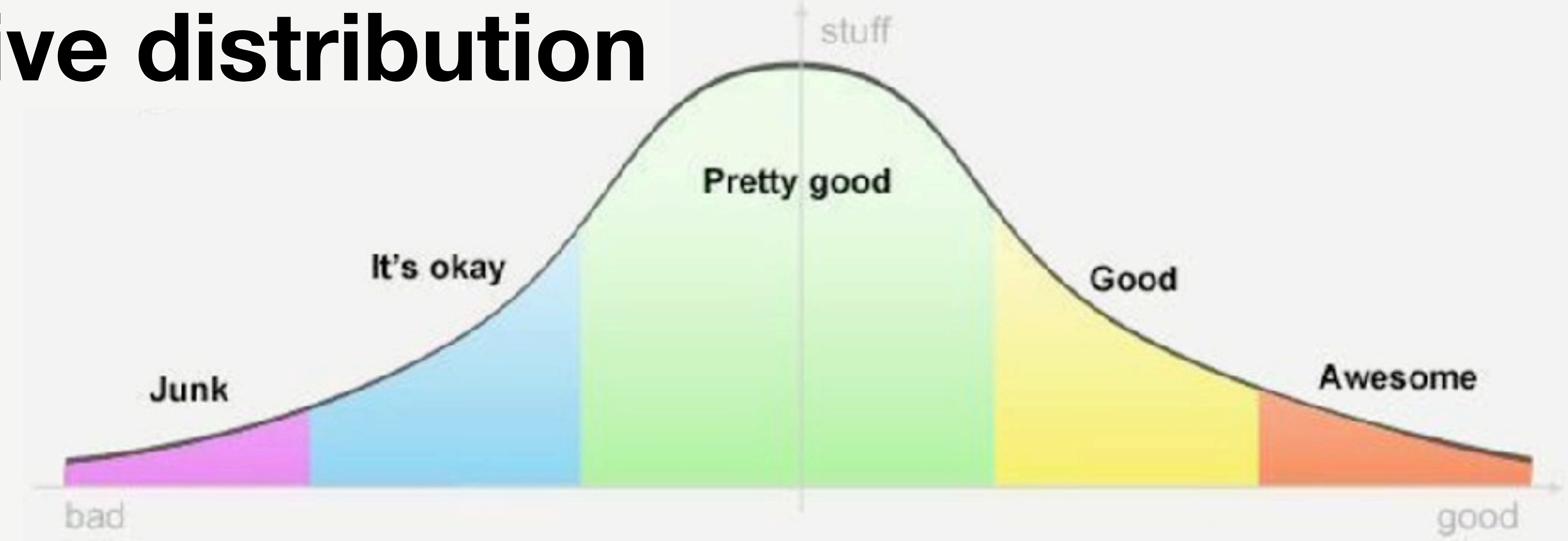


According to
Americans

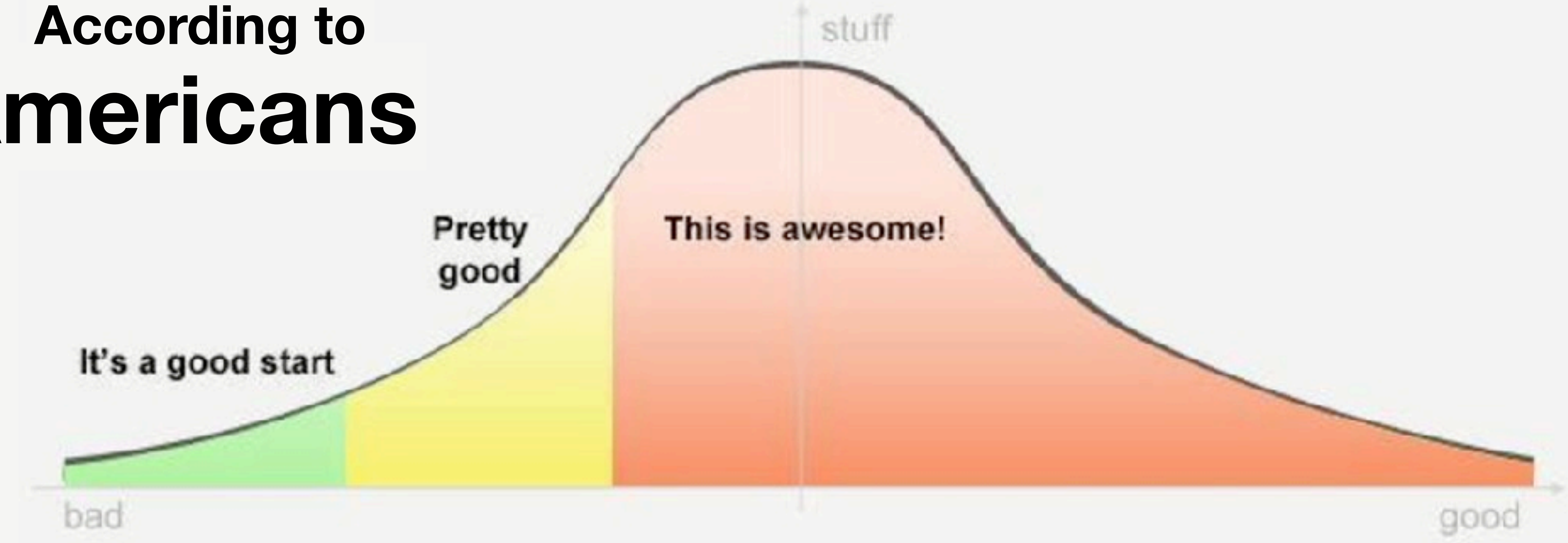


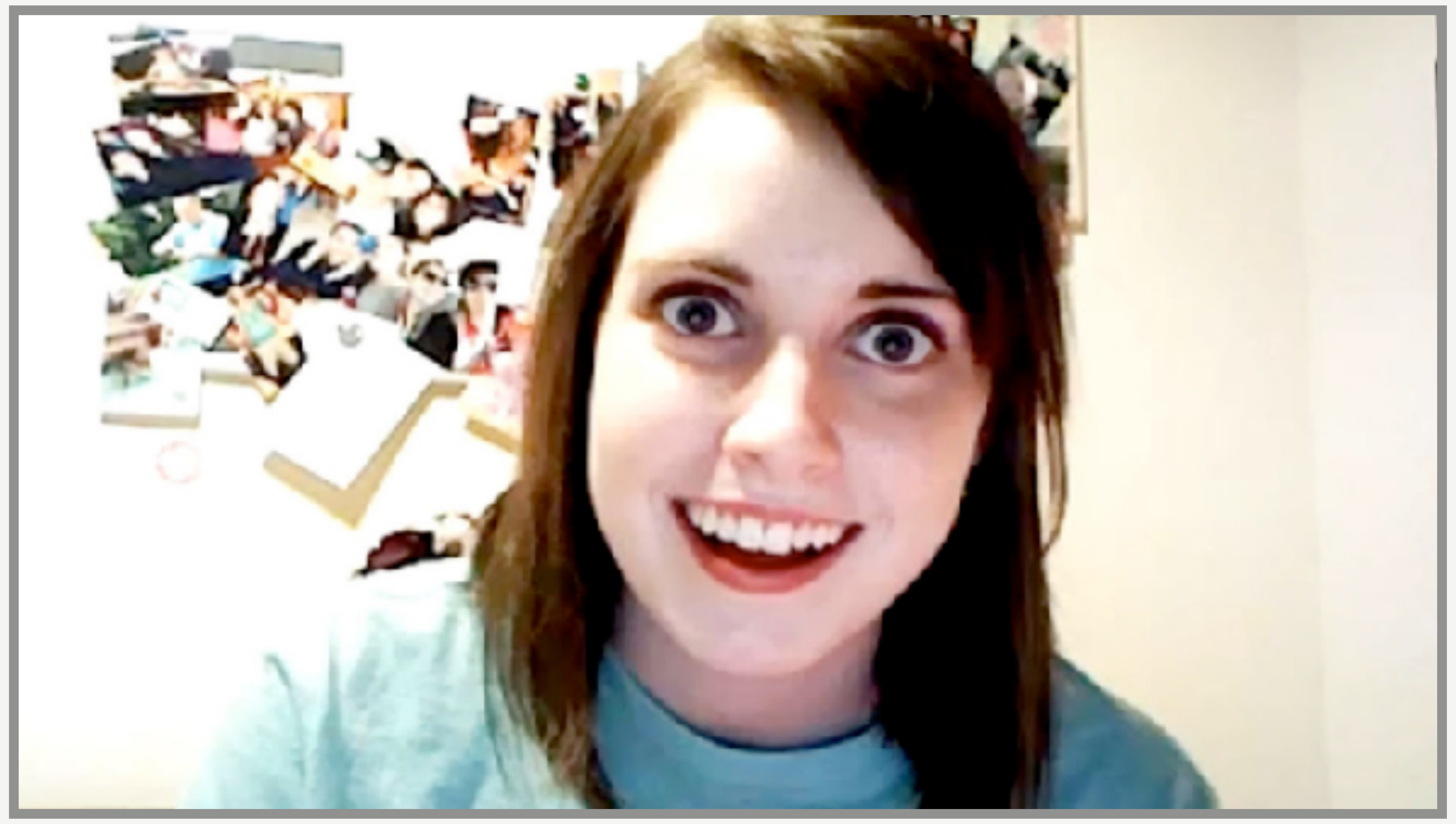


Objective distribution

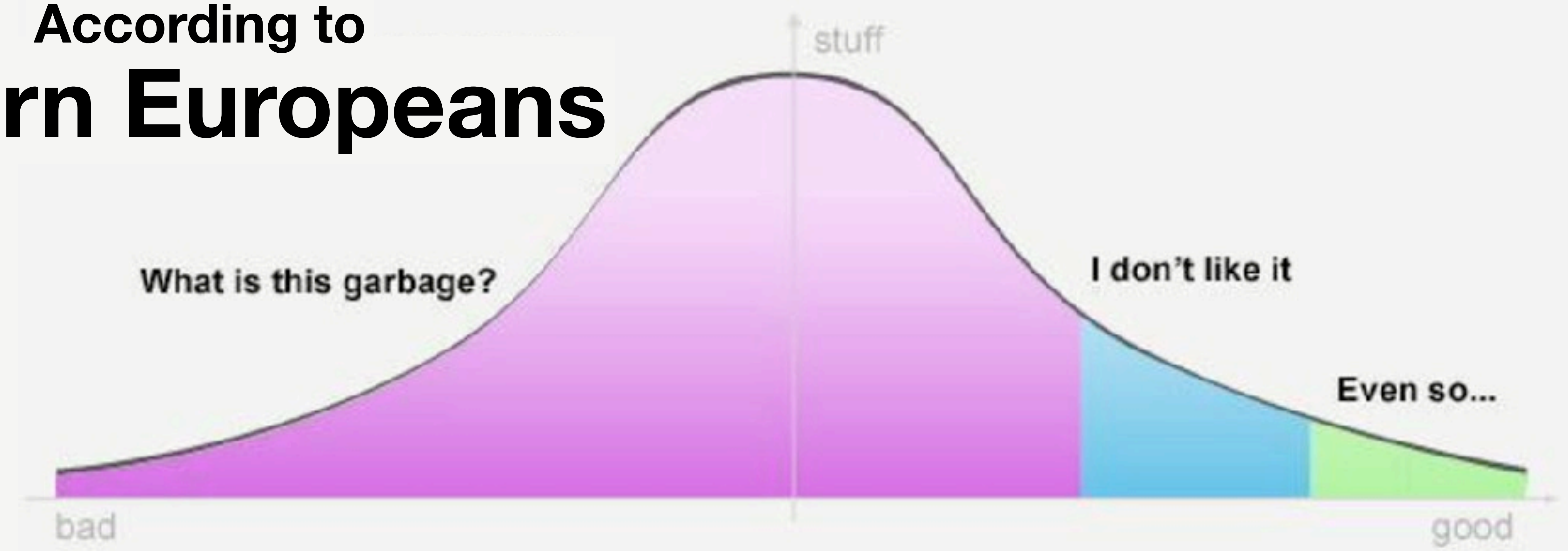


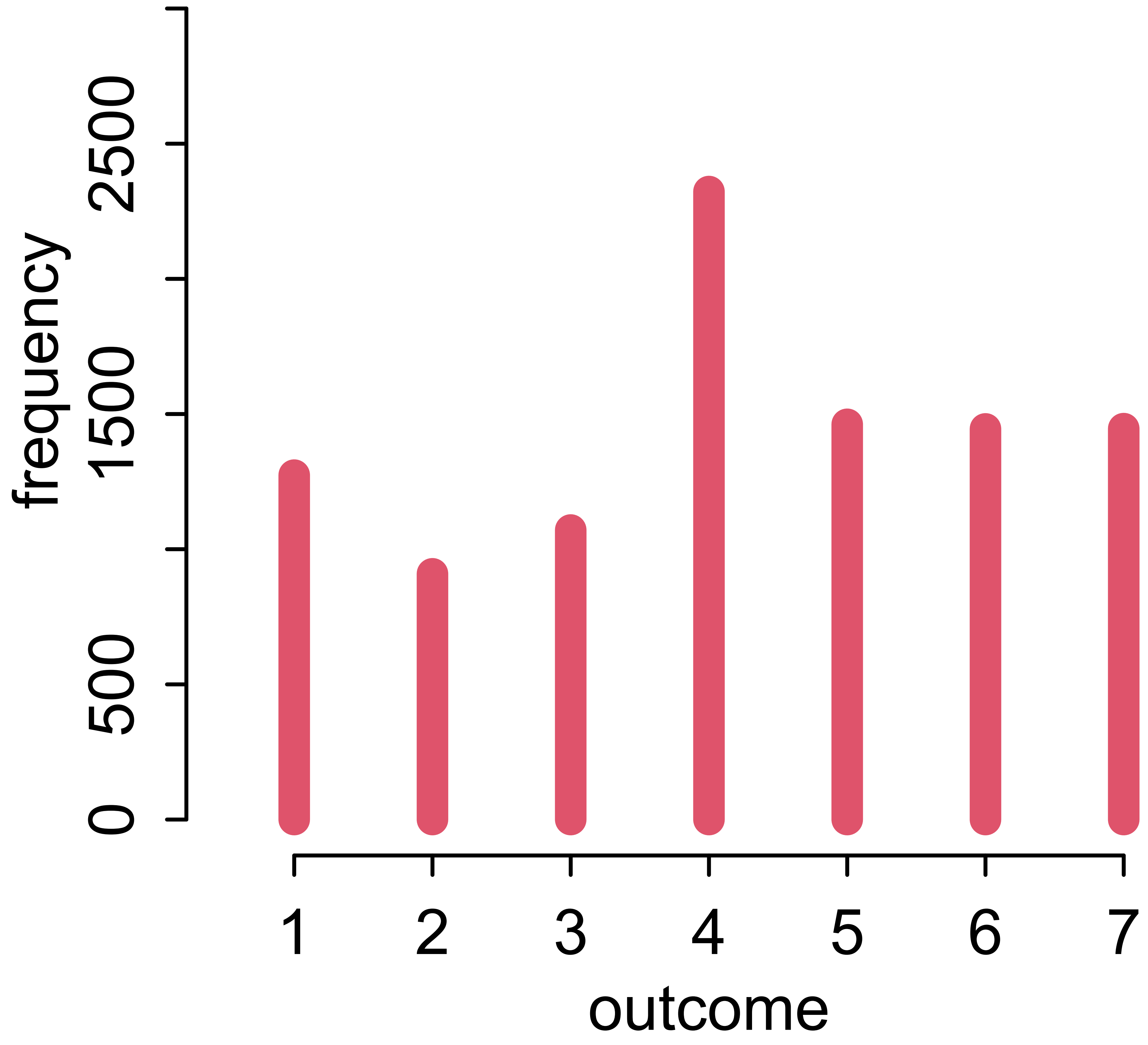
According to
Americans

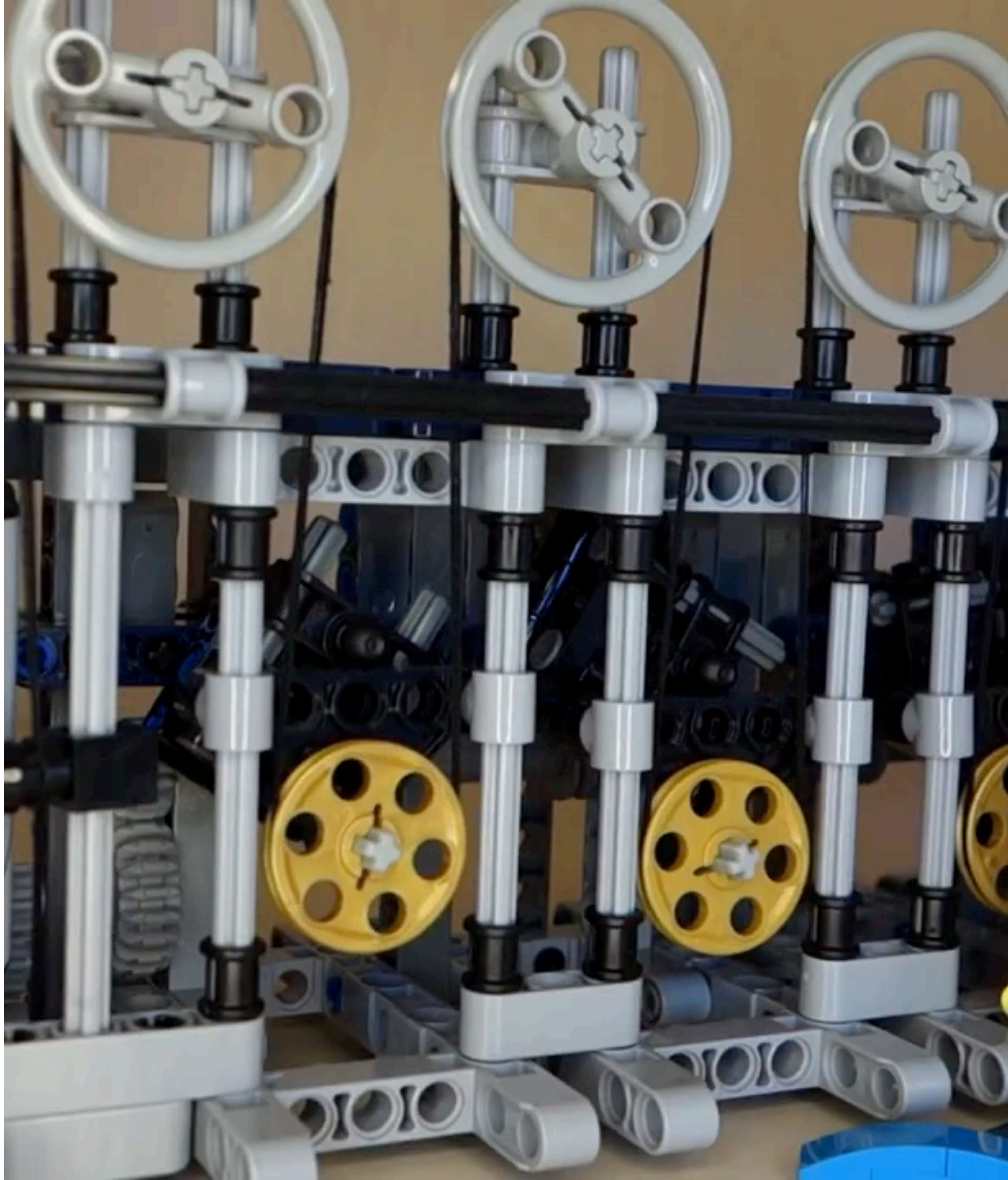




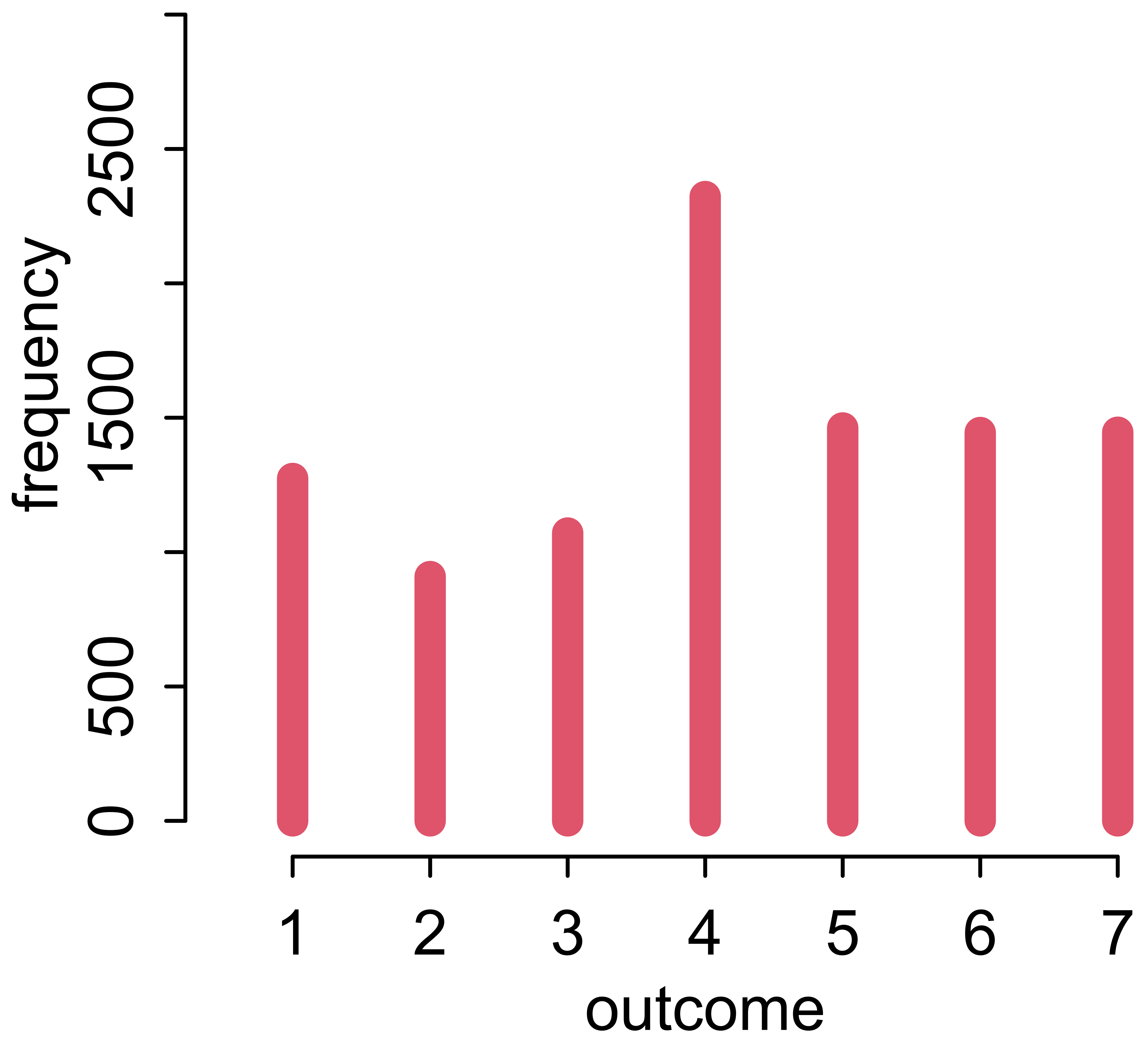
According to
Eastern Europeans

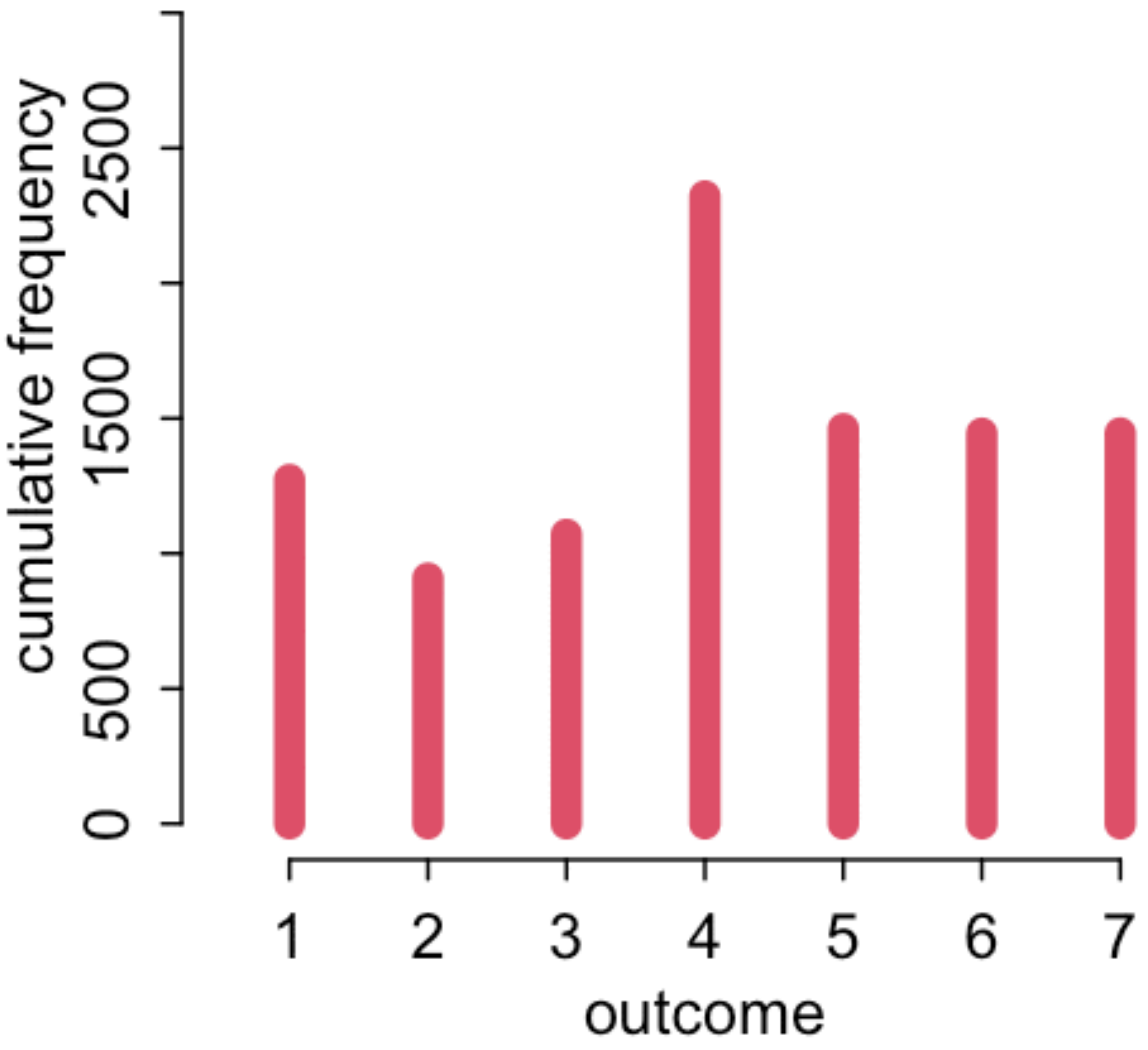


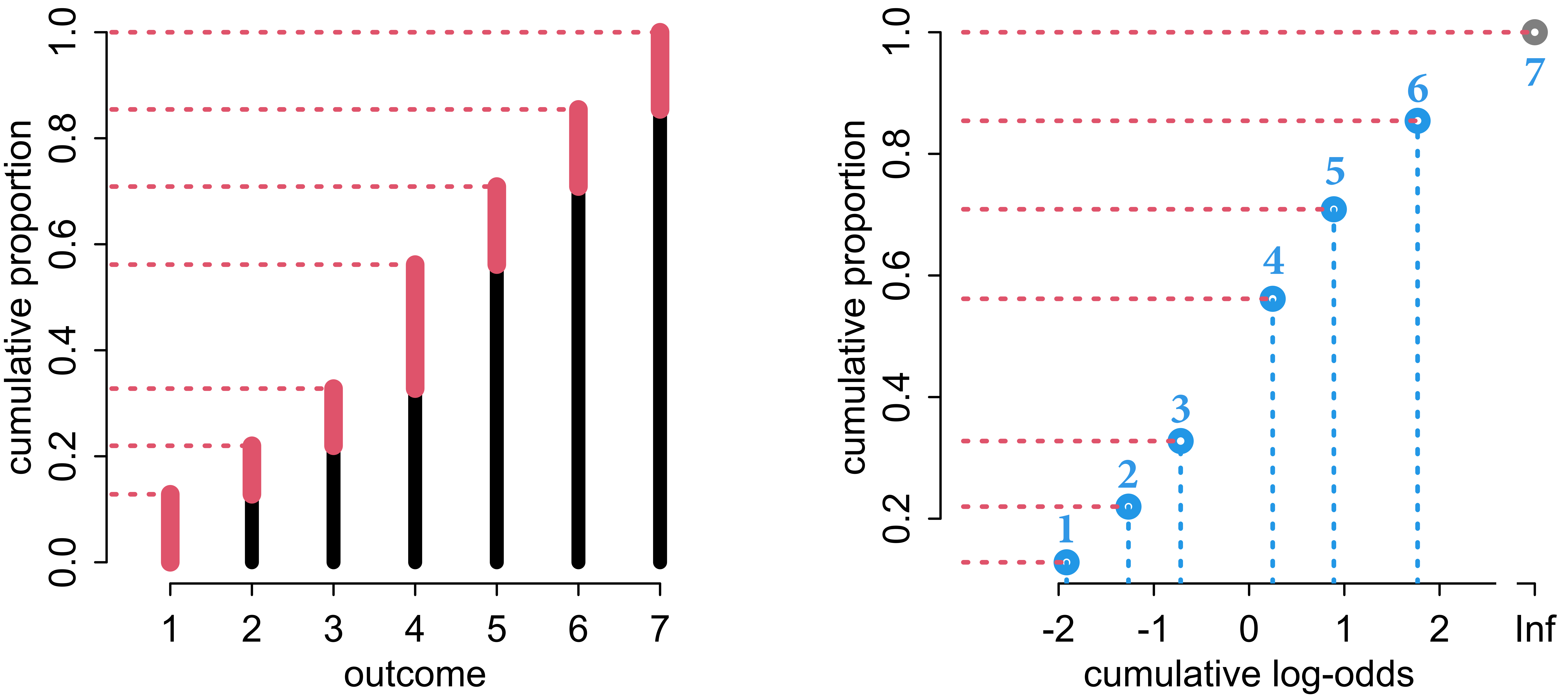

frequency
0
500
1500
2500
1
2
3
4
5
6
7
outcome

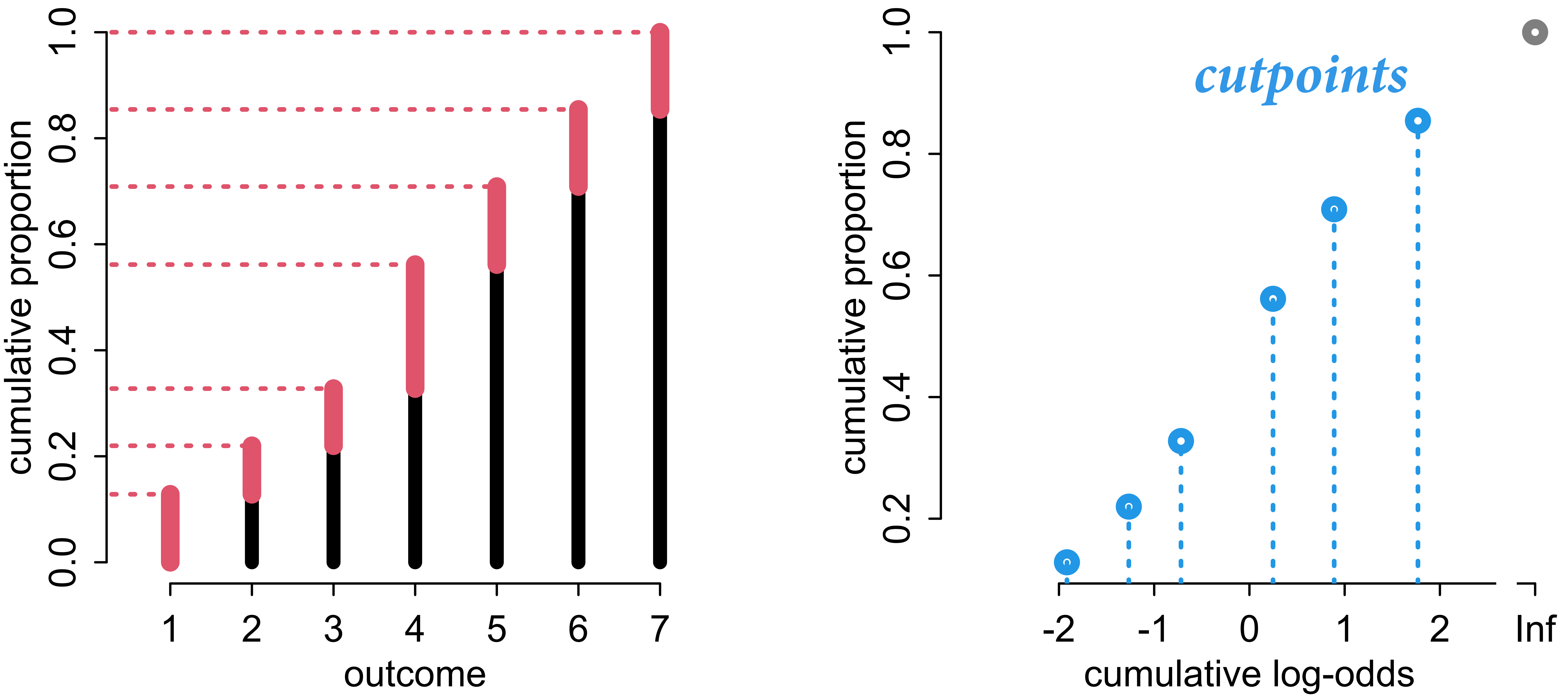


Ordered = Cumulative

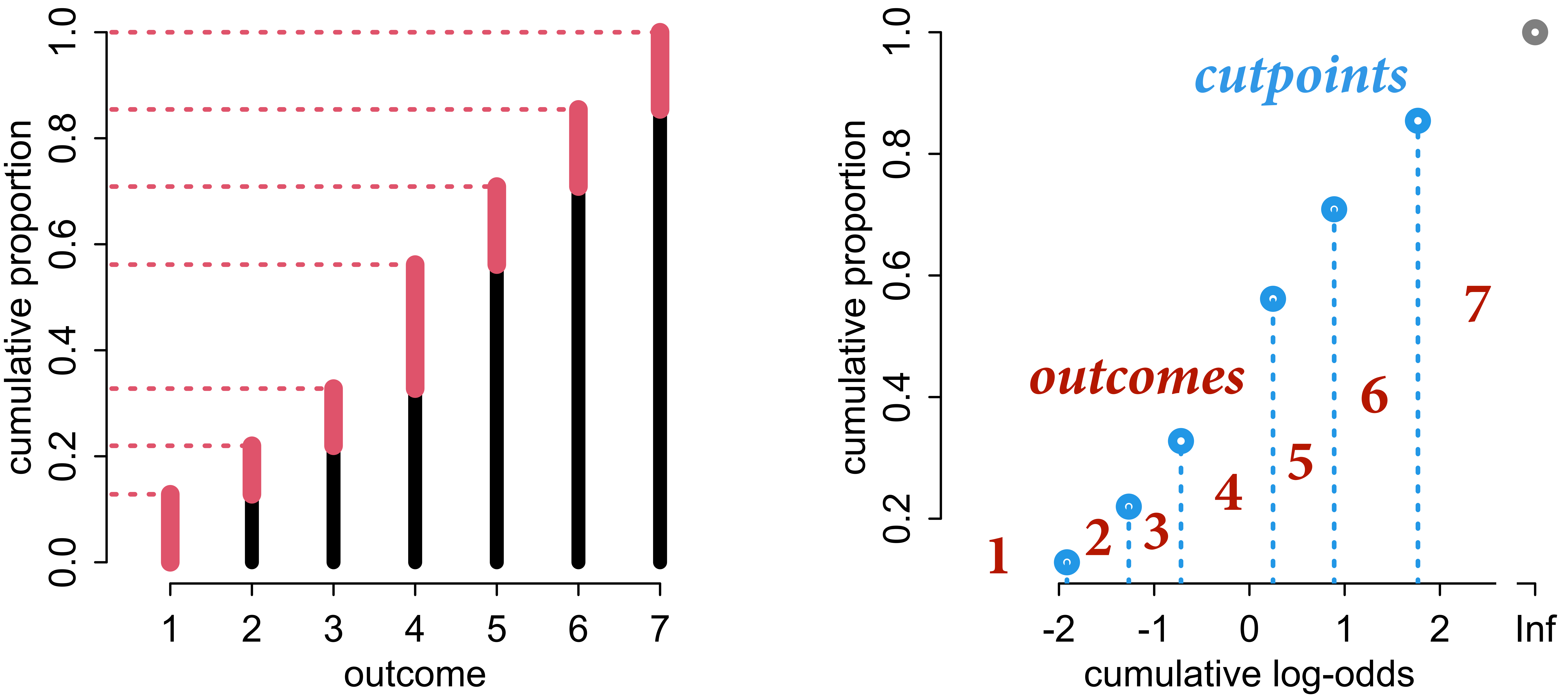




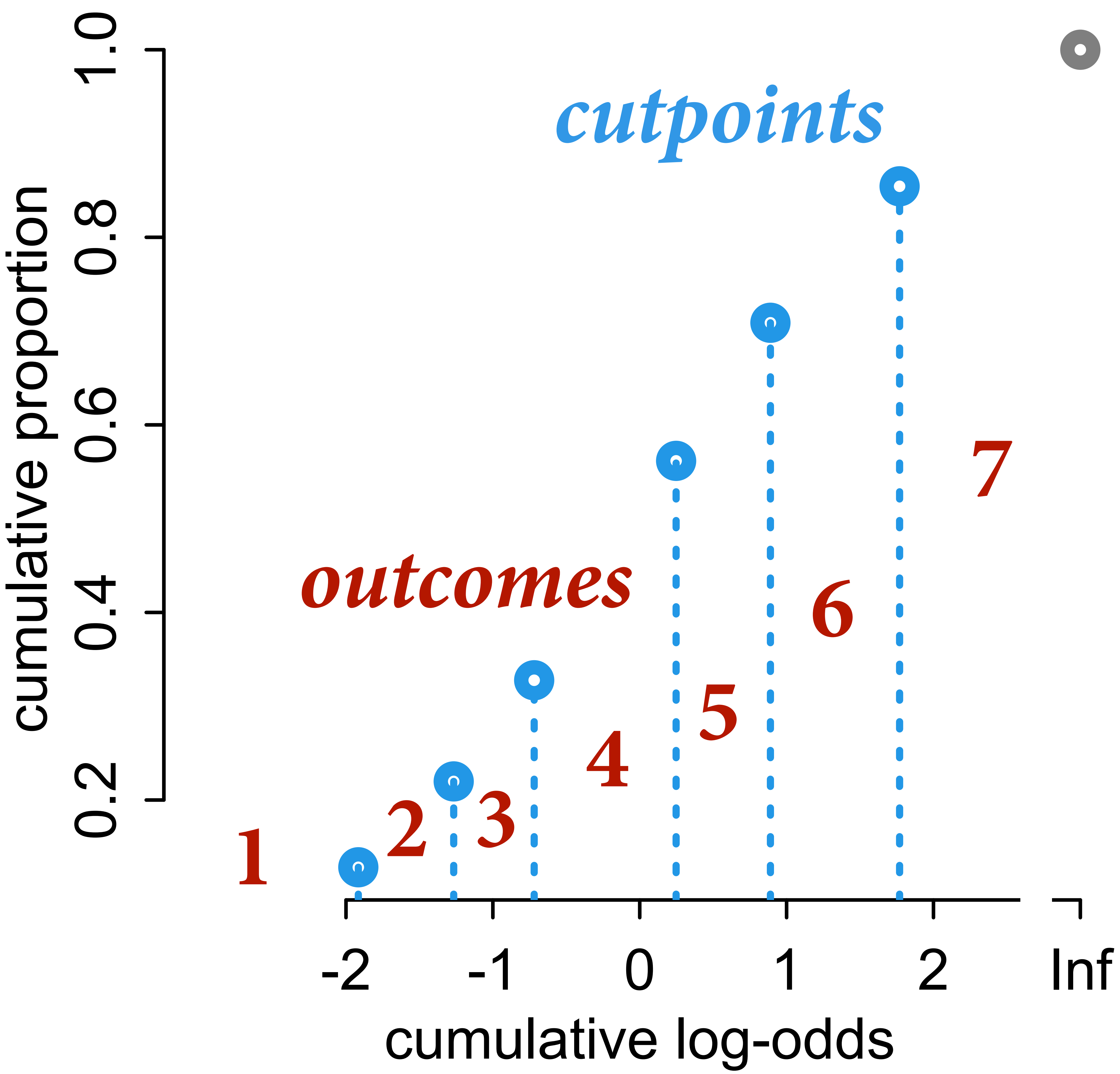

cumulative proportion
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
outcome
cumulative proportion
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
-2
-1
0
1
2
Inf
cumulative log-odds



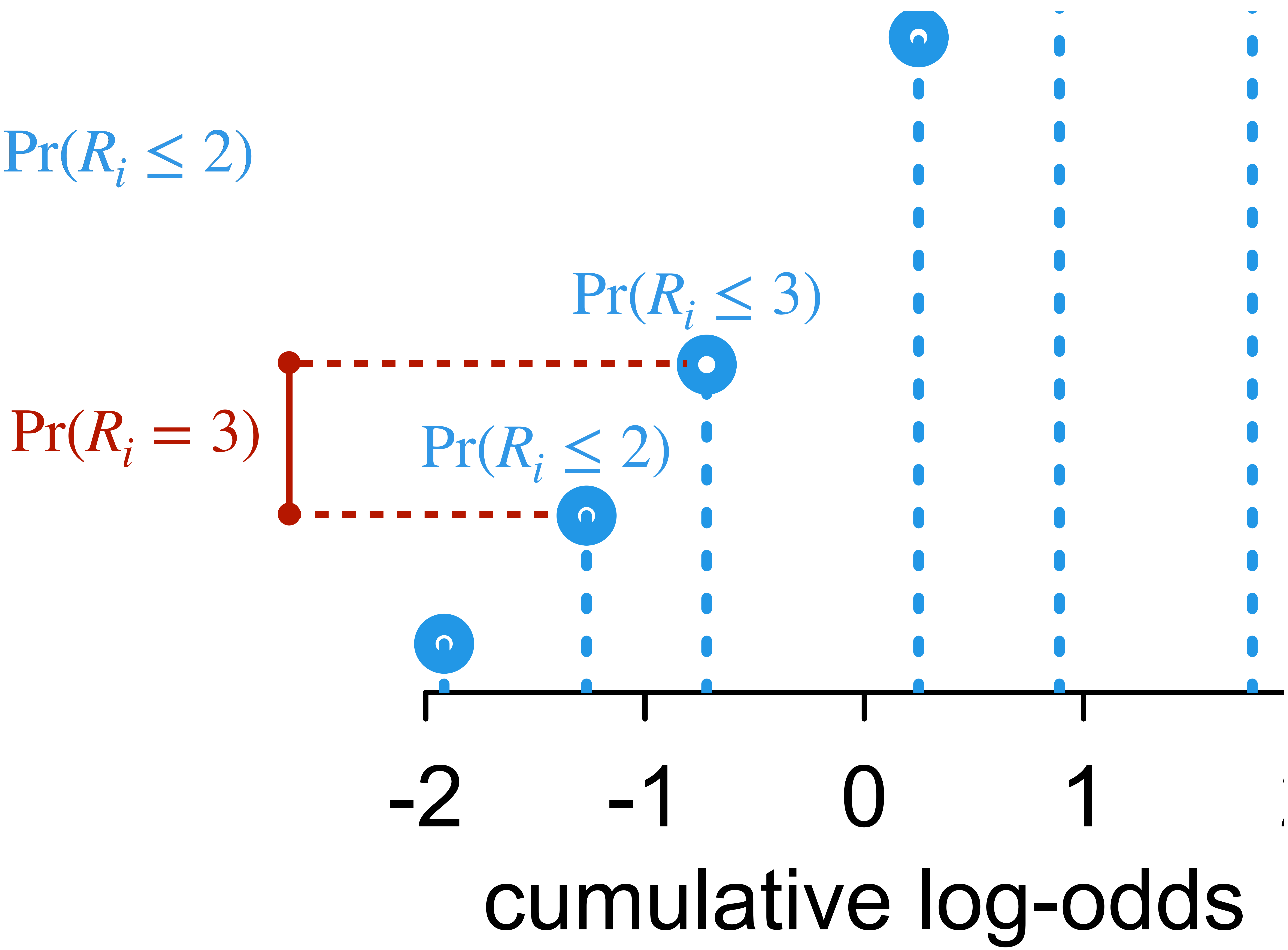
cumulative proportion
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
outcome
cutpoints
cumulative proportion
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Inf
cumulative log-odds


cumulative proportion
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
outcome
cumulative proportion
0.2
0.4
0.6
0.8
1.0
cutpoints
outcomes
1
2
3
4
5
6
7
-2
-1
0
1
2
Inf
cumulative log-odds

$$\Pr(R_i = k) = \Pr(R_i \leq k) - \Pr(R_i \leq k - 1)$$



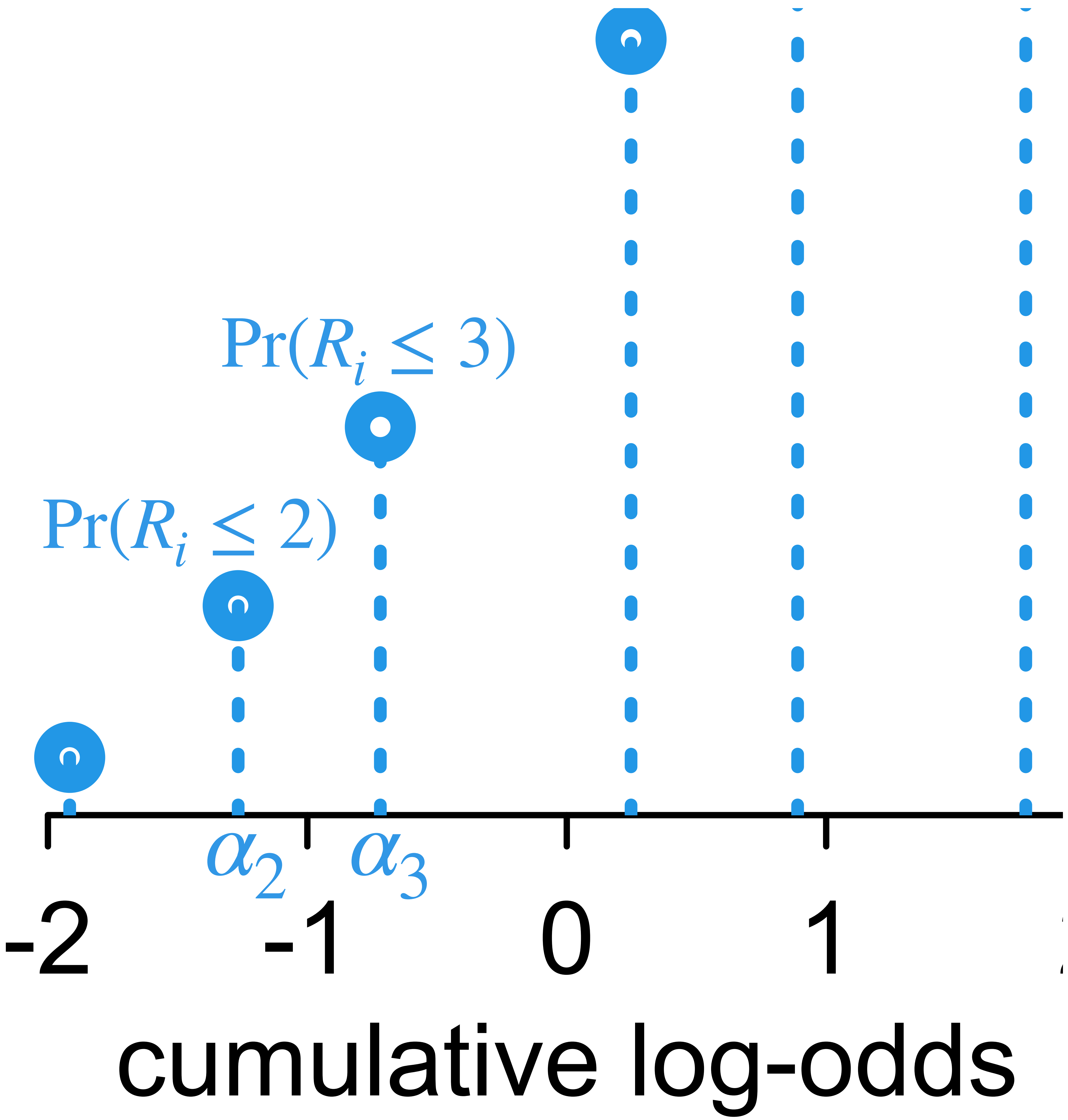
$$\Pr(R_i = 3) = \Pr(R_i \leq 3) - \Pr(R_i \leq 2)$$



$$\Pr(R_i = 3) = \Pr(R_i \leq 3) - \Pr(R_i \leq 2)$$

$$\underbrace{\log \frac{\Pr(R_i \leq k)}{1 - \Pr(R_i \leq k)}}_{\textit{cumulative log-odds}} = \alpha_k$$

cutpoint
(to estimate)



Where's the GLM?

So far just estimating the histogram

How to make it a function of variables?

(1) Stratify cutpoints

(2) Offset each cutpoint by value of linear model ϕ_i

Where's the GLM?

So far just estimating the histogram

$$\phi_i = \beta x_i$$

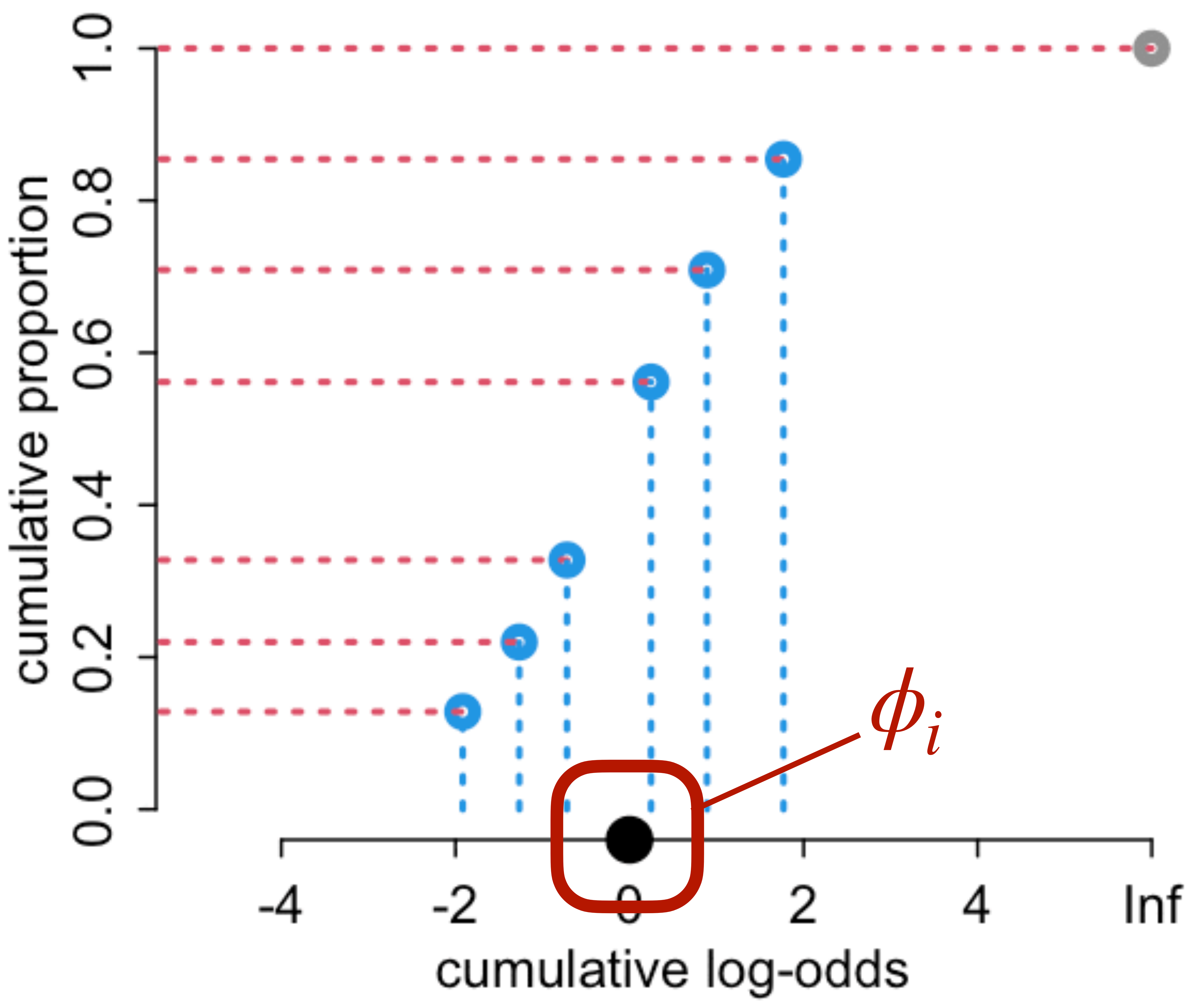
How to make it a function of variables?

$$\log \frac{\Pr(R_i \leq k)}{1 - \Pr(R_i \leq k)} = \alpha_k + \phi_i$$

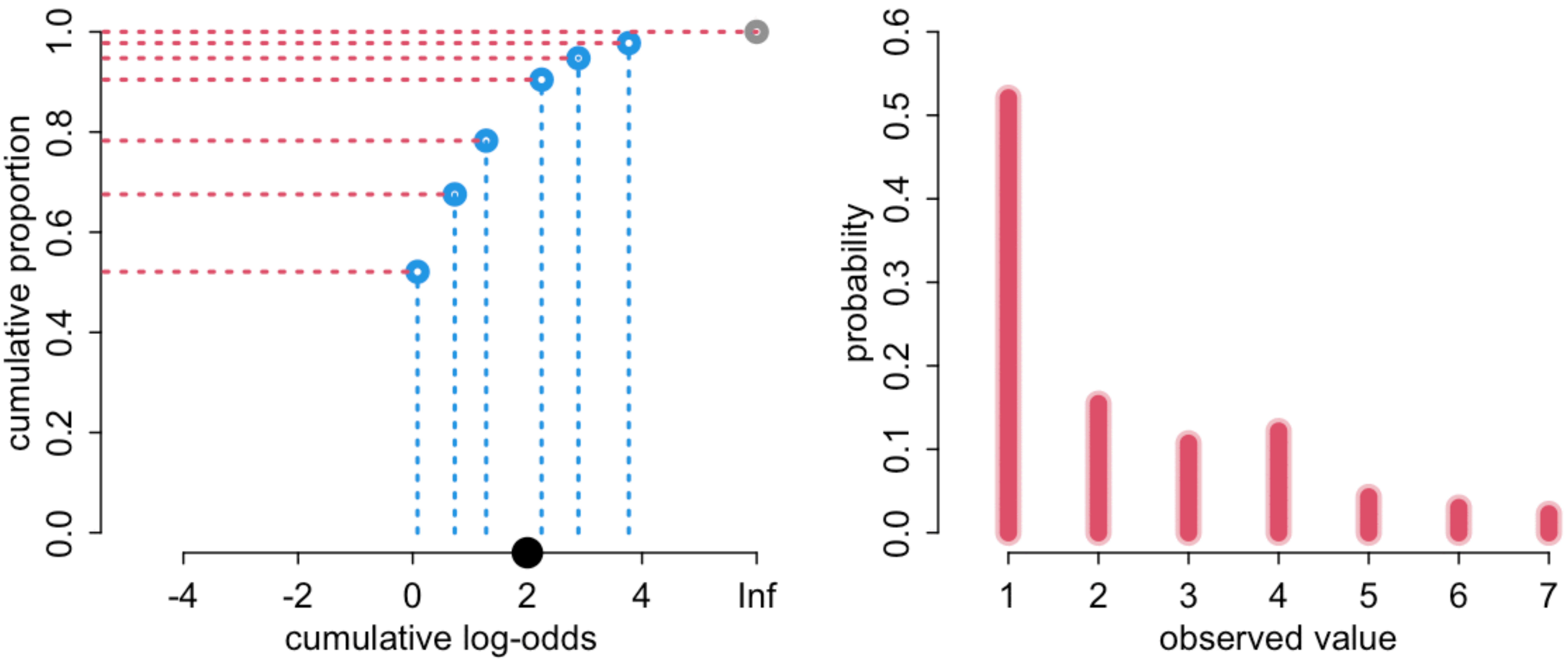
(1) Stratify cutpoints

(2) Offset each cutpoint by value of linear model ϕ_i

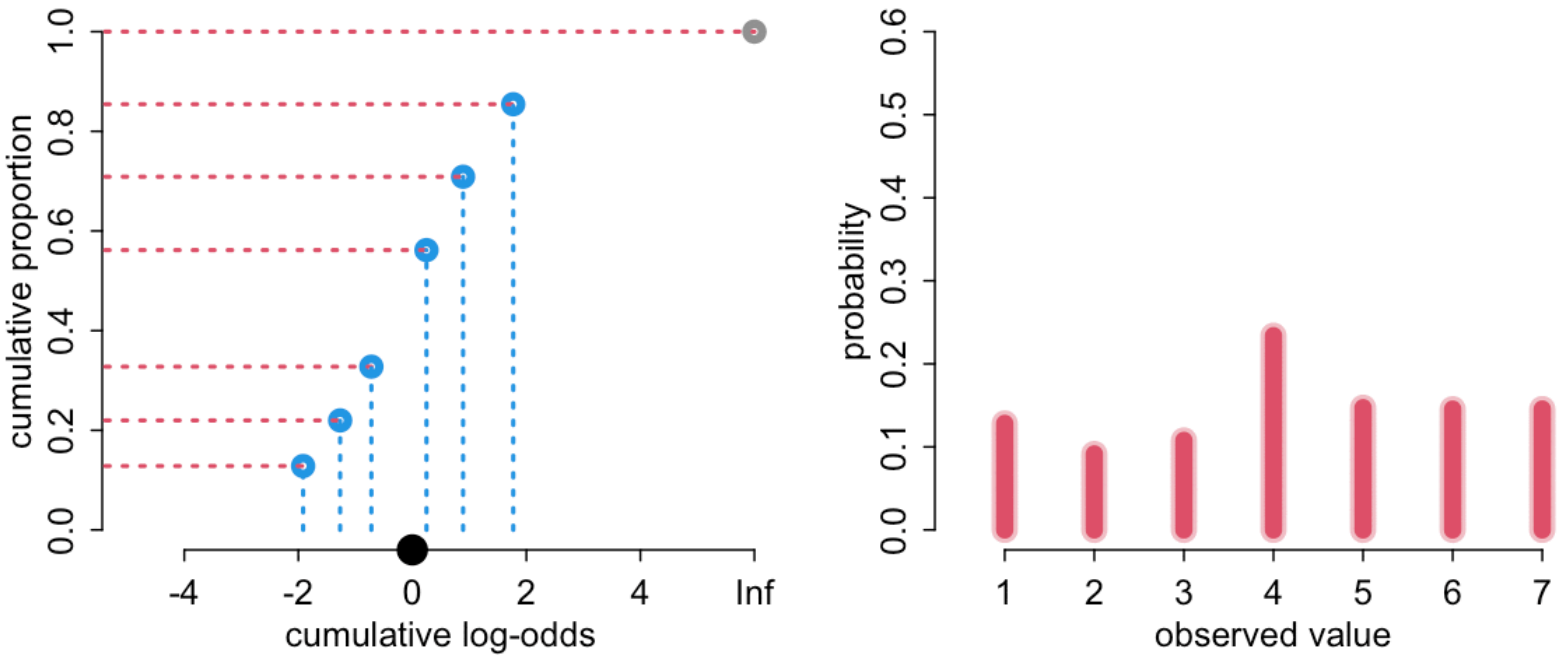
$$R_i \sim \text{OrderedLogit}(\phi_i, \alpha)$$



$$\log \frac{\Pr(R_i \leq k)}{1 - \Pr(R_i \leq k)} = \alpha_k + \phi_i$$

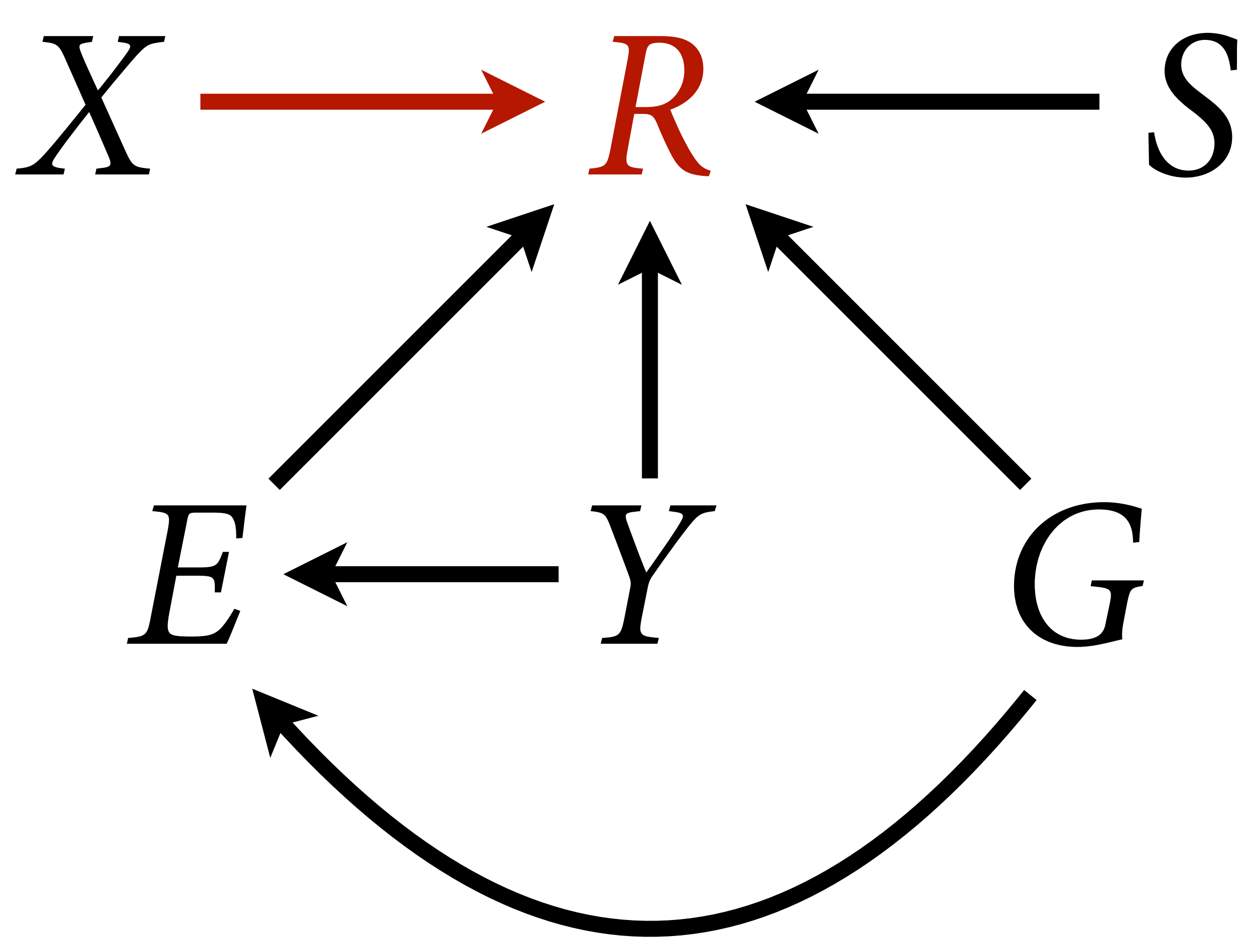


cumulative proportion
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Inf
cumulative log-odds
probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1
2
3
4
5
6
7
observed value



cumulative proportion
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Inf
cumulative log-odds
probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1
2
3
4
5
6
7
observed value

Start off easy:



$$R_i \sim \text{OrderedLogit}(\phi_i, \alpha)$$
$$\phi_i = \beta_A A_i + \beta_C C_i + \beta_I I_i$$
$$\beta_{\_} \sim \text{Normal}(0,0.5)$$
$$\alpha_j \sim \text{Normal}(0,1)$$

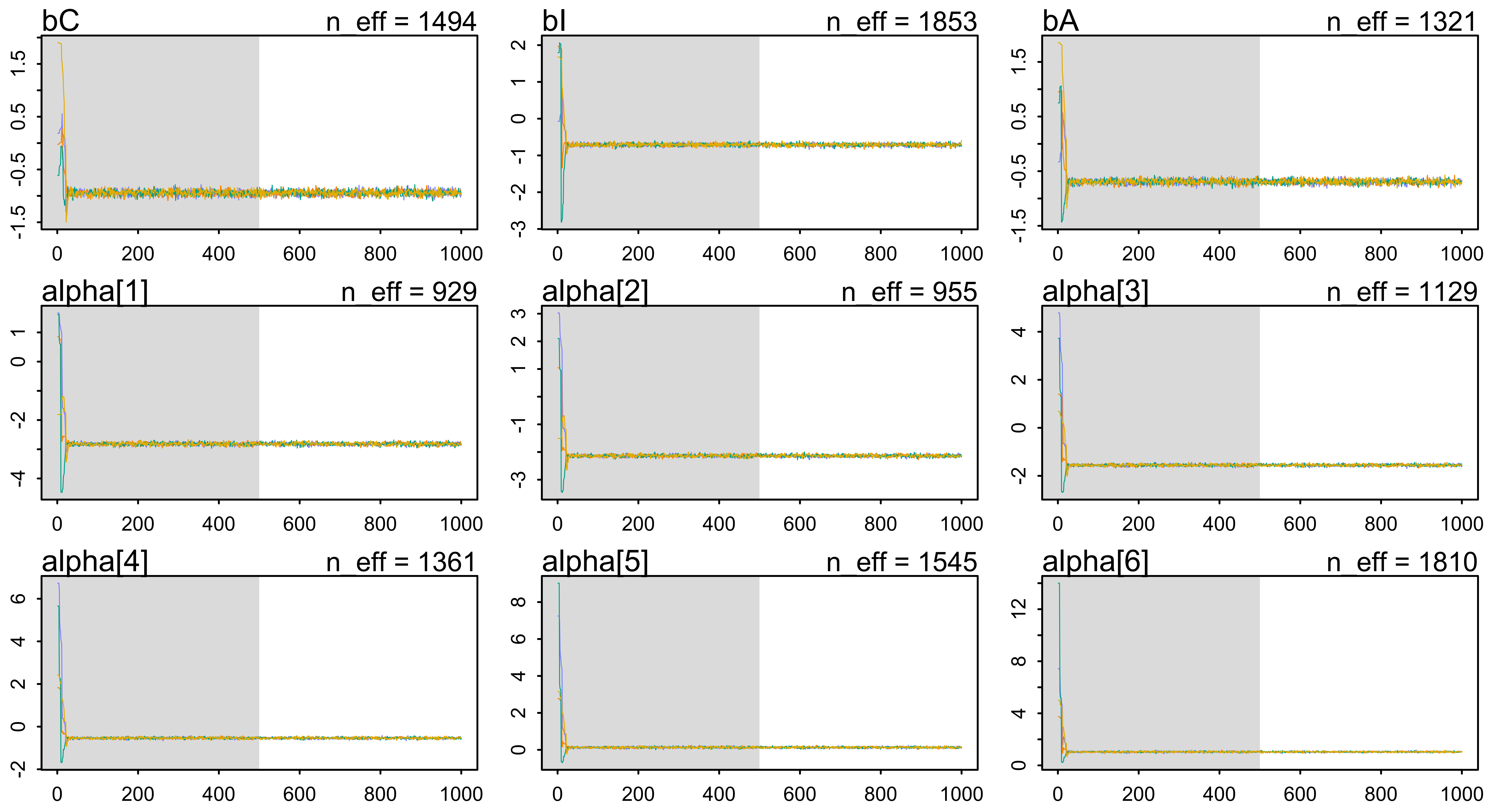
```r
data(Trolley)
d <- Trolley
dat <- list(
    R = d$response,
    A = d$action,
    I = d$intention,
    C = d$contact
)

mRX <- ulam(
    alist(

        R ~ dordlogit(phi,alpha),
        phi <- bA*A + bI*I + bC*C,
        c(bA,bI,bC) ~ normal(0,0.5),
        alpha ~ normal(0,1)

    ) , data=dat , chains=4 , cores=4 )
```

$$R_i \sim \text{OrderedLogit}(\phi_i, \alpha)$$

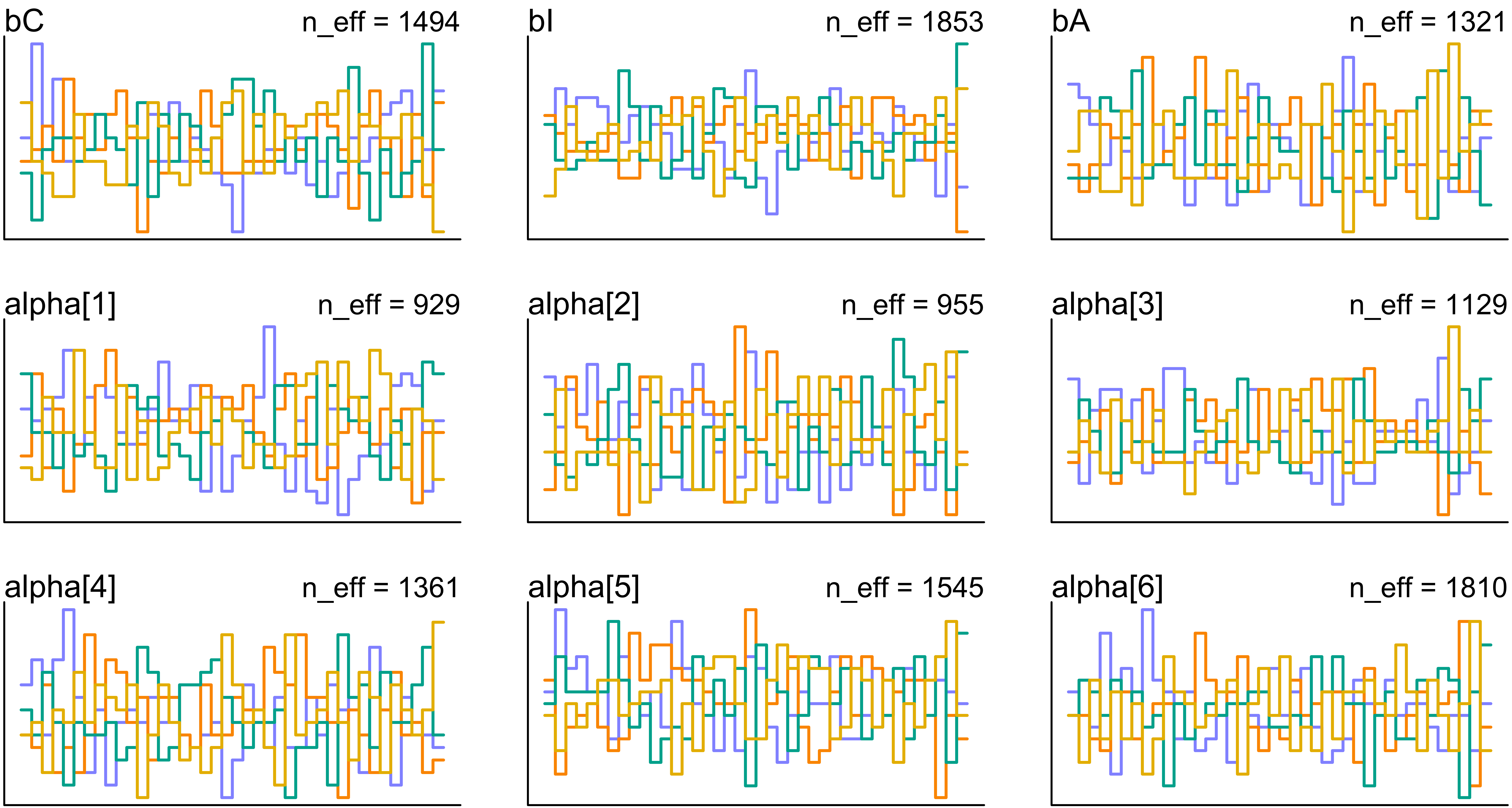
$$\phi_i = \beta_A A_i + \beta_C C_i + \beta_I I_i$$

$$\beta_{\_} \sim \text{Normal}(0{,}0.5)$$

$$\alpha_j \sim \text{Normal}(0{,}1)$$



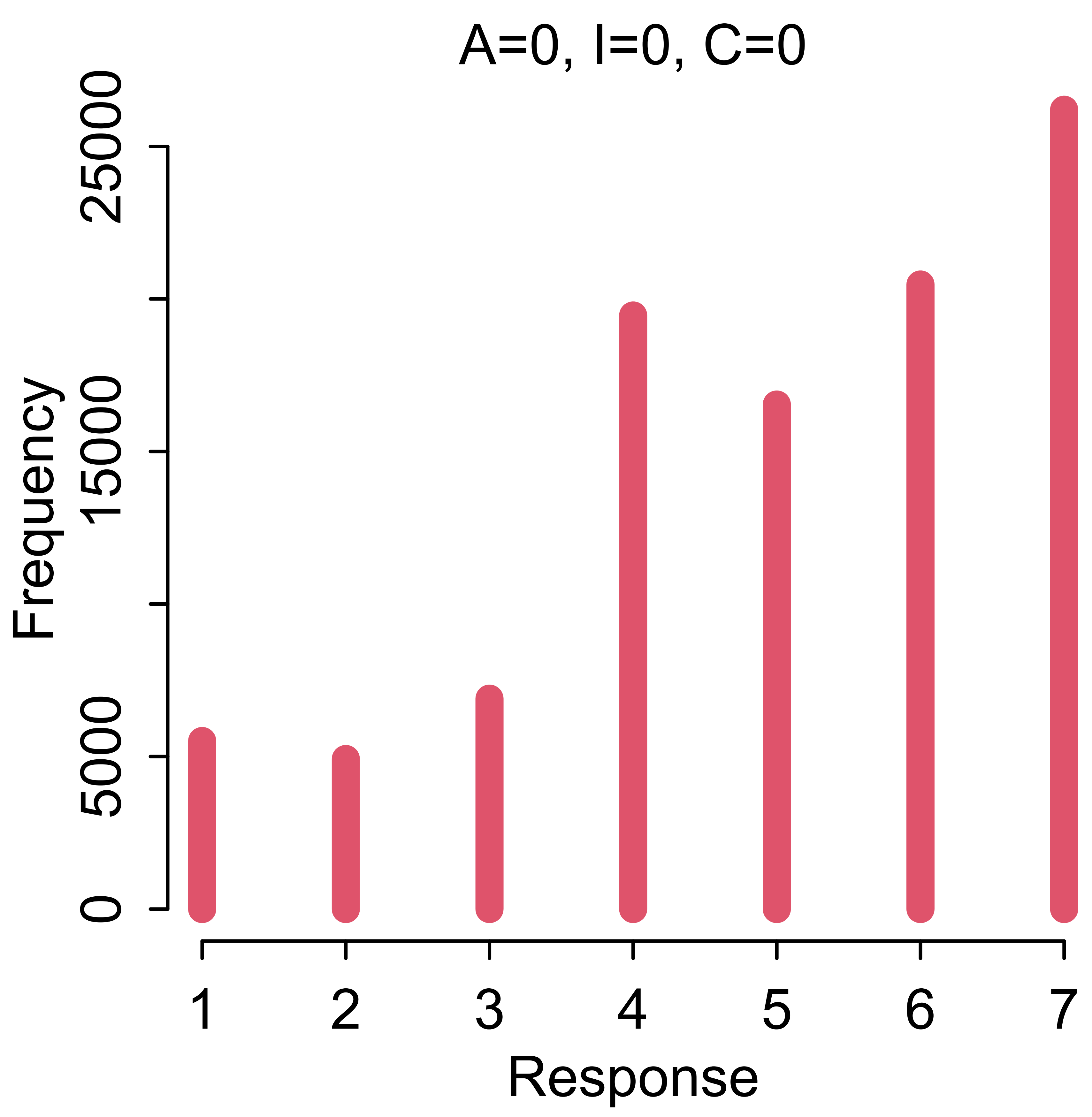
bC
n_eff = 1494
bI
n_eff = 1853
bA
n_eff = 1321
alpha[1]
n_eff = 929
alpha[2]
n_eff = 955
alpha[3]
n_eff = 1129
alpha[4]
n_eff = 1361
alpha[5]
n_eff = 1545
alpha[6]
n_eff = 1810

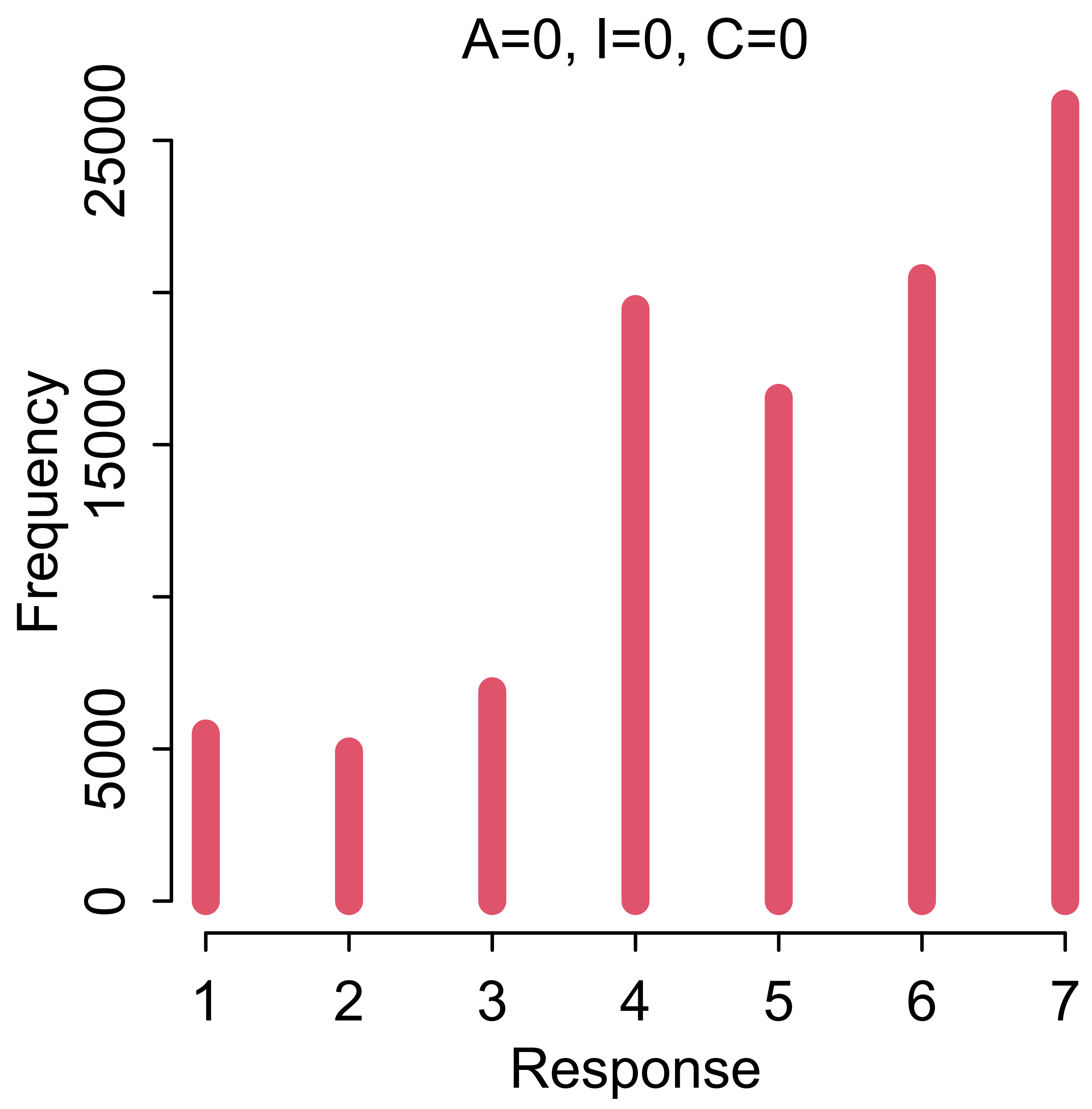

bC
n_eff = 1494
bI
n_eff = 1853
bA
n_eff = 1321
alpha[1]
n_eff = 929
alpha[2]
n_eff = 955
alpha[3]
n_eff = 1129
alpha[4]
n_eff = 1361
alpha[5]
n_eff = 1545
alpha[6]
n_eff = 1810

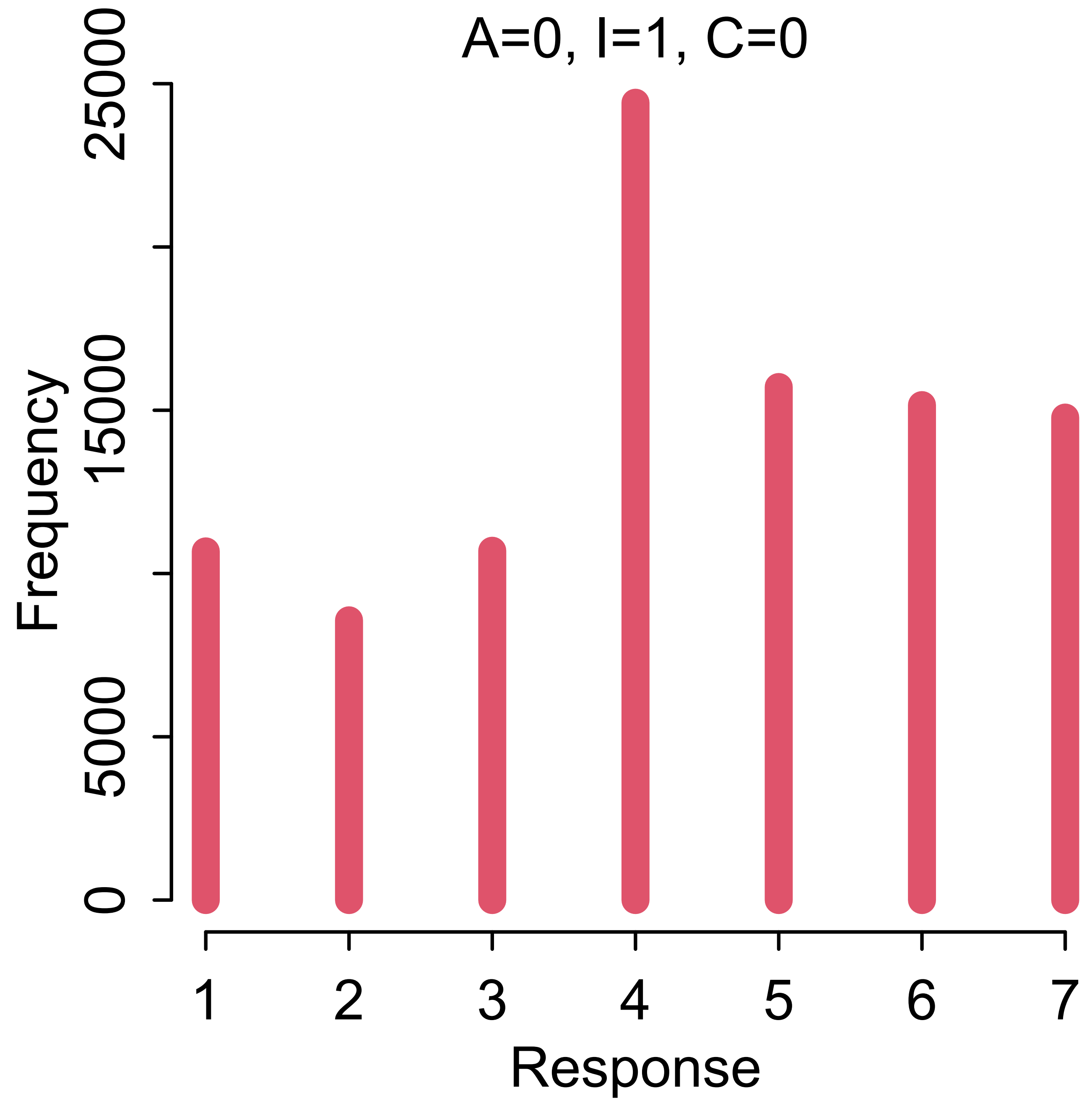
```r
data(Trolley)
d <- Trolley
dat <- list(
    R = d$response,
    A = d$action,
    I = d$intention,
    C = d$contact
)

mRX <- ulam(
    alist(

        R ~ dordlogit(phi,alpha),
        phi <- bA*A + bI*I + bC*C,
        c(bA,bI,bC) ~ normal(0,0.5),
        alpha ~ normal(0,1)

    ) , data=dat , chains=4 , cores=4 )
```

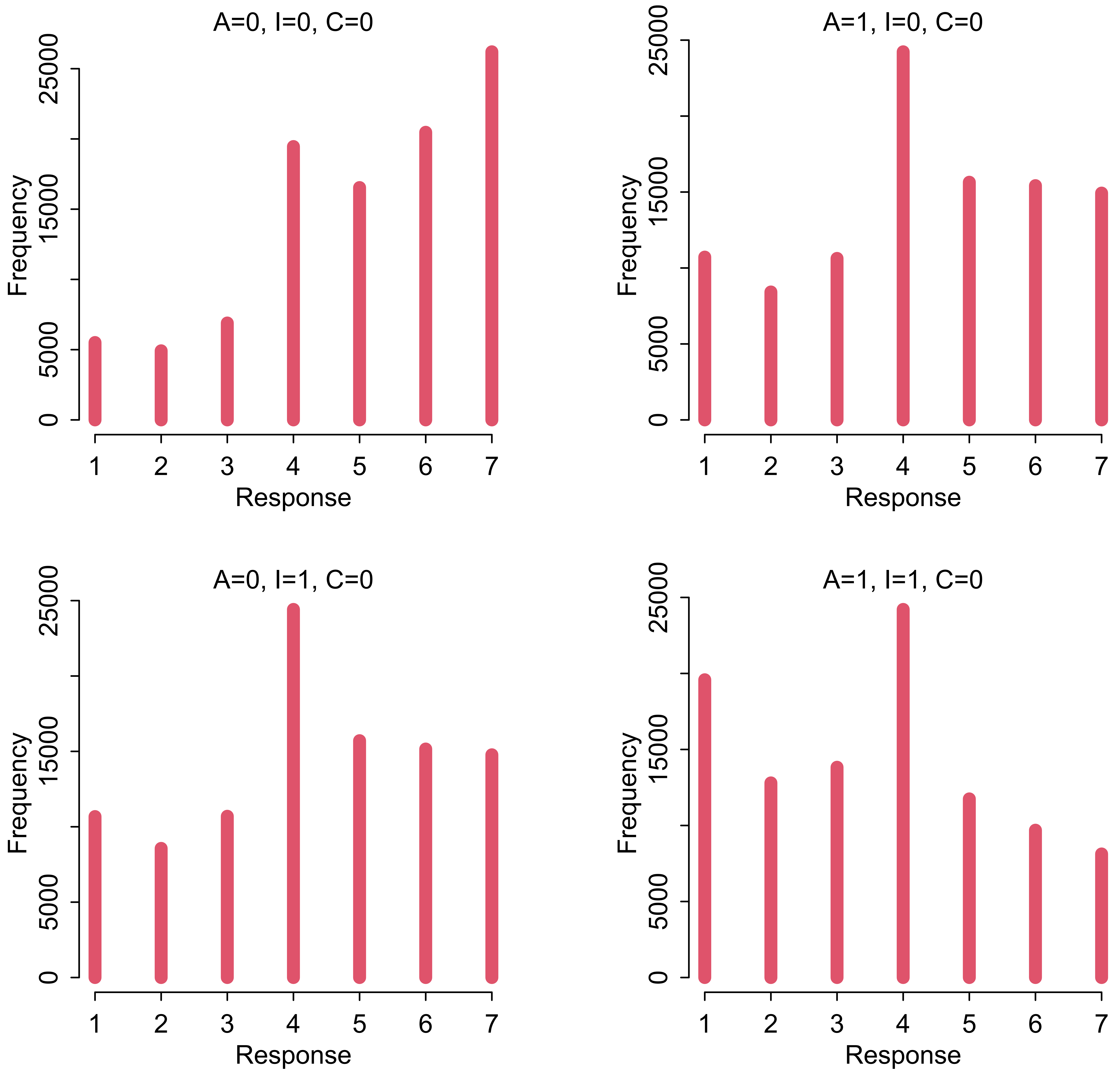
```
> precis(mRX,2)
          mean   sd  5.5% 94.5% n_eff Rhat4
bC       -0.94 0.05 -1.02 -0.87  1494     1
bI       -0.71 0.04 -0.77 -0.65  1853     1
bA       -0.69 0.04 -0.76 -0.63  1321     1
alpha[1] -2.82 0.05 -2.89 -2.74   929     1
alpha[2] -2.14 0.04 -2.20 -2.07   955     1
alpha[3] -1.56 0.04 -1.62 -1.49  1129     1
alpha[4] -0.54 0.04 -0.59 -0.48  1361     1
alpha[5]  0.13 0.04  0.07  0.19  1545     1
alpha[6]  1.04 0.04  0.97  1.10  1810     1
```

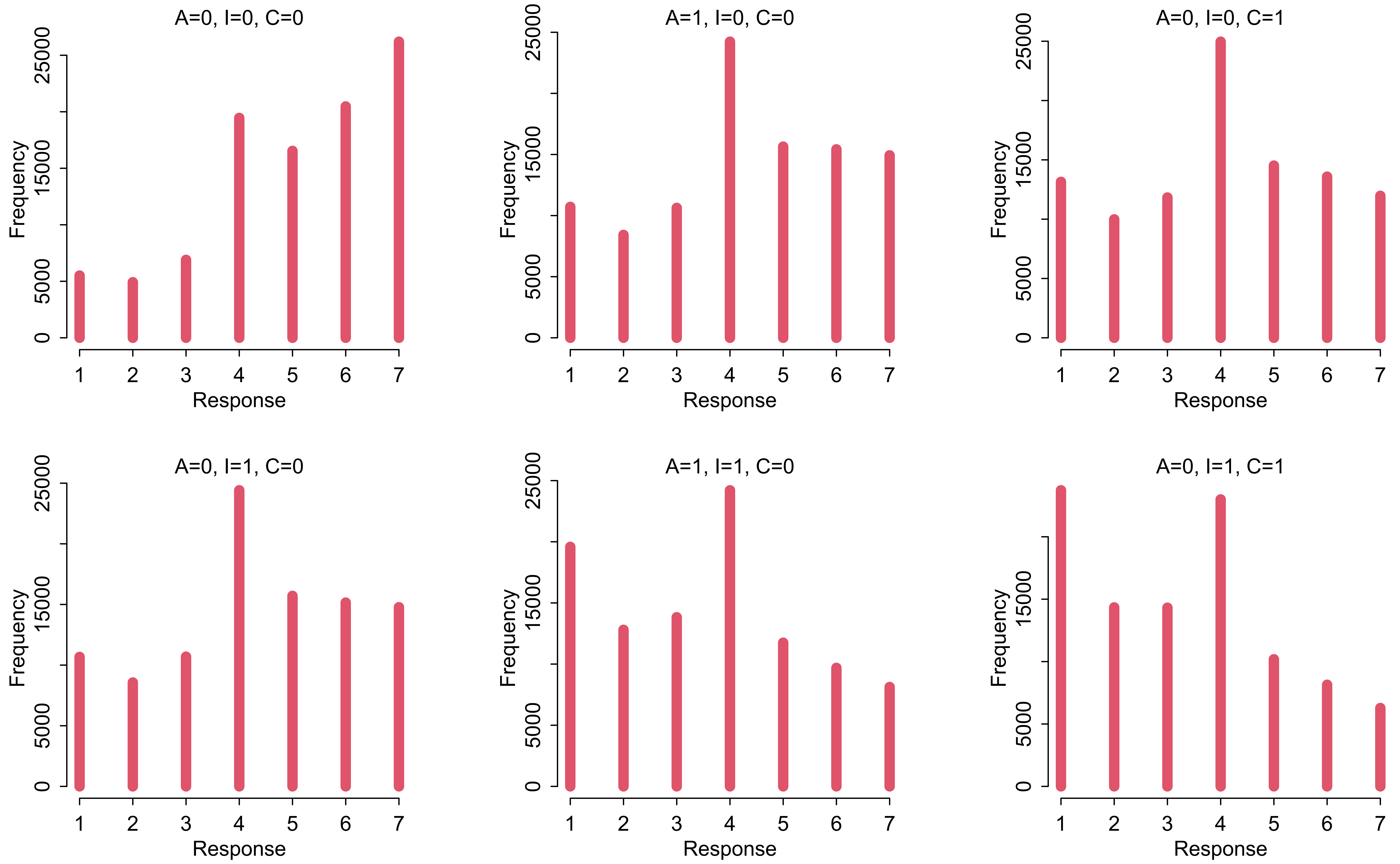
```r
# plot predictive distributions for each treatment

vals <- c(0,0,0)
Rsim <- mcreplicate( 100 ,
sim(mRX,data=list(A=vals[1],I=vals[2],C=vals[3])) ,
mc.cores=6 )

simplehist(as.vector(Rsim),lwd=8,col=2,xlab="Response")
mtext(concat("A=",vals[1],", I=",vals[2],",
C=",vals[3]))
```



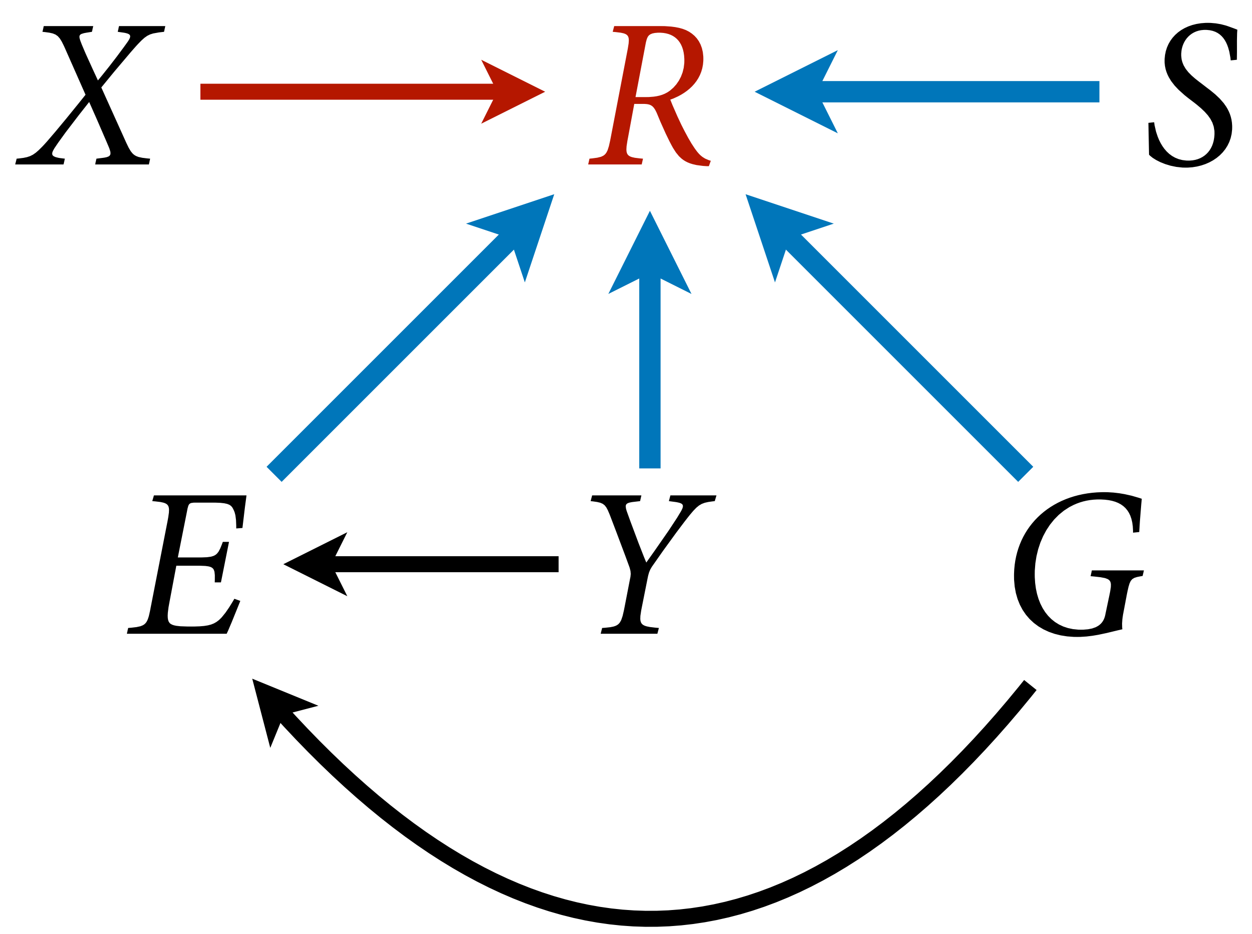

A=0, I=0, C=0
Frequency
0
5000
15000
25000
1
2
3
4
5
6
7
Response


A=0, I=1, C=0
Frequency
0
5000
15000
25000
1
2
3
4
5
6
7
Response


A=0, I=0, C=0
A=1, I=0, C=0
A=0, I=1, C=0
A=1, I=1, C=0
Frequency
Response
0
5000
15000
25000
1
2
3
4
5
6
7


A=0, I=0, C=0
A=1, I=0, C=0
A=0, I=0, C=1
A=0, I=1, C=0
A=1, I=1, C=0
A=0, I=1, C=1
Frequency
Response
0
5000
15000
25000
1 2 3 4 5 6 7

What about the competing causes?



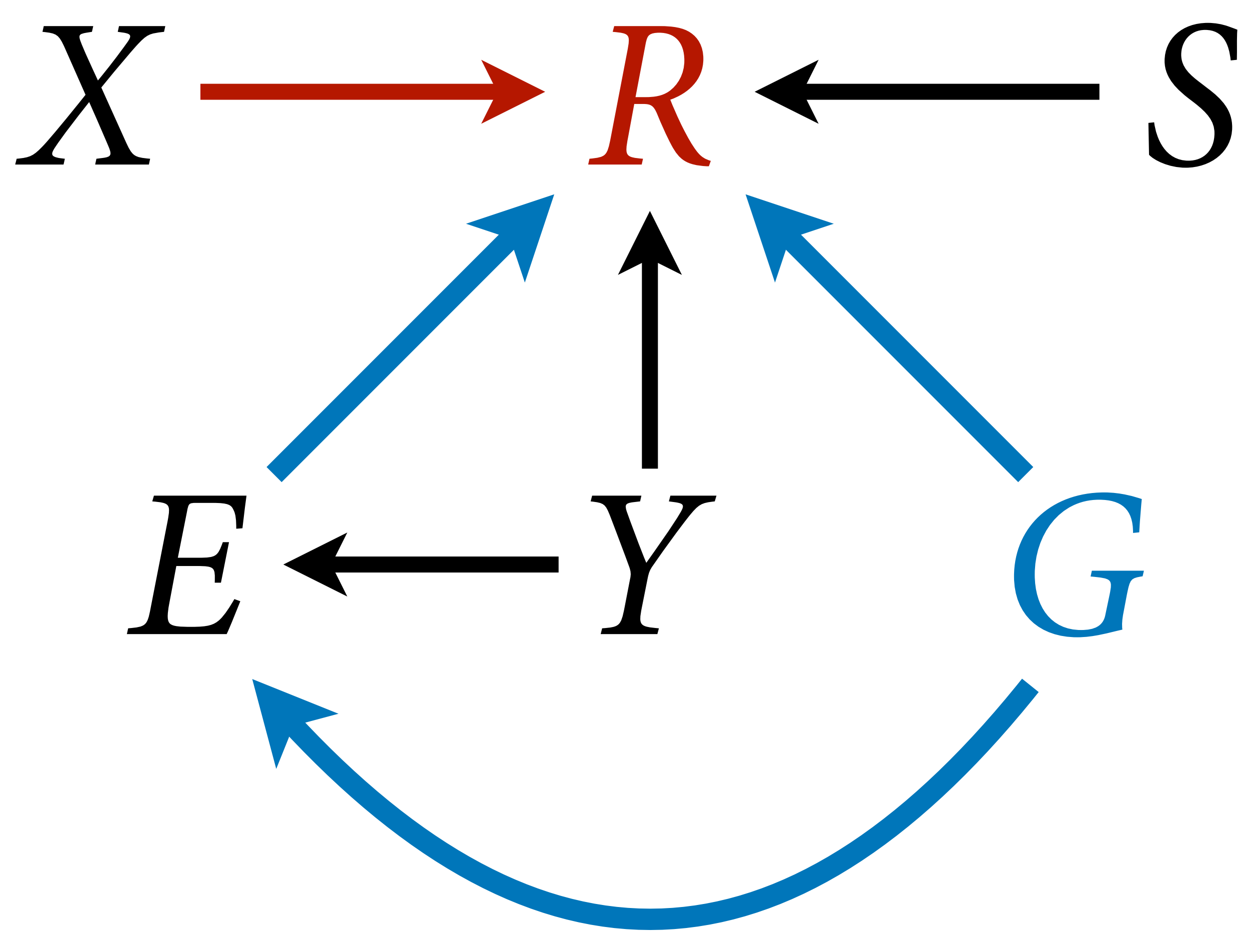
$$R_i \sim \text{OrderedLogit}(\phi_i, \alpha)$$

$$\phi_i = \beta_A A_i + \beta_C C_i + \beta_I I_i$$

$$\beta \sim \text{Normal}(0,0.5)$$

$$\alpha_j \sim \text{Normal}(0,1)$$

Total effect of gender:



$$R_i \sim \text{OrderedLogit}(\phi_i, \alpha)$$

$$\phi_i = \beta_{A,G[i]}A_i + \beta_{C,G[i]}C_i + \beta_{I,G[i]}I_i$$

$$\beta\_ \sim \text{Normal}(0,0.5)$$

$$\alpha_j \sim \text{Normal}(0,1)$$

```
# total effect of gender
dat$G <- ifelse(d$male==1,2,1)
mRXG <- ulam(
    alist(

        R ~ dordlogit(phi,alpha),
        phi <- bA[G]*A + bI[G]*I + bC[G]*C,
        bA[G] ~ normal(0,0.5),
        bI[G] ~ normal(0,0.5),
        bC[G] ~ normal(0,0.5),
        alpha ~ normal(0,1)

    ) , data=dat , chains=4 , cores=4 )
```

$$R_i \sim \text{OrderedLogit}(\phi_i, \alpha)$$

$$\phi_i = \beta_{A,G[i]} A_i + \beta_{C,G[i]} C_i + \beta_{I,G[i]} I_i$$

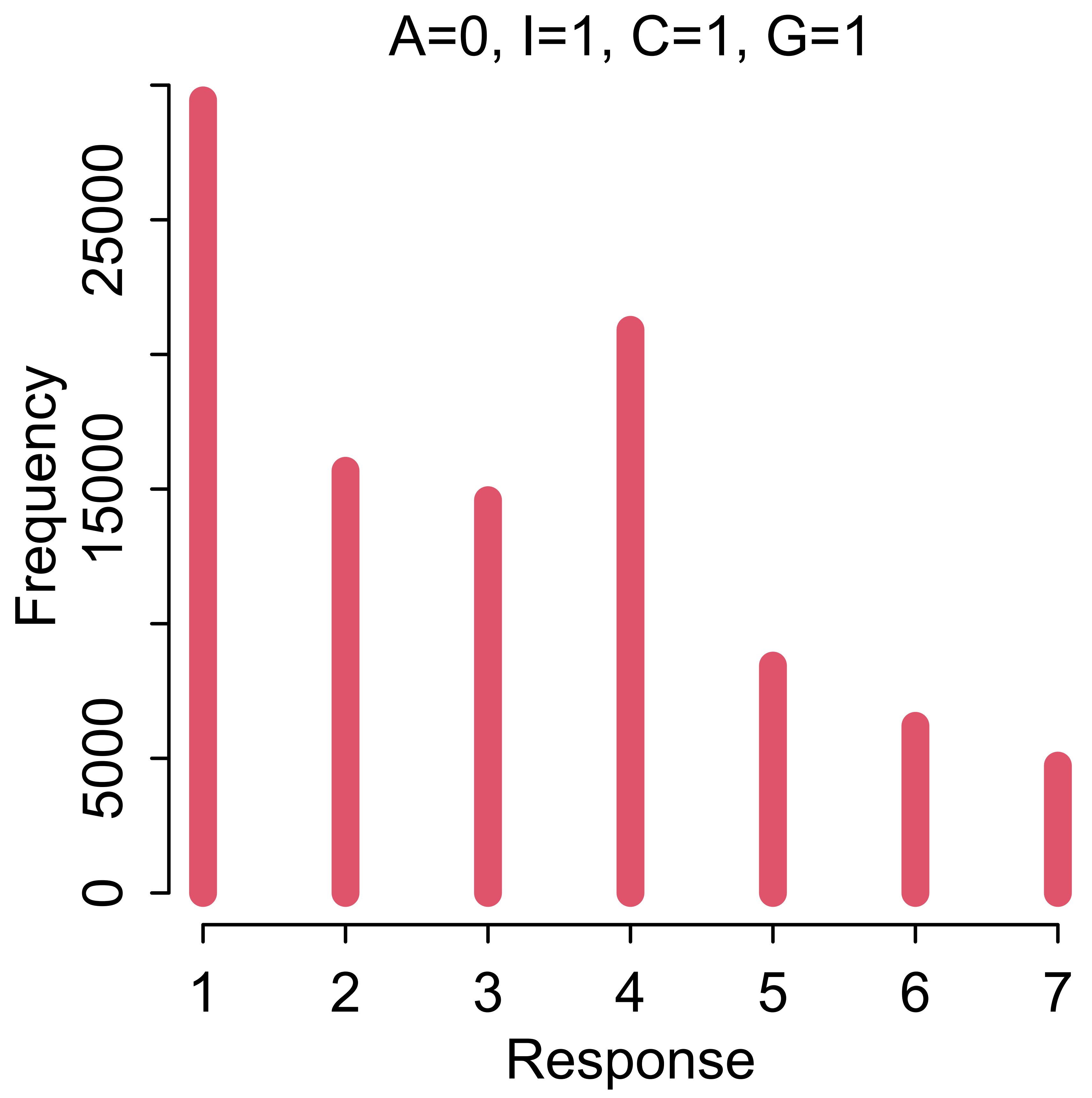
$$\beta_{\_} \sim \text{Normal}(0,0.5)$$

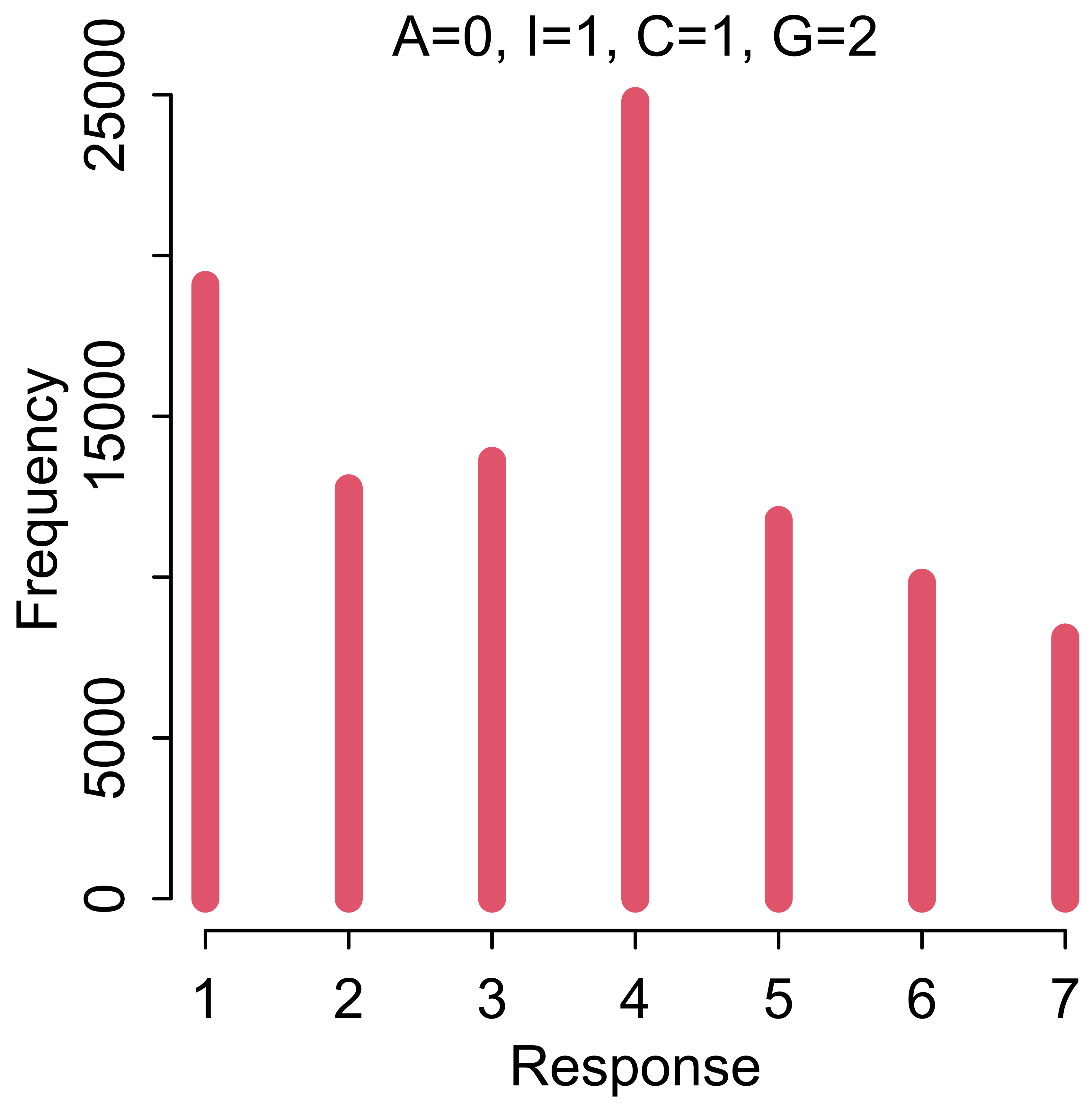
$$\alpha_j \sim \text{Normal}(0,1)$$

```
# total effect of gender
dat$G <- ifelse(d$male==1,2,1)
mRXG <- ulam(
    alist(

        R ~ dordlogit(phi,alpha),
        phi <- bA[G]*A + bI[G]*I + bC[G]*C,
        bA[G] ~ normal(0,0.5),
        bI[G] ~ normal(0,0.5),
        bC[G] ~ normal(0,0.5),
        alpha ~ normal(0,1)

    ) , data=dat , chains=4 , cores=4 )
```

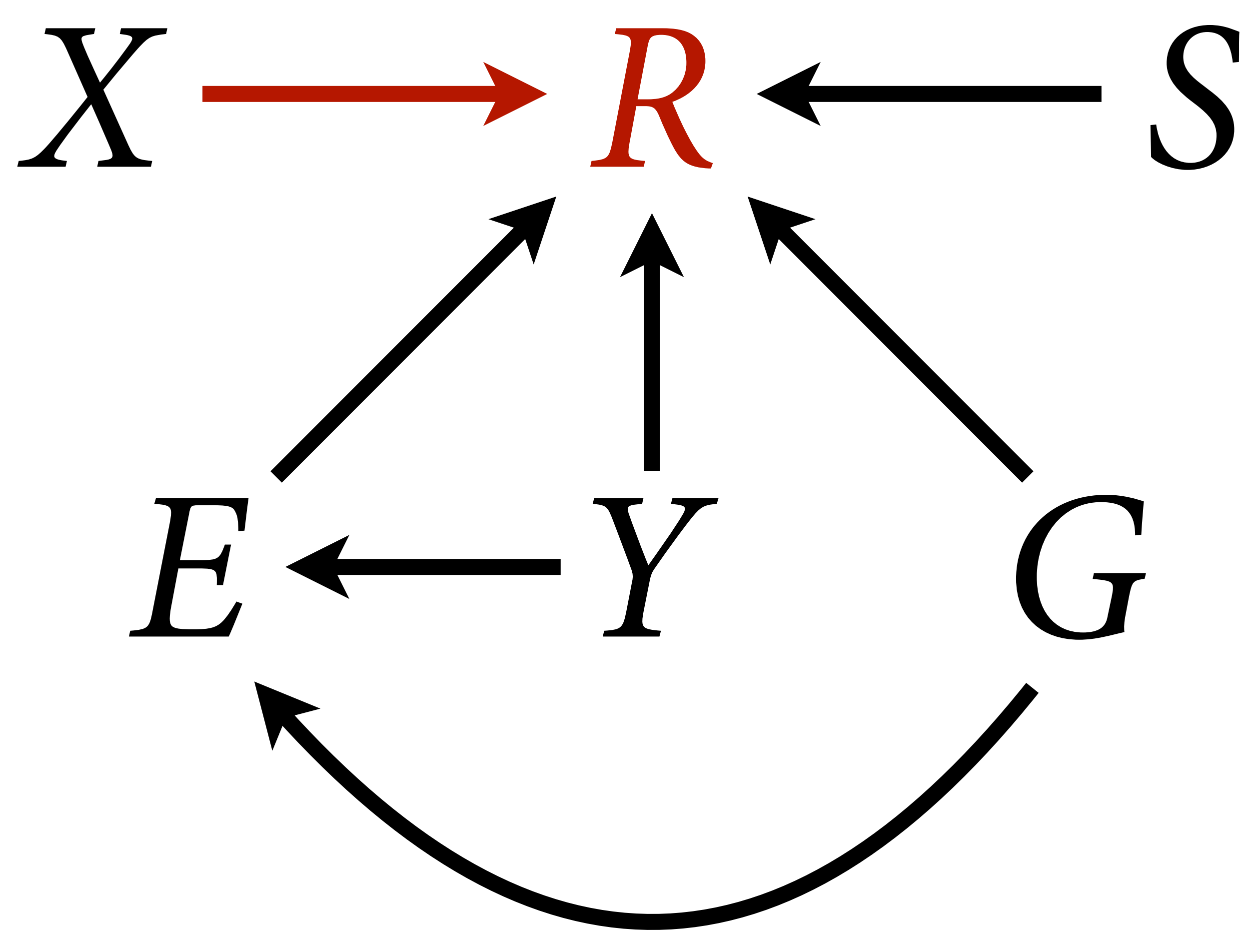
```
> precis(mRXG,2)
          mean   sd  5.5% 94.5% n_eff Rhat4
bA[1]    -0.88 0.05 -0.96 -0.80  1858  1.00
bA[2]    -0.53 0.05 -0.61 -0.45  1724  1.00
bI[1]    -0.90 0.05 -0.97 -0.82  2189  1.00
bI[2]    -0.55 0.05 -0.63 -0.48  2382  1.00
bC[1]    -1.06 0.07 -1.17 -0.95  2298  1.00
bC[2]    -0.84 0.06 -0.94 -0.74  2000  1.00
alpha[1] -2.83 0.05 -2.90 -2.75  1054  1.01
alpha[2] -2.15 0.04 -2.21 -2.08  1104  1.00
alpha[3] -1.56 0.04 -1.62 -1.50  1076  1.00
alpha[4] -0.53 0.04 -0.59 -0.47  1080  1.00
alpha[5]  0.14 0.04  0.09  0.20  1216  1.00
alpha[6]  1.06 0.04  1.00  1.12  1532  1.00
```

$G[i],i$

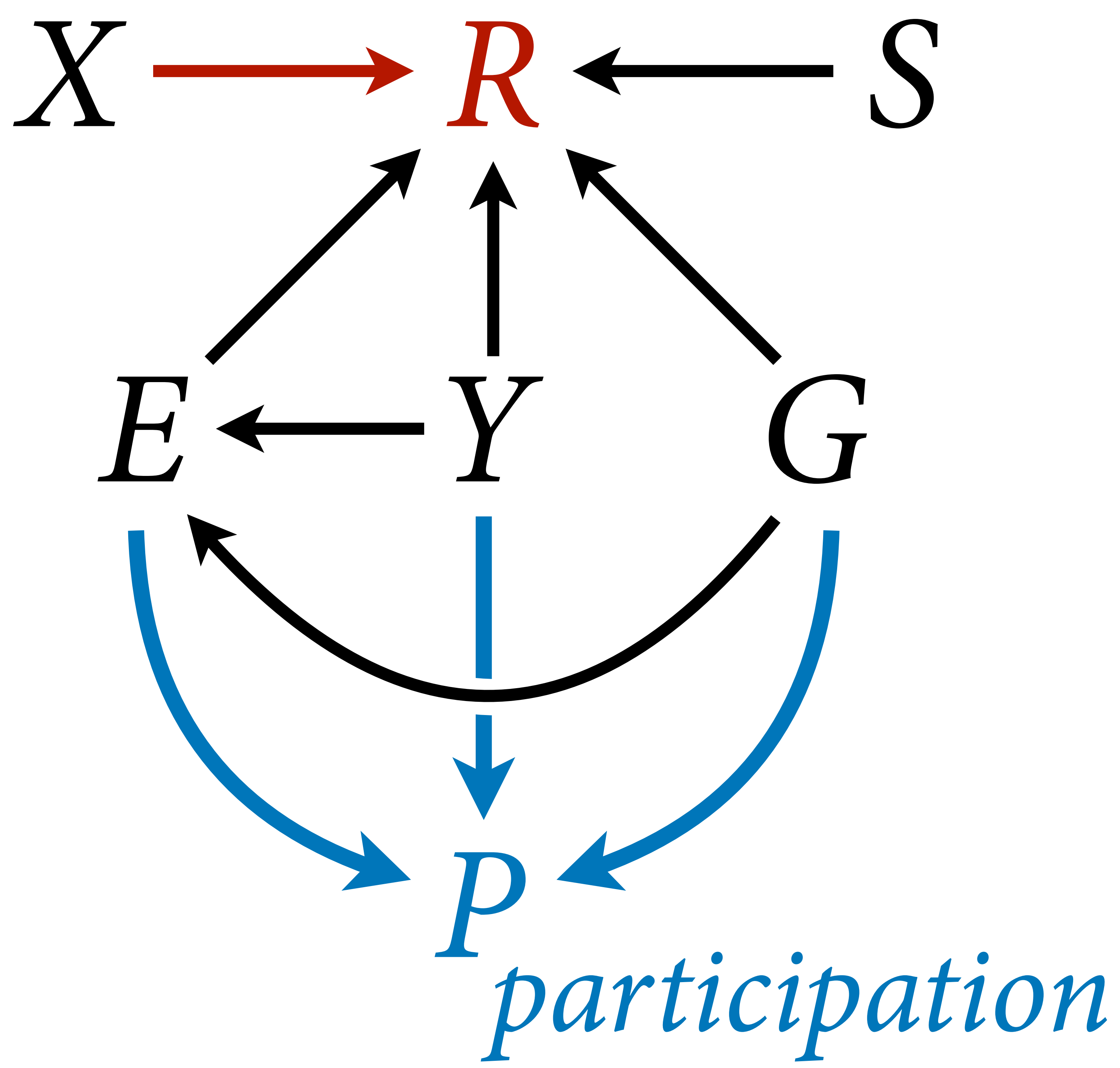

A=0, I=1, C=1, G=1
Frequency
0
5000
15000
25000
1
2
3
4
5
6
7
Response


A=0, I=1, C=1, G=2
Frequency
0
5000
15000
25000
1
2
3
4
5
6
7
Response

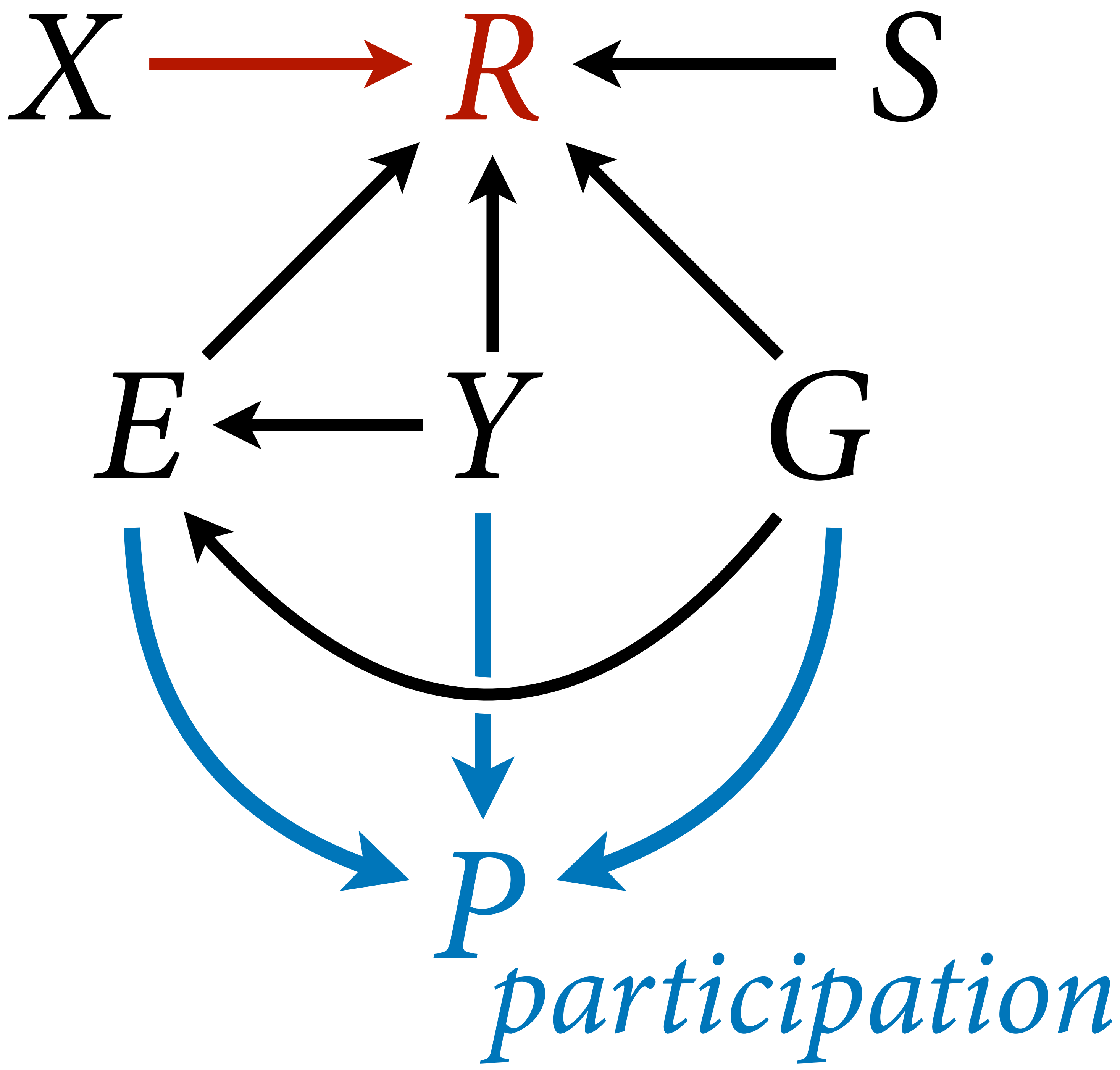
Hang on! This is a voluntary sample



Hang on! This is a voluntary sample



Hang on! This is a voluntary sample



Conditioning on P makes E,Y,G covary in sample

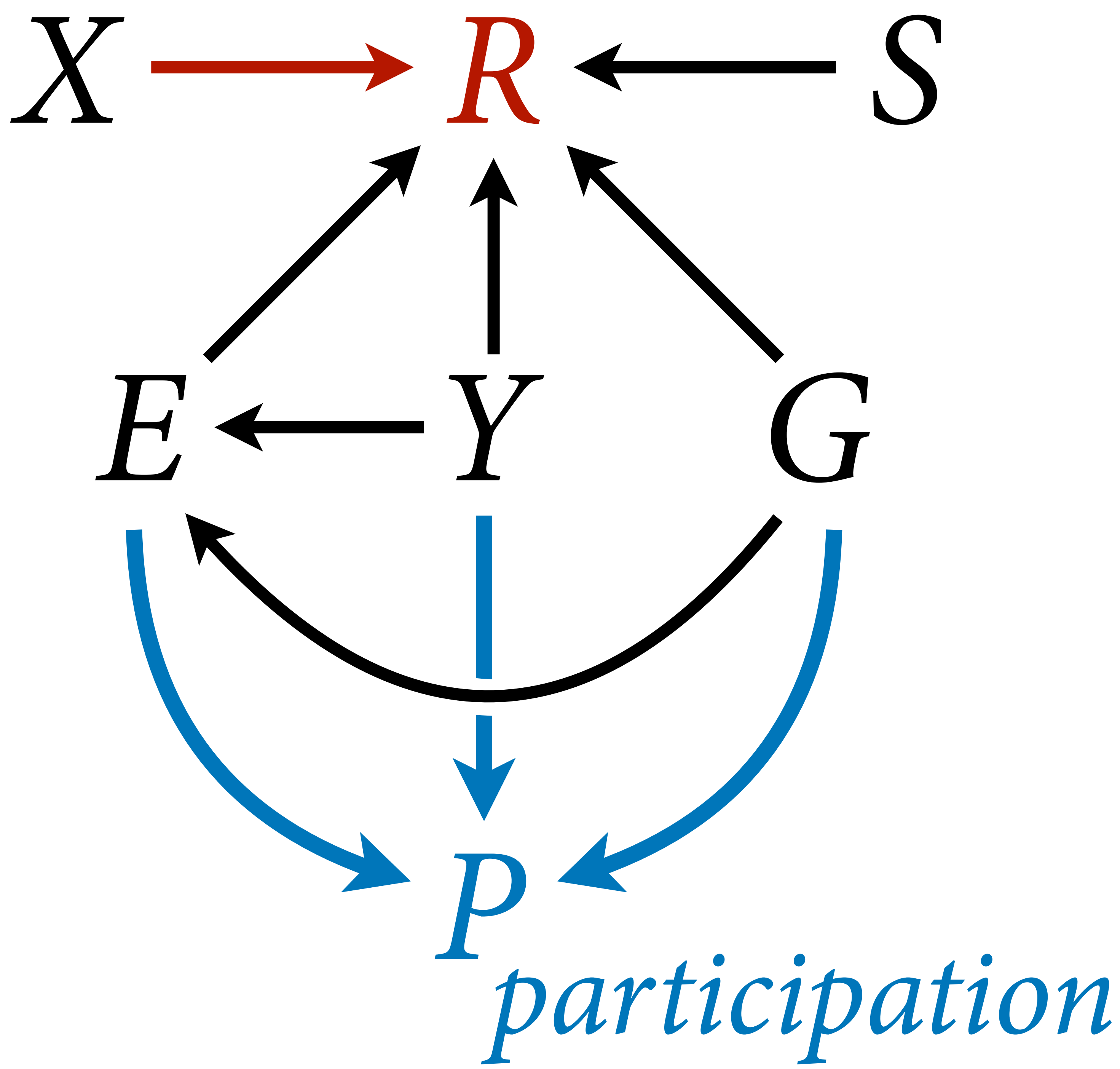
Endogenous selection

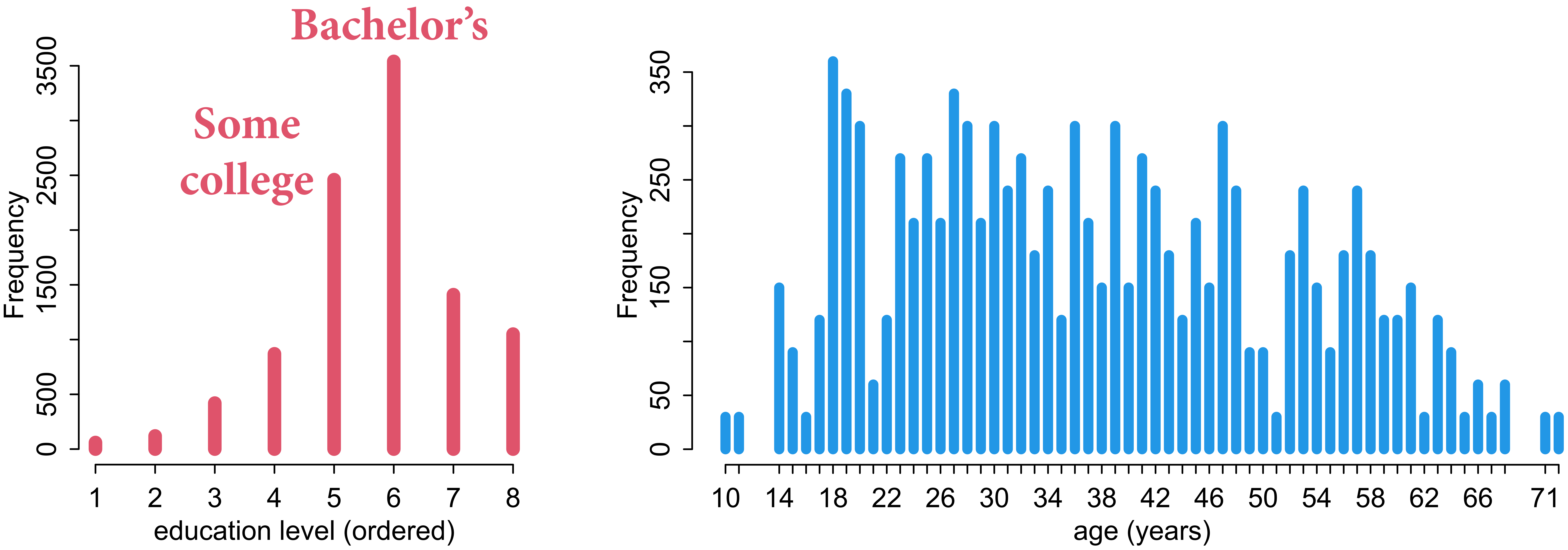
Sample is selected on a collider

Induces misleading associations among variables

Not possible here to estimate total effect of G, BUT can get direct effect

Need to stratify by E and Y and G





Bachelor's
Some college
Frequency
0
500
1500
2500
3500
1 2 3 4 5 6 7 8
education level (ordered)
Frequency
0
50
150
250
350
10 14 18 22 26 30 34 38 42 46 50 54 58 62 66 71
age (years)

PAUSE

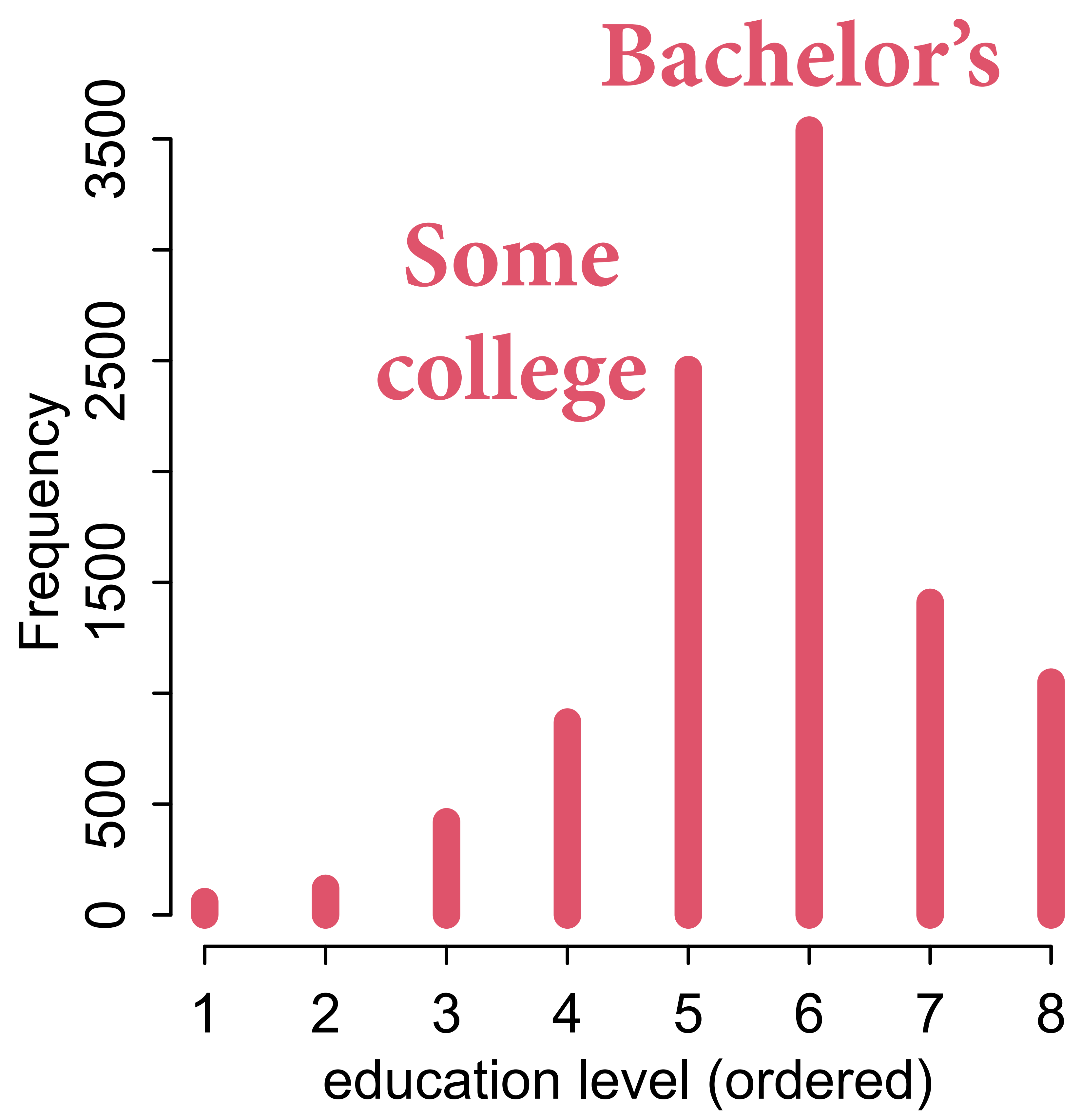
Ordered monotonic predictors

Education is an ordered category

Unlikely that each level has same effect

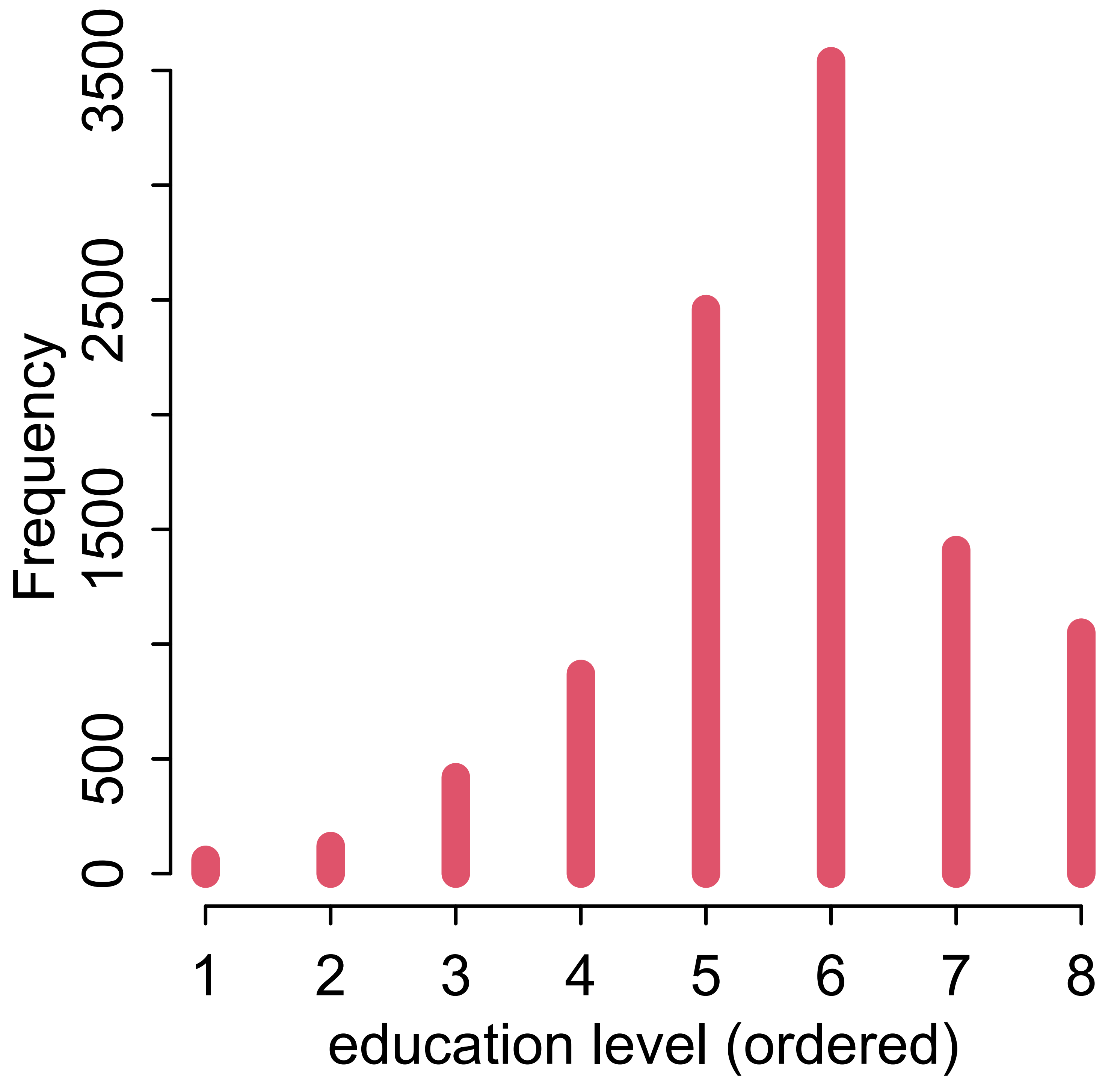
Want a parameter for each level

But how to enforce ordering, so that each level has larger (or smaller) effect than previous?



Ordered monotonic predictors

1 (elementary)	$\phi_i = 0$
2 (middle school)	$\phi_i = \delta_1$
3 (some high school)	$\phi_i = \delta_1 + \delta_2$
4 (high school)	$\phi_i = \delta_1 + \delta_2 + \delta_3$
5 (some college)	$\phi_i = \delta_1 + \delta_2 + \delta_3 + \delta_4$
6 (college)	$\phi_i = \delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5$
7 (master's)	$\phi_i = \delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6$
8 (doctorate)	$\phi_i = \delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6 + \delta_7$



Ordered monotonic predictors

1 (elementary) $\phi_i = 0$

2 (middle school) $\phi_i = \delta_1$

3 (some high school) $\phi_i = \delta_1 + \delta_2$

4 (high school) $\phi_i = \delta_1 + \delta_2 + \delta_3$

5 (some college) $\phi_i = \delta_1 + \delta_2 + \delta_3 + \delta_4$

6 (college) $\phi_i = \delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5$

7 (master's) $\phi_i = \delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6$

8 (doctorate) $\phi_i = \delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6 + \delta_7 = \boxed{\beta_E}$

maximum effect of education

Ordered monotonic predictors

1 (elementary)

2 (middle school)

3 (some high school)

4 (high school)

5 (some college)

6 (college)

7 (master's)

8 (doctorate)

$$\delta_0 = 0$$

$$\sum_{j=0}^{7} \delta_j = 1$$

Ordered monotonic predictors

1 (elementary)

2 (middle school)

3 (some high school)

4 (high school)

5 (some college)

6 (college)

7 (master's)

8 (doctorate)

education level

$$\phi_i = \beta_E \sum_{j=0}^{E_i - 1} \delta_j$$

maximum effect

proportion of maximum effect

Ordered monotonic *priors*

How do we set priors for the delta parameters?

delta parameters form a simplex

Simplex: vector that sums to 1

$$R_i \sim \text{OrderedLogit}(\phi_i, \alpha)$$

$$\phi_i = \beta_E \sum_{j=0}^{E_i - 1} \delta_j + \ldots$$

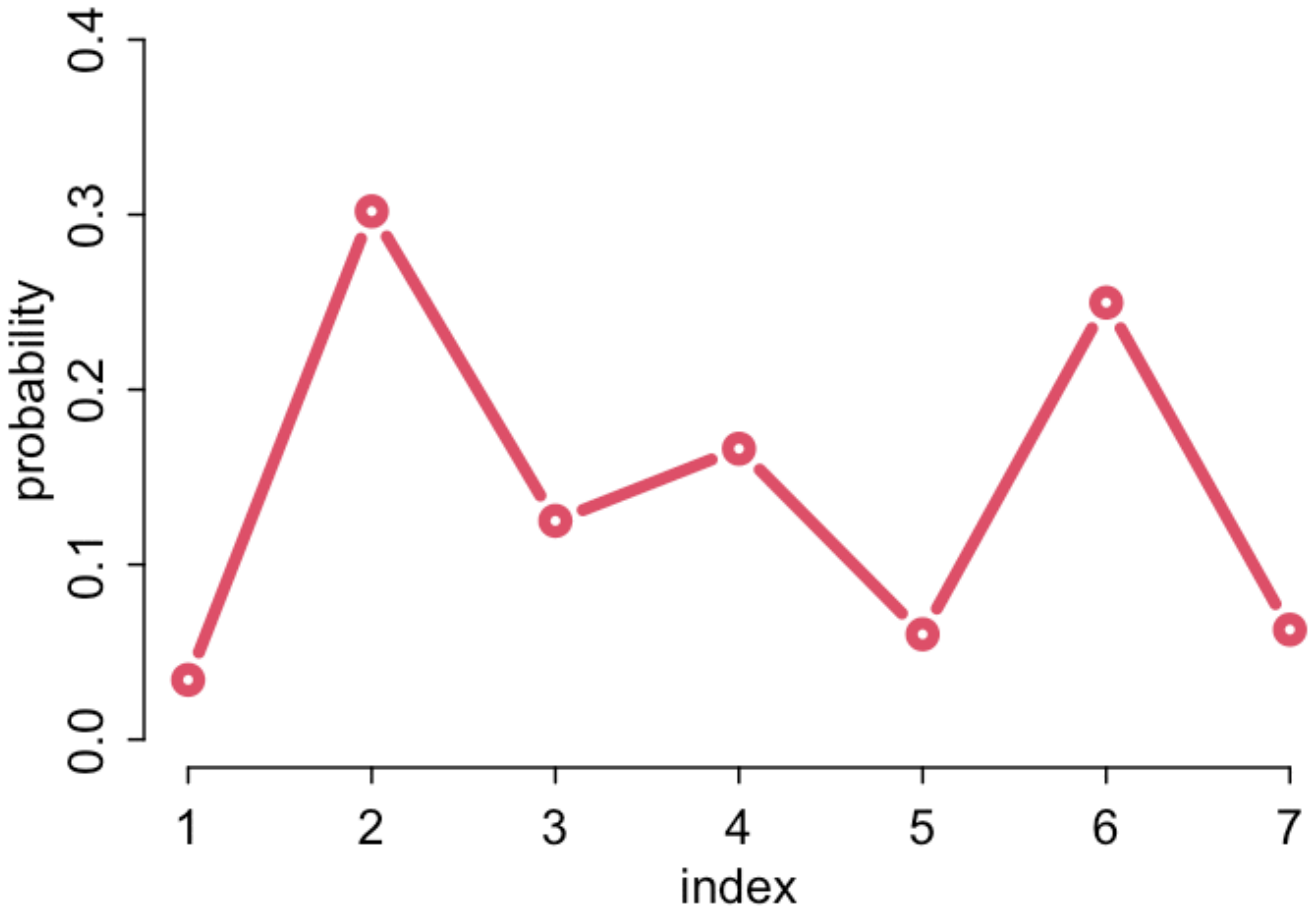
$$\alpha_j \sim \text{Normal}(0,1)$$

$$\beta\_ \sim \text{Normal}(0,0.5)$$

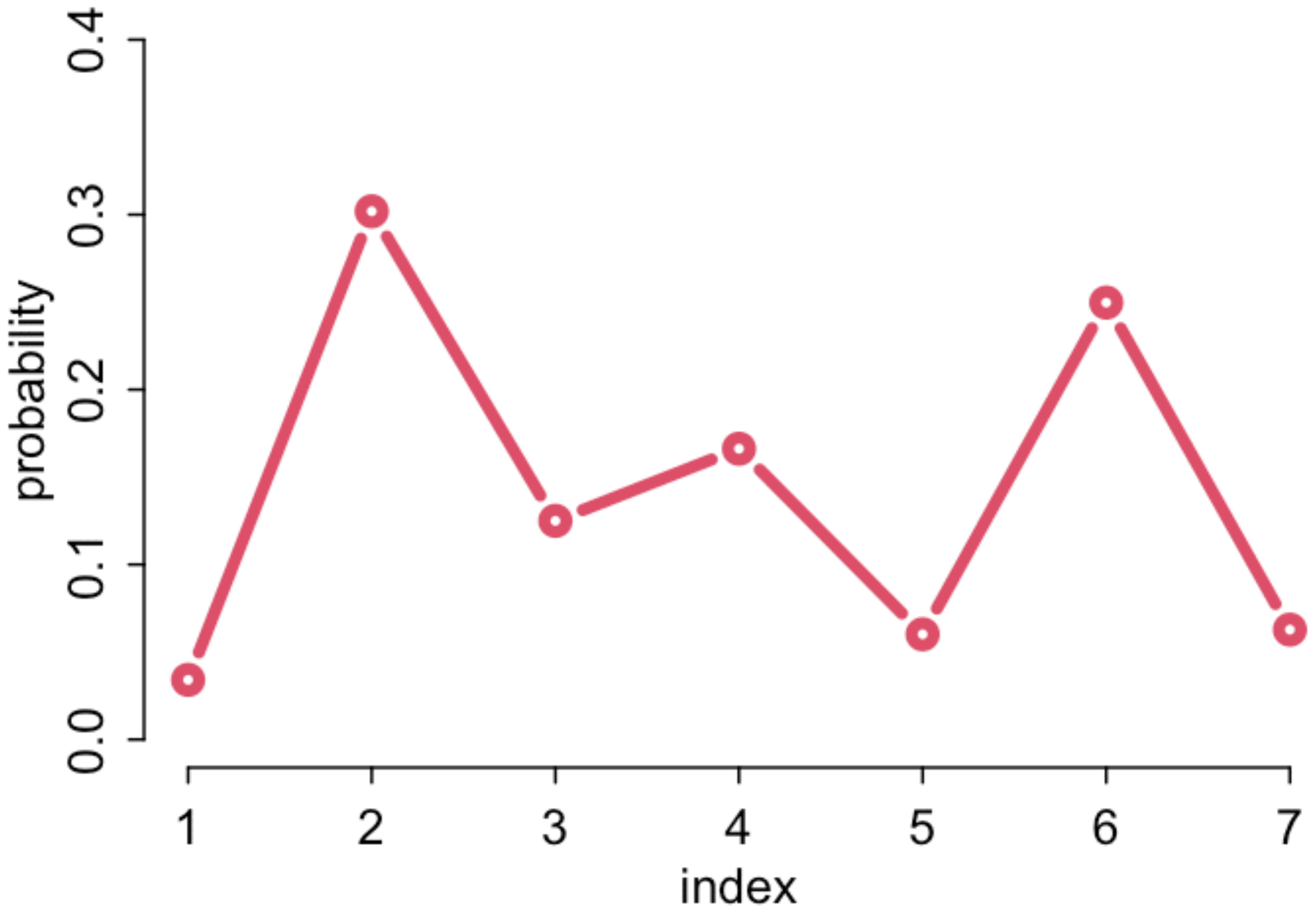
$$\delta_j \sim ?$$

$$\delta \sim \text{Dirichlet}(a)$$

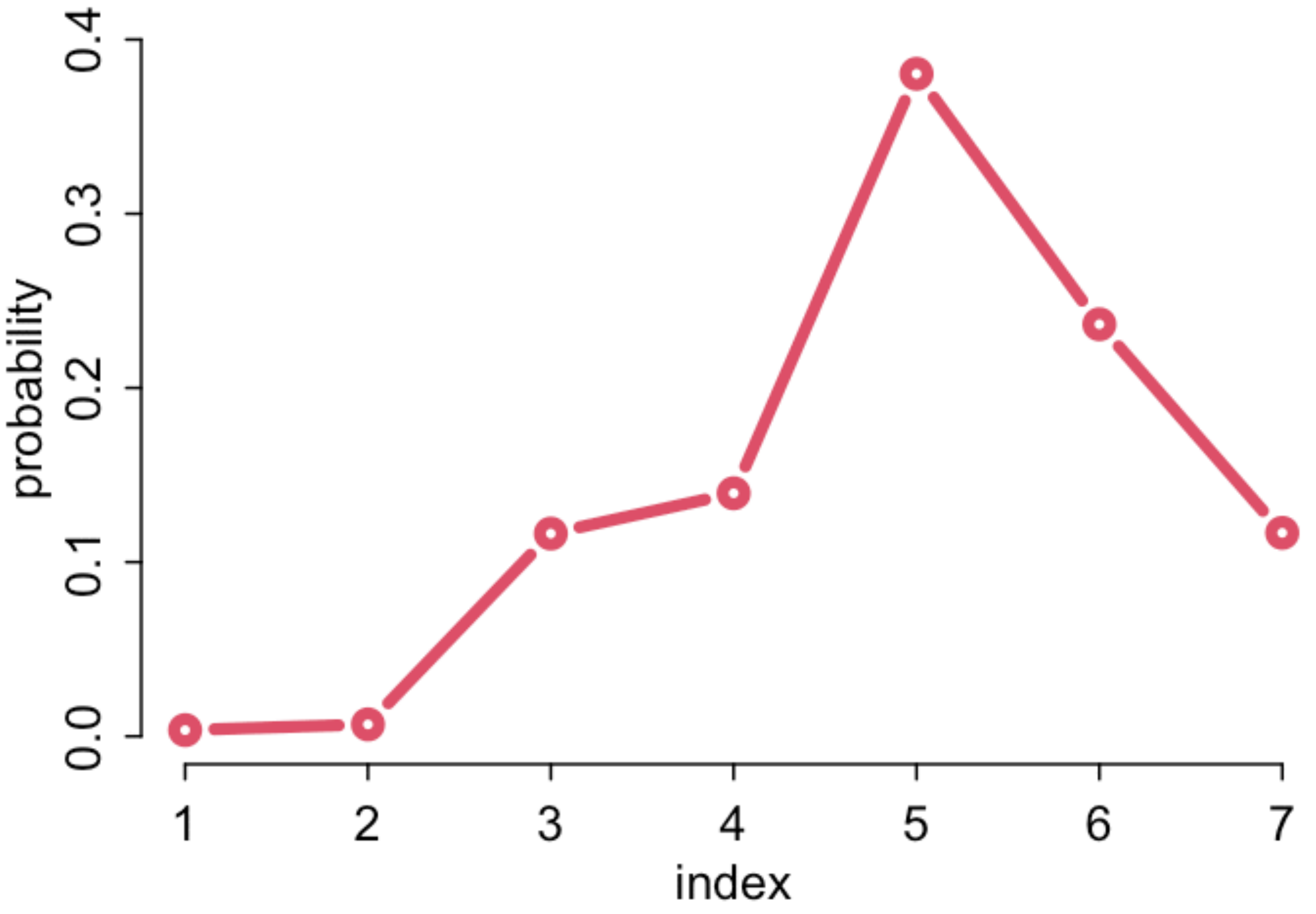
$$a = [2,2,2,2,2,2,2]$$



$$\delta \sim \text{Dirichlet}(a)$$
$$a = [2,2,2,2,2,2,2]$$



$$\delta \sim \text{Dirichlet}(a)$$
$$a = [1,2,3,4,5,6,7]$$



```r
edu_levels <- c( 6 , 1 , 8 , 4 , 7 , 2 , 5 , 3 )
edu_new <- edu_levels[ d$edu ]

dat$E <- edu_new
dat$a <- rep(2,7) # dirichlet prior

mRXE <- ulam(
    alist(

        R ~ ordered_logistic( phi , alpha ),
        phi <- bE*sum( delta_j[1:E] ) +
               bA*A + bI*I + bC*C,
        alpha ~ normal( 0 , 1 ),
        c(bA,bI,bC,bE) ~ normal( 0 , 0.5 ),
        vector[8]: delta_j <<- append_row( 0 , delta ),
        simplex[7]: delta ~ dirichlet( a )

    ), data=dat , chains=4 , cores=4 )
```

$$R_i \sim \text{OrderedLogit}(\phi_i, \alpha)$$

$$\phi_i = \beta_E \sum_{j=0}^{E_i - 1} \delta_j + \ldots$$

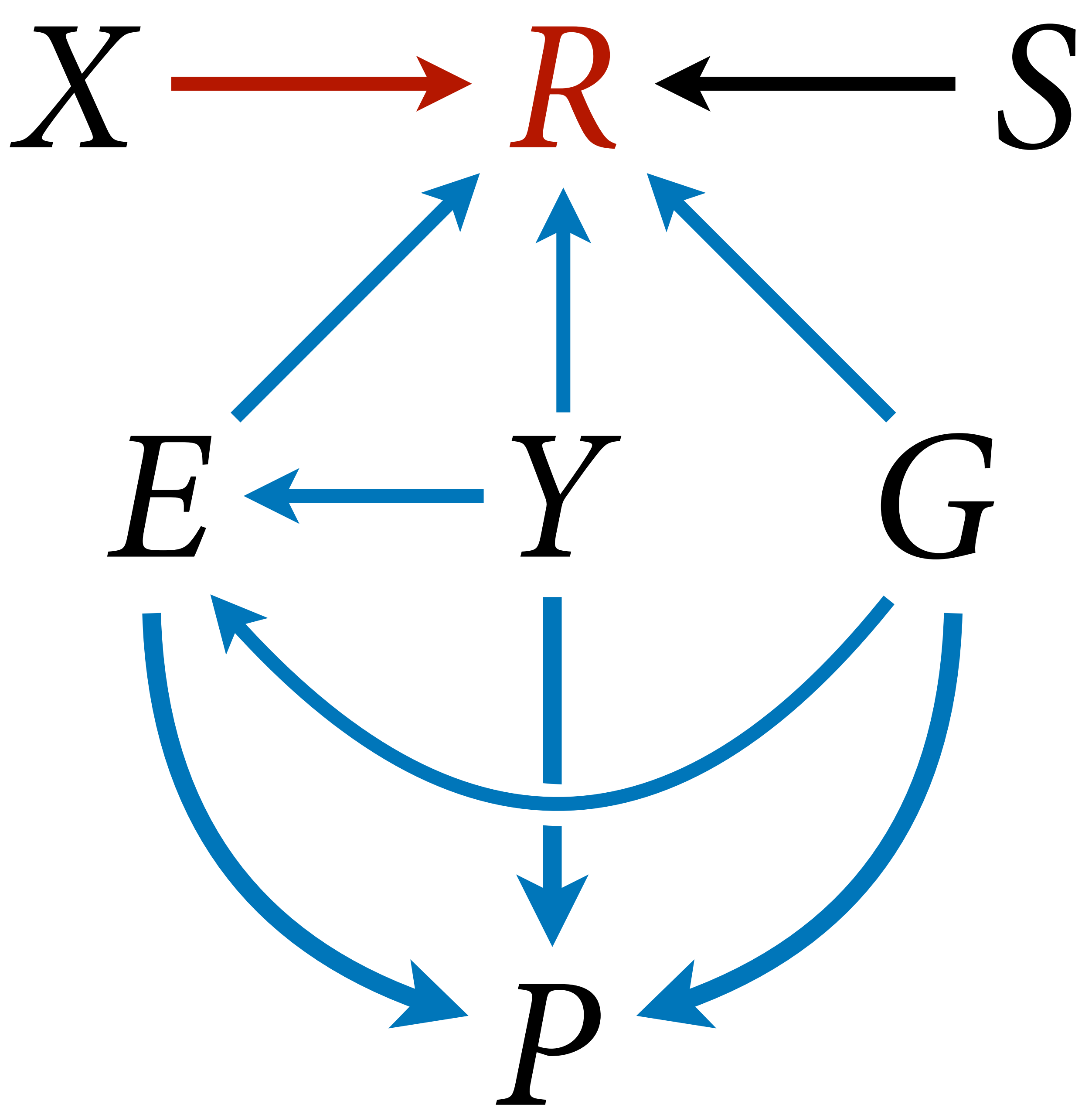
$$\alpha_j \sim \text{Normal}(0,1)$$

$$\beta_{\_} \sim \text{Normal}(0,0.5)$$

$$\delta \sim \text{Dirichlet}(a)$$

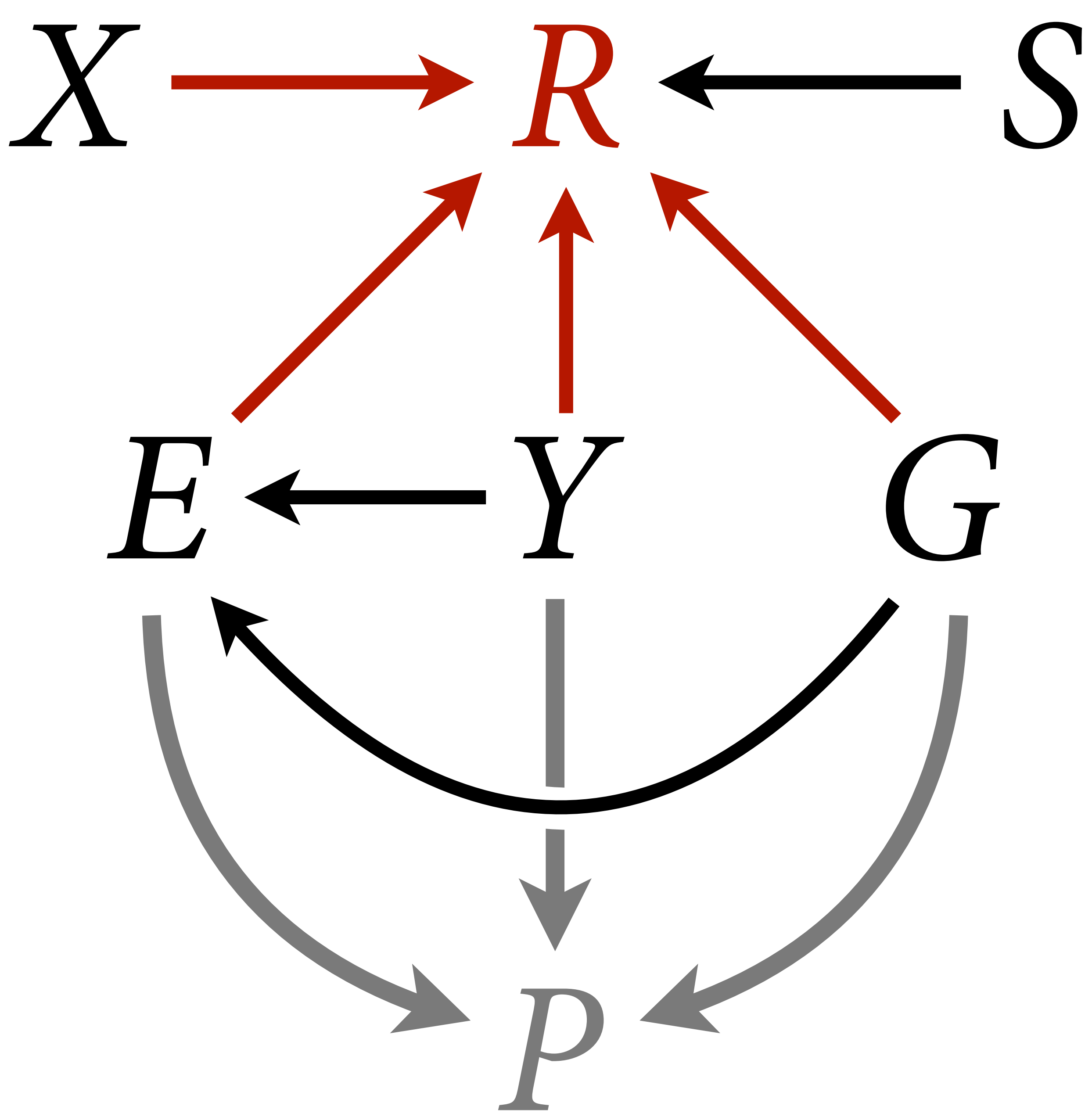
```
edu_levels <- c( 6 , 1 , 8 , 4 , 7 , 2 , 5 , 3 )
edu_new <- edu_levels[ d$edu ]

dat$E <- edu_new
dat$a <- rep(2,7) # dirichlet prior

mRXE <- ulam(
    alist(

        R ~ ordered_logistic( phi , alpha ),
        phi <- bE*sum( delta_j[1:E] ) +
               bA*A + bI*I + bC*C,
        alpha ~ normal( 0 , 1 ),
        c(bA,bI,bC,bE) ~ normal( 0 , 0.5 ),
        vector[8]: delta_j <<- append_row( 0
        simplex[7]: delta ~ dirichlet( a )

    ), data=dat , chains=4 , cores=4 )
```

```
> precis(mRXE,2)
          mean   sd  5.5% 94.5% n_eff Rhat4
alpha[1] -3.07 0.14 -3.32 -2.86   793     1
alpha[2] -2.39 0.14 -2.63 -2.17   804     1
alpha[3] -1.81 0.14 -2.05 -1.60   811     1
alpha[4] -0.79 0.14 -1.03 -0.57   799     1
alpha[5] -0.12 0.14 -0.36  0.10   804     1
alpha[6]  0.79 0.14  0.54  1.00   831     1
bE       -0.31 0.16 -0.57 -0.06   838     1
bC       -0.96 0.05 -1.04 -0.88  1757     1
bI       -0.72 0.04 -0.77 -0.66  1982     1
bA       -0.70 0.04 -0.77 -0.64  1779     1
delta[1]  0.22 0.13  0.05  0.47  1227     1
delta[2]  0.14 0.09  0.03  0.31  2258     1
delta[3]  0.20 0.11  0.05  0.38  2256     1
delta[4]  0.17 0.09  0.04  0.34  1926     1
delta[5]  0.04 0.05  0.01  0.12   945     1
delta[6]  0.10 0.07  0.02  0.23  1870     1
delta[7]  0.13 0.08  0.03  0.27  2335     1
```



bE not interpretable

```
> precis(mRXE,2)
           mean   sd  5.5% 94.5% n_eff Rhat4
alpha[1] -3.07 0.14 -3.32 -2.86   793     1
alpha[2] -2.39 0.14 -2.63 -2.17   804     1
alpha[3] -1.81 0.14 -2.05 -1.60   811     1
alpha[4] -0.79 0.14 -1.03 -0.57   799     1
alpha[5] -0.12 0.14 -0.36  0.10   804     1
alpha[6]  0.79 0.14  0.54  1.00   831     1
bE       -0.31 0.16 -0.57 -0.06   838     1
bC       -0.96 0.05 -1.04 -0.88  1757     1
bI       -0.72 0.04 -0.77 -0.66  1982     1
bA       -0.70 0.04 -0.77 -0.64  1779     1
delta[1]  0.22 0.13  0.05  0.47  1227     1
delta[2]  0.14 0.09  0.03  0.31  2258     1
delta[3]  0.20 0.11  0.05  0.38  2256     1
delta[4]  0.17 0.09  0.04  0.34  1926     1
delta[5]  0.04 0.05  0.01  0.12   945     1
delta[6]  0.10 0.07  0.02  0.23  1870     1
delta[7]  0.13 0.08  0.03  0.27  2335     1
```



$$R_i \sim \text{OrderedLogit}(\phi_i, \alpha)$$

$$\phi_i = \beta_{E,G[i]} \sum_{j=0}^{E_i - 1} \delta_j +$$

$$\beta_{A,G[i]} A_i + \beta_{I,G[i]} I_i + \beta_{C,G[i]} C_i +$$

$$\beta_{Y,G[i]} Y_i$$

$$\alpha_j \sim \text{Normal}(0,1)$$

$$\beta_{\_} \sim \text{Normal}(0,0.5)$$

$$\delta \sim \text{Dirichlet}(a)$$

```
dat$Y <- standardize(d$age)

mRXEYGt <- ulam(
    alist(
        R ~ ordered_logistic( phi , alpha ),
        phi <- bE[G]*sum( delta_j[1:E] ) +
               bA[G]*A + bI[G]*I + bC[G]*C +
               bY[G]*Y,
        alpha ~ normal( 0 , 1 ),
        bA[G] ~ normal( 0 , 0.5 ),
        bI[G] ~ normal( 0 , 0.5 ),
        bC[G] ~ normal( 0 , 0.5 ),
        bE[G] ~ normal( 0 , 0.5 ),
        bY[G] ~ normal( 0 , 0.5 ),
        vector[8]: delta_j <<- append_row( 0 , delta ),
        simplex[7]: delta ~ dirichlet( a )
    ), data=dat , chains=4 , cores=4 , threads=2 )
```

$$R_i \sim \mathrm{OrderedLogit}(\phi_i, \alpha)$$

$$\phi_i = \beta_{E,G[i]} \sum_{j=0}^{E_i-1} \delta_j + \beta_{A,G[i],i} A_i + \beta_{I,G[i]} I_i + \beta_{C,G[i]} C_i + \beta_{Y,G[i]} Y_i$$

$$\alpha_j \sim \mathrm{Normal}(0,1)$$

$$\beta\_ \sim \mathrm{Normal}(0,0.5)$$

$$\delta \sim \mathrm{Dirichlet}(a)$$

```
dat$Y <- standardize(d$age)

mRXEYGt <- ulam(
    alist(
        R ~ ordered_logistic( phi , alpha ),
        phi <- bE[G]*sum( delta_j[1:E] ) +
               bA[G]*A + bI[G]*I + bC[G]*C +
               bY[G]*Y,
        alpha ~ normal( 0 , 1 ),
        bA[G] ~ normal( 0 , 0.5 ),
        bI[G] ~ normal( 0 , 0.5 ),
        bC[G] ~ normal( 0 , 0.5 ),
        bE[G] ~ normal( 0 , 0.5 ),
        bY[G] ~ normal( 0 , 0.5 ),
        vector[8]: delta_j <<- append_row( 0 , delta ),
        simplex[7]: delta ~ dirichlet( a )
    ), data=dat , chains=4 , cores=4 , threads=2 )
```

$$R_i \sim \text{OrderedLogit}(\phi_i, \alpha)$$

$$\phi_i = \beta_{E,G[i]} \sum_{j=0}^{E_i-1} \delta_j + \beta_{A,G[i],i} A_i + \beta_{I,G[i]} I_i + \beta_{C,G[i]} C_i + \beta_{Y,G[i]} Y_i$$

$$\alpha_j \sim \text{Normal}(0,1)$$

$$\beta_{\_} \sim \text{Normal}(0,0.5)$$

$$\delta \sim \text{Dirichlet}(a)$$

4 chains times 2 threads each = 8 cores

```
dat$Y <- standardize(d$age)

mRXEYGt <- ulam(
    alist(
        R ~ ordered_logistic( phi , alpha ),
        phi <- bE[G]*sum( delta_j[1:E] ) +
               bA[G]*A + bI[G]*I + bC[G]*C +
               bY[G]*Y,
        alpha ~ normal( 0 , 1 ),
        bA[G] ~ normal( 0 , 0.5 ),
        bI[G] ~ normal( 0 , 0.5 ),
        bC[G] ~ normal( 0 , 0.5 ),
        bE[G] ~ normal( 0 , 0.5 ),
        bY[G] ~ normal( 0 , 0.5 ),
        vector[8]: delta_j <<- append_row( 0 , delta ),
        simplex[7]: delta ~ dirichlet( a )
    ), data=dat , chains=4 , cores=4 , threads=2 )
```

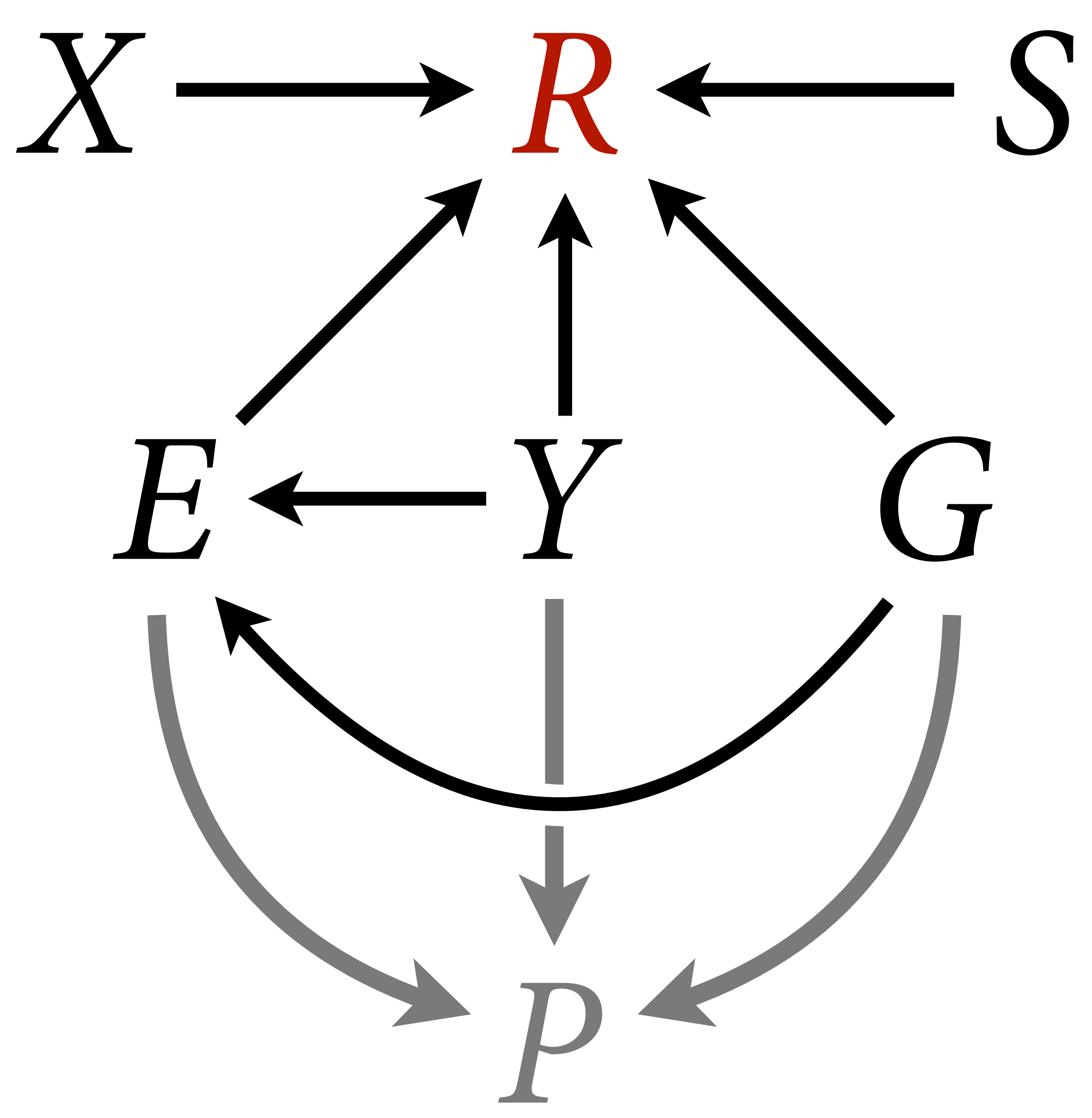
4 chains times 2 threads each = 8 cores

1 thread each

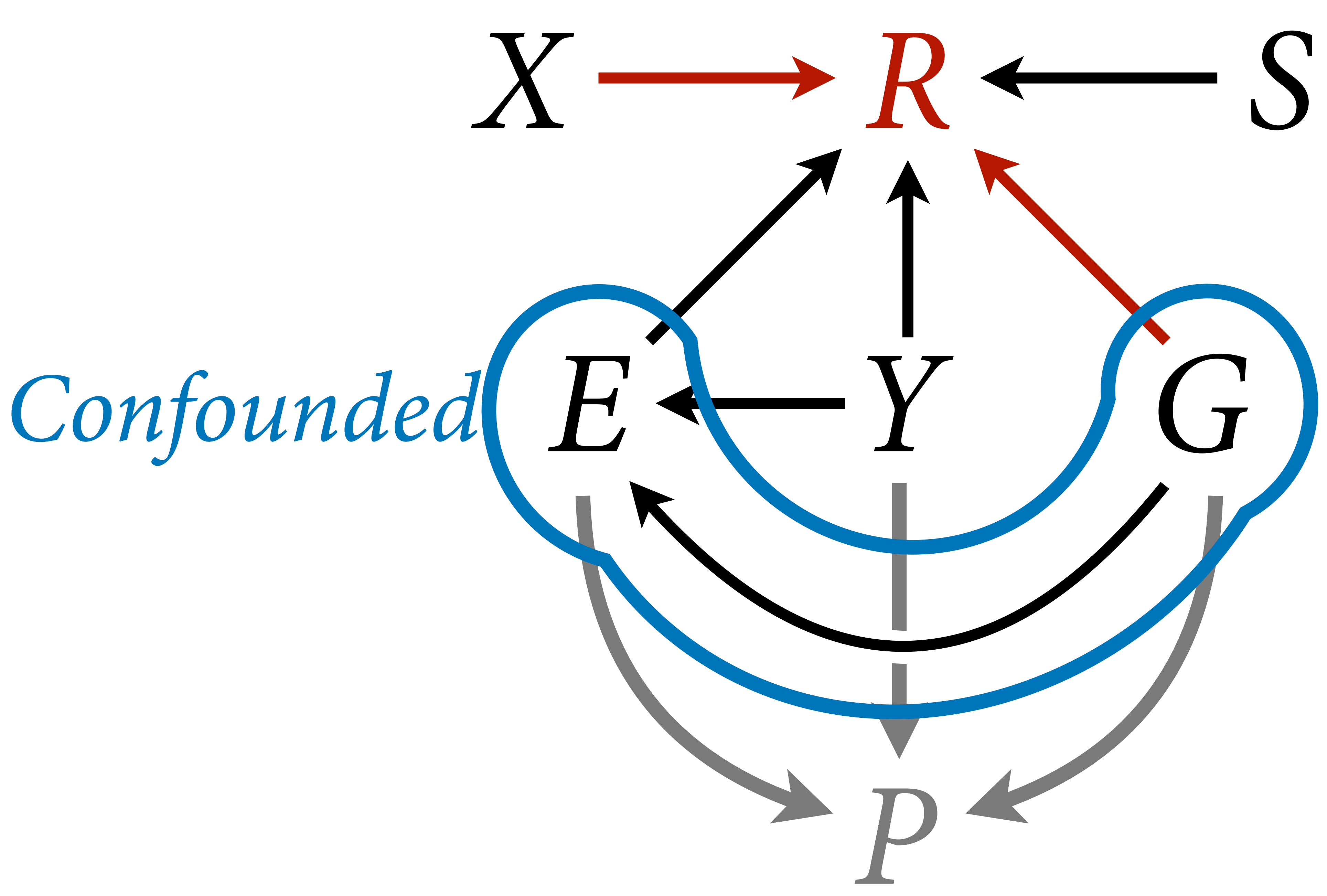
```
Sampling durations (minutes):
        warmup sample total
chain:1   6.53   3.99 10.52
chain:2   7.33   2.66  9.99
chain:3   6.88   3.70 10.58
chain:4   6.40   2.63  9.03
```

2 threads each

```
Sampling durations (minutes):
        warmup sample total
chain:1   4.41   1.80  6.21
chain:2   4.69   1.87  6.56
chain:3   5.14   1.56  6.70
chain:4   4.21   1.84  6.05
```

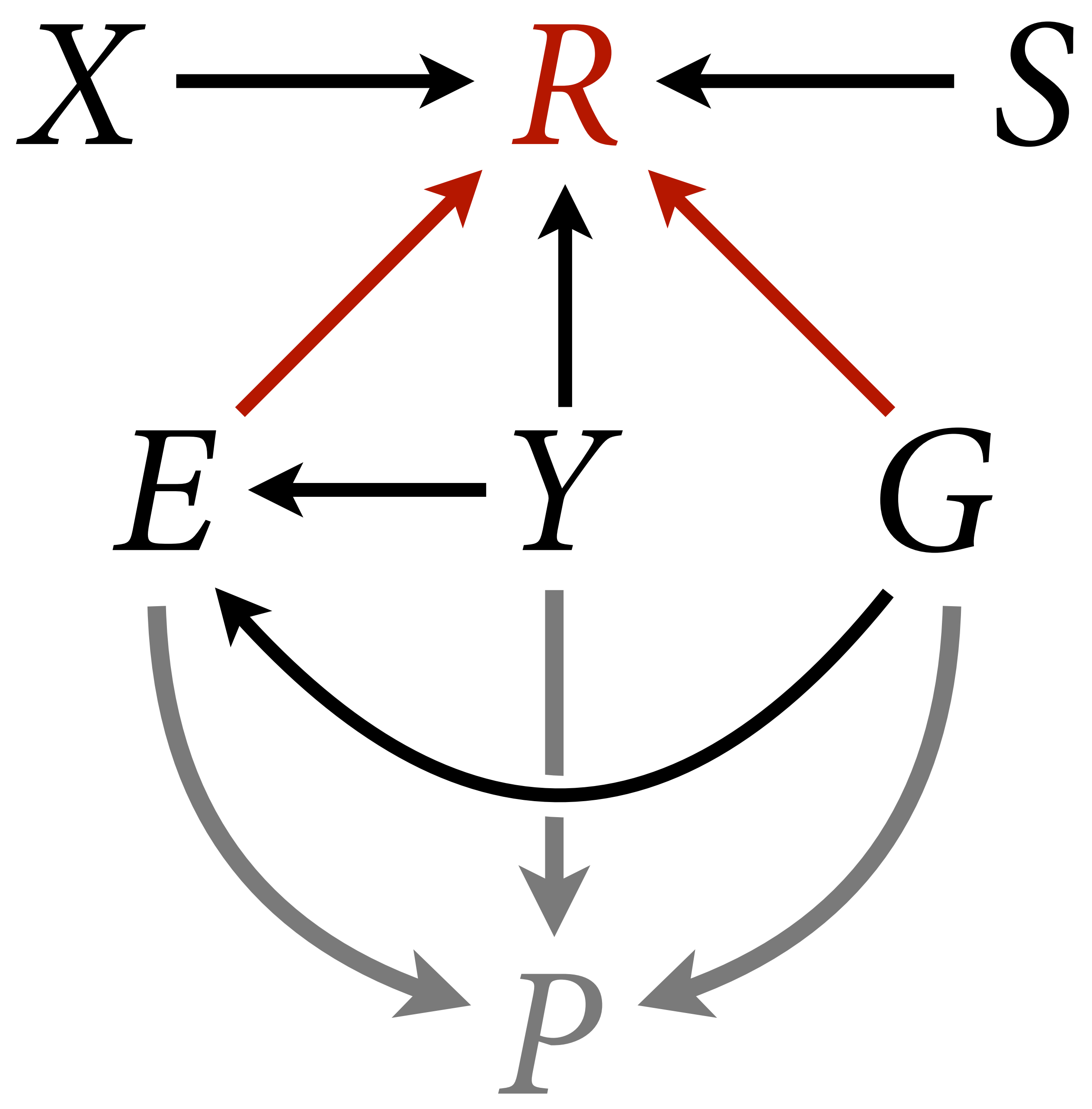



```
> precis(mRXEYGt,2)
          mean   sd  5.5% 94.5% n_eff Rhat4
alpha[1] -2.89 0.10 -3.06 -2.73   729     1
alpha[2] -2.21 0.10 -2.37 -2.06   728     1
alpha[3] -1.62 0.10 -1.78 -1.47   724     1
alpha[4] -0.58 0.10 -0.74 -0.43   729     1
alpha[5]  0.11 0.10 -0.05  0.26   726     1
alpha[6]  1.03 0.10  0.87  1.18   746     1
bA[1]    -0.56 0.06 -0.65 -0.47  1932     1
bA[2]    -0.81 0.05 -0.90 -0.73  2013     1
bI[1]    -0.66 0.05 -0.74 -0.58  2539     1
bI[2]    -0.76 0.05 -0.84 -0.68  2283     1
bC[1]    -0.77 0.07 -0.88 -0.65  2029     1
bC[2]    -1.09 0.07 -1.20 -0.99  2012     1
bE[1]    -0.63 0.14 -0.85 -0.42   810     1
bE[2]     0.41 0.14  0.19  0.62   795     1
bY[1]     0.00 0.03 -0.05  0.05  2740     1
bY[2]    -0.13 0.03 -0.18 -0.09  1426     1
delta[1]  0.15 0.08  0.04  0.31  1759     1
delta[2]  0.15 0.09  0.04  0.30  2440     1
delta[3]  0.29 0.11  0.11  0.46  2001     1
delta[4]  0.08 0.05  0.02  0.17  2414     1
delta[5]  0.06 0.04  0.01  0.14  1087     1
delta[6]  0.24 0.07  0.13  0.34  2301     1
delta[7]  0.04 0.02  0.01  0.08  2755     1
```

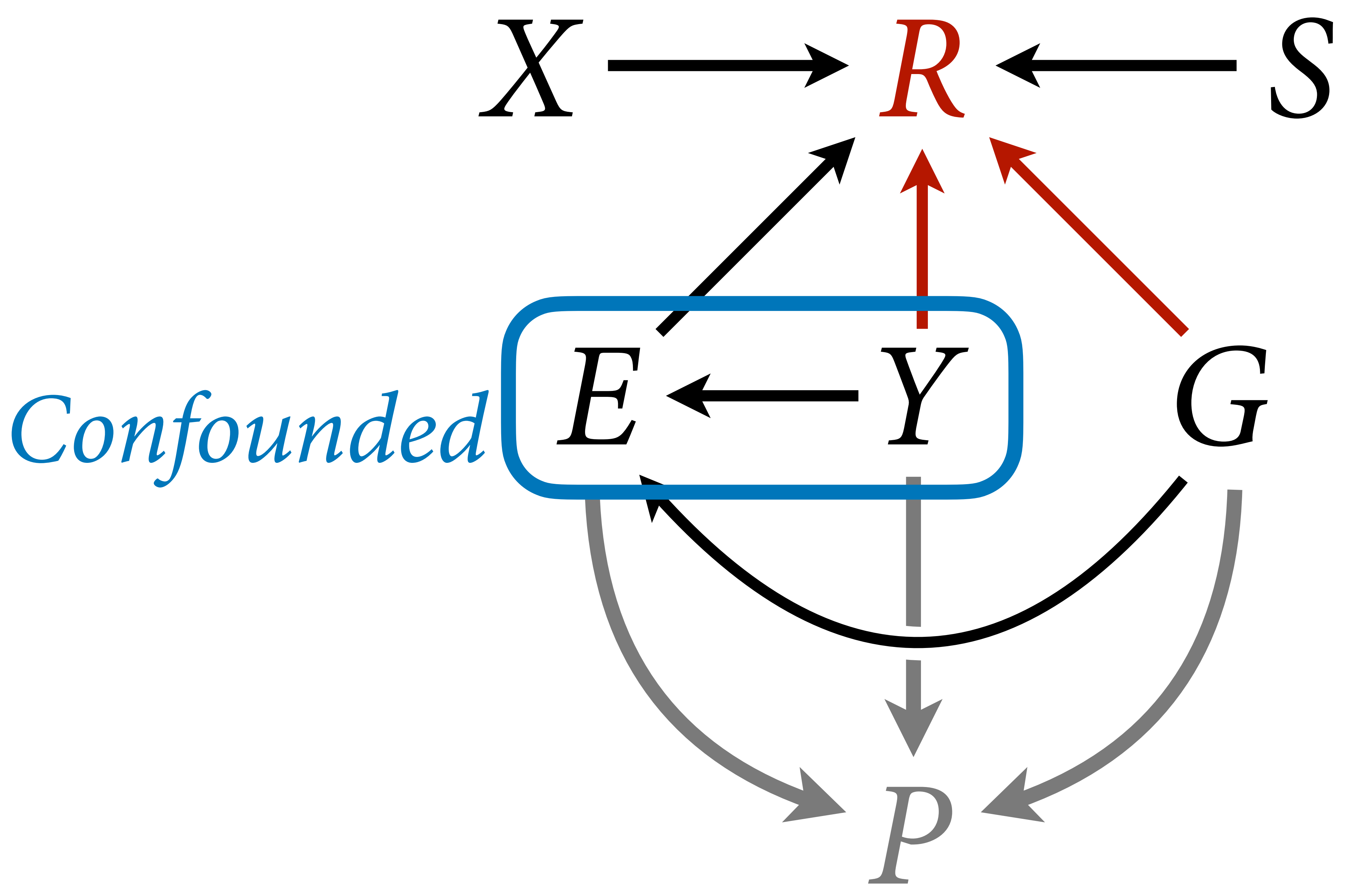


Only direct effect G

```
> precis(mRXEYGt,2)
           mean   sd  5.5% 94.5% n_eff Rhat4
alpha[1] -2.89 0.10 -3.06 -2.73   729     1
alpha[2] -2.21 0.10 -2.37 -2.06   728     1
alpha[3] -1.62 0.10 -1.78 -1.47   724     1
alpha[4] -0.58 0.10 -0.74 -0.43   729     1
alpha[5]  0.11 0.10 -0.05  0.26   726     1
alpha[6]  1.03 0.10  0.87  1.18   746     1
bA[1]    -0.56 0.06 -0.65 -0.47  1932     1
bA[2]    -0.81 0.05 -0.90 -0.73  2013     1
bI[1]    -0.66 0.05 -0.74 -0.58  2539     1
bI[2]    -0.76 0.05 -0.84 -0.68  2283     1
bC[1]    -0.77 0.07 -0.88 -0.65  2029     1
bC[2]    -1.09 0.07 -1.20 -0.99  2012     1
bE[1]    -0.63 0.14 -0.85 -0.42   810     1
bE[2]     0.41 0.14  0.19  0.62   795     1
bY[1]     0.00 0.03 -0.05  0.05  2740     1
bY[2]    -0.13 0.03 -0.18 -0.09  1426     1
delta[1]  0.15 0.08  0.04  0.31  1759     1
delta[2]  0.15 0.09  0.04  0.30  2440     1
delta[3]  0.29 0.11  0.11  0.46  2001     1
delta[4]  0.08 0.05  0.02  0.17  2414     1
delta[5]  0.06 0.04  0.01  0.14  1087     1
delta[6]  0.24 0.07  0.13  0.34  2301     1
delta[7]  0.04 0.02  0.01  0.08  2755     1
```

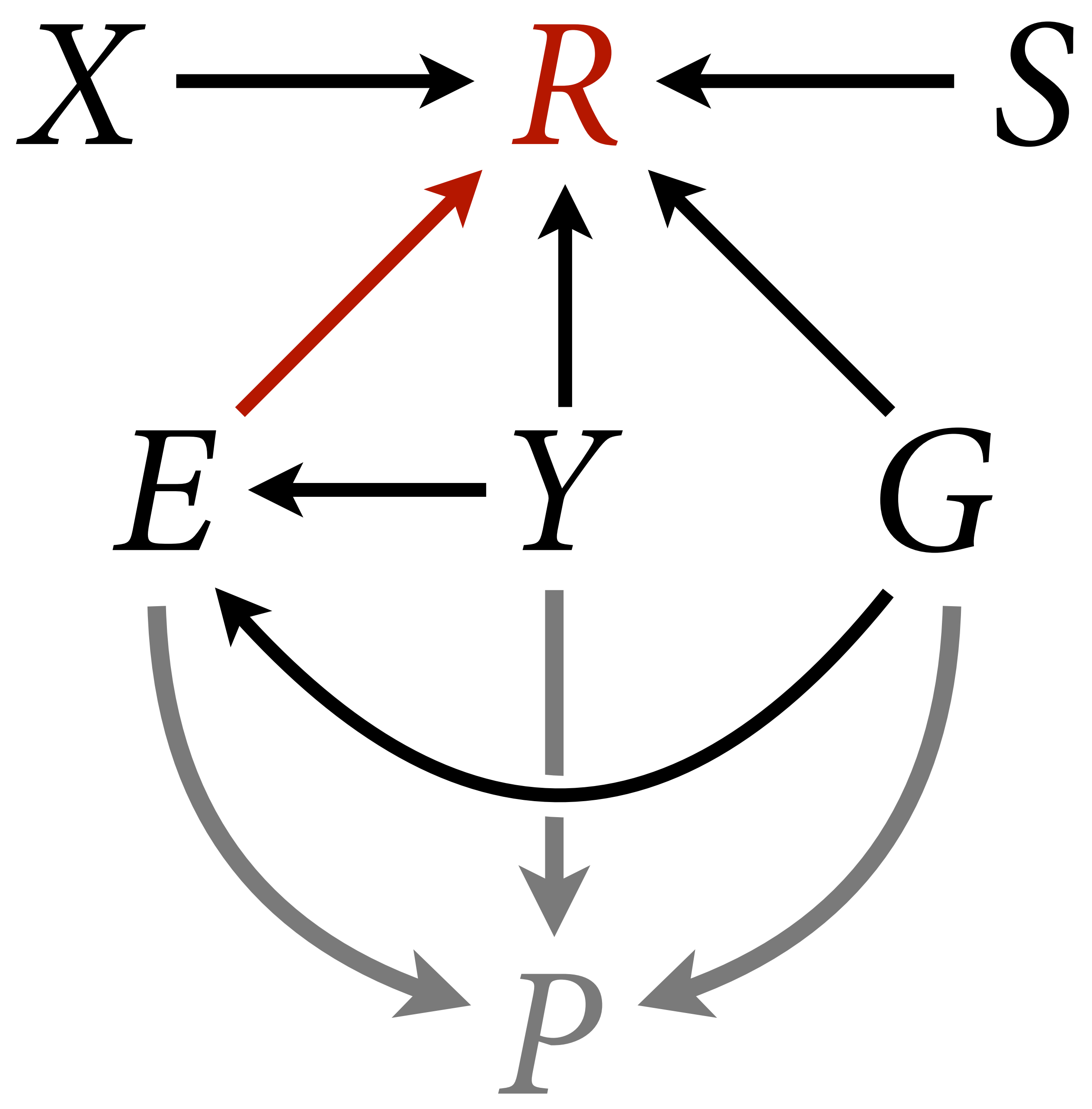



```
> precis(mRXEYGt,2)
           mean   sd  5.5% 94.5% n_eff Rhat4
alpha[1]  -2.89 0.10 -3.06 -2.73   729     1
alpha[2]  -2.21 0.10 -2.37 -2.06   728     1
alpha[3]  -1.62 0.10 -1.78 -1.47   724     1
alpha[4]  -0.58 0.10 -0.74 -0.43   729     1
alpha[5]   0.11 0.10 -0.05  0.26   726     1
alpha[6]   1.03 0.10  0.87  1.18   746     1
bA[1]     -0.56 0.06 -0.65 -0.47  1932     1
bA[2]     -0.81 0.05 -0.90 -0.73  2013     1
bI[1]     -0.66 0.05 -0.74 -0.58  2539     1
bI[2]     -0.76 0.05 -0.84 -0.68  2283     1
bC[1]     -0.77 0.07 -0.88 -0.65  2029     1
bC[2]     -1.09 0.07 -1.20 -0.99  2012     1
bE[1]     -0.63 0.14 -0.85 -0.42   810     1
bE[2]      0.41 0.14  0.19  0.62   795     1
bY[1]      0.00 0.03 -0.05  0.05  2740     1
bY[2]     -0.13 0.03 -0.18 -0.09  1426     1
delta[1]   0.15 0.08  0.04  0.31  1759     1
delta[2]   0.15 0.09  0.04  0.30  2440     1
delta[3]   0.29 0.11  0.11  0.46  2001     1
delta[4]   0.08 0.05  0.02  0.17  2414     1
delta[5]   0.06 0.04  0.01  0.14  1087     1
delta[6]   0.24 0.07  0.13  0.34  2301     1
delta[7]   0.04 0.02  0.01  0.08  2755     1
```



Only direct effect

```
> precis(mRXEYGt,2)
            mean   sd  5.5% 94.5% n_eff Rhat4
alpha[1]   -2.89 0.10 -3.06 -2.73   729     1
alpha[2]   -2.21 0.10 -2.37 -2.06   728     1
alpha[3]   -1.62 0.10 -1.78 -1.47   724     1
alpha[4]   -0.58 0.10 -0.74 -0.43   729     1
alpha[5]    0.11 0.10 -0.05  0.26   726     1
alpha[6]    1.03 0.10  0.87  1.18   746     1
bA[1]      -0.56 0.06 -0.65 -0.47  1932     1
bA[2]      -0.81 0.05 -0.90 -0.73  2013     1
bI[1]      -0.66 0.05 -0.74 -0.58  2539     1
bI[2]      -0.76 0.05 -0.84 -0.68  2283     1
bC[1]      -0.77 0.07 -0.88 -0.65  2029     1
bC[2]      -1.09 0.07 -1.20 -0.99  2012     1
bE[1]      -0.63 0.14 -0.85 -0.42   810     1
bE[2]       0.41 0.14  0.19  0.62   795     1
bY[1]       0.00 0.03 -0.05  0.05  2740     1
bY[2]      -0.13 0.03 -0.18 -0.09  1426     1
delta[1]    0.15 0.08  0.04  0.31  1759     1
delta[2]    0.15 0.09  0.04  0.30  2440     1
delta[3]    0.29 0.11  0.11  0.46  2001     1
delta[4]    0.08 0.05  0.02  0.17  2414     1
delta[5]    0.06 0.04  0.01  0.14  1087     1
delta[6]    0.24 0.07  0.13  0.34  2301     1
delta[7]    0.04 0.02  0.01  0.08  2755     1
```




```
> precis(mRXEYGt,2)
            mean   sd  5.5% 94.5% n_eff Rhat4
alpha[1] -2.89 0.10 -3.06 -2.73   729     1
alpha[2] -2.21 0.10 -2.37 -2.06   728     1
alpha[3] -1.62 0.10 -1.78 -1.47   724     1
alpha[4] -0.58 0.10 -0.74 -0.43   729     1
alpha[5]  0.11 0.10 -0.05  0.26   726     1
alpha[6]  1.03 0.10  0.87  1.18   746     1
bA[1]    -0.56 0.06 -0.65 -0.47  1932     1
bA[2]    -0.81 0.05 -0.90 -0.73  2013     1
bI[1]    -0.66 0.05 -0.74 -0.58  2539     1
bI[2]    -0.76 0.05 -0.84 -0.68  2283     1
bC[1]    -0.77 0.07 -0.88 -0.65  2029     1
bC[2]    -1.09 0.07 -1.20 -0.99  2012     1
bE[1]    -0.63 0.14 -0.85 -0.42   810     1
bE[2]     0.41 0.14  0.19  0.62   795     1
bY[1]     0.00 0.03 -0.05  0.05  2740     1
bY[2]    -0.13 0.03 -0.18 -0.09  1426     1
delta[1]  0.15 0.08  0.04  0.31  1759     1
delta[2]  0.15 0.09  0.04  0.30  2440     1
delta[3]  0.29 0.11  0.11  0.46  2001     1
delta[4]  0.08 0.05  0.02  0.17  2414     1
delta[5]  0.06 0.04  0.01  0.14  1087     1
delta[6]  0.24 0.07  0.13  0.34  2301     1
delta[7]  0.04 0.02  0.01  0.08  2755     1
```

Complex causal effects

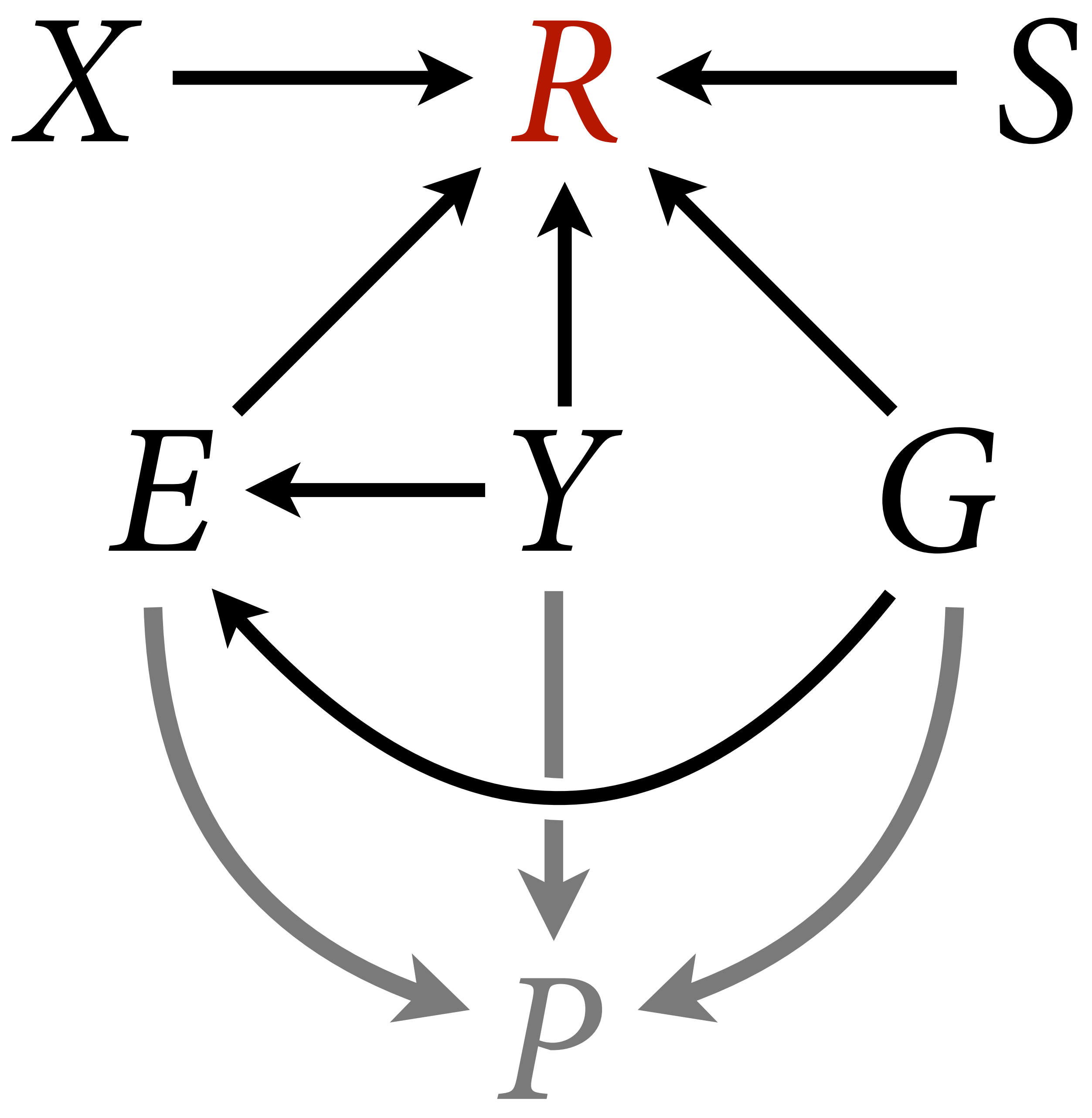
Causal effects (predicted consequences of intervention) require marginalization

Example: Causal effect of E requires distribution of Y and G to average over

Problem 1: Should not marginalize over this sample—*cursed P!* Post-stratify to new target.

Problem 2: Should not set all Y to same E

Example: Causal effect of Y requires effect of Y on E, which we cannot estimate (P again!)



Complex causal effects

Causal effects (predicted consequences of
interve

Examp
of Y an

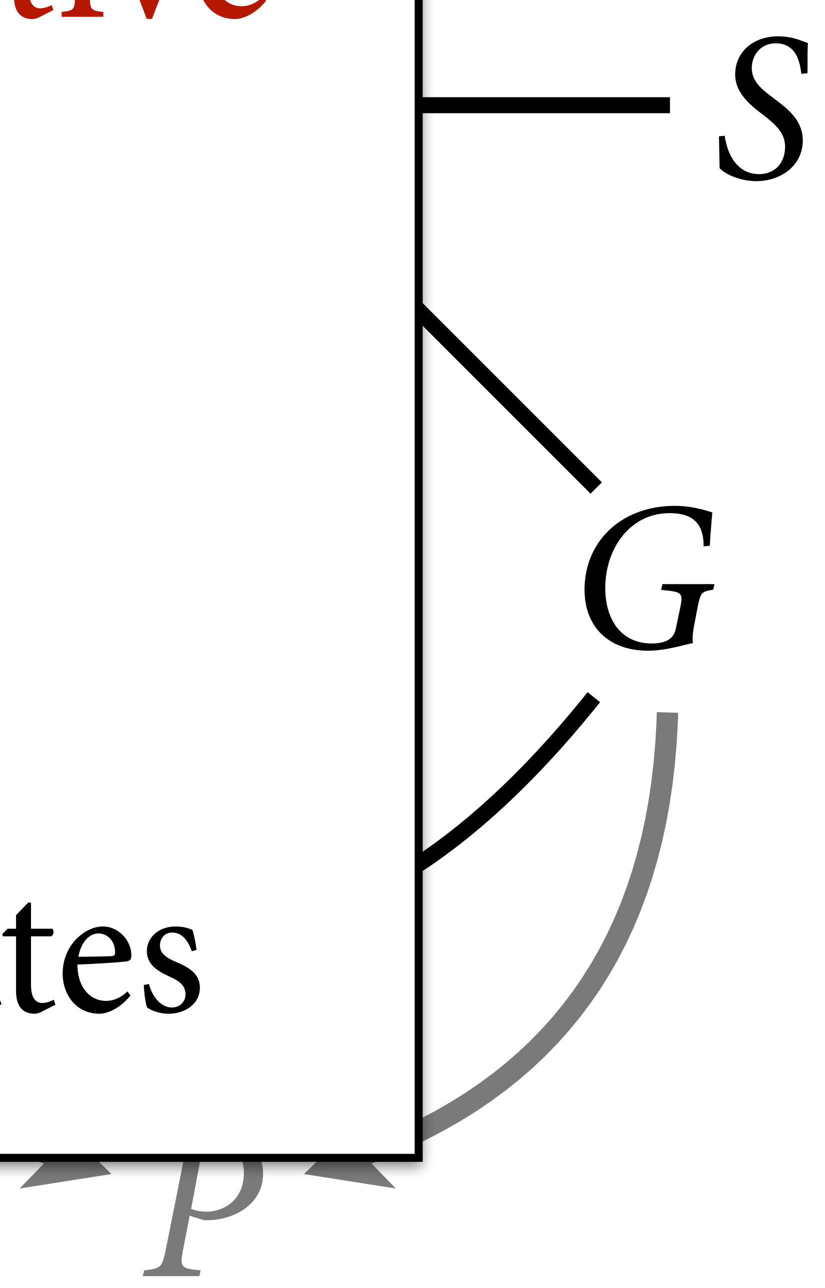
Proble
sample

Proble

No matter how complex, still just a generative simulation using posterior samples

Need generative model to plan estimation

Need generative model to compute estimates

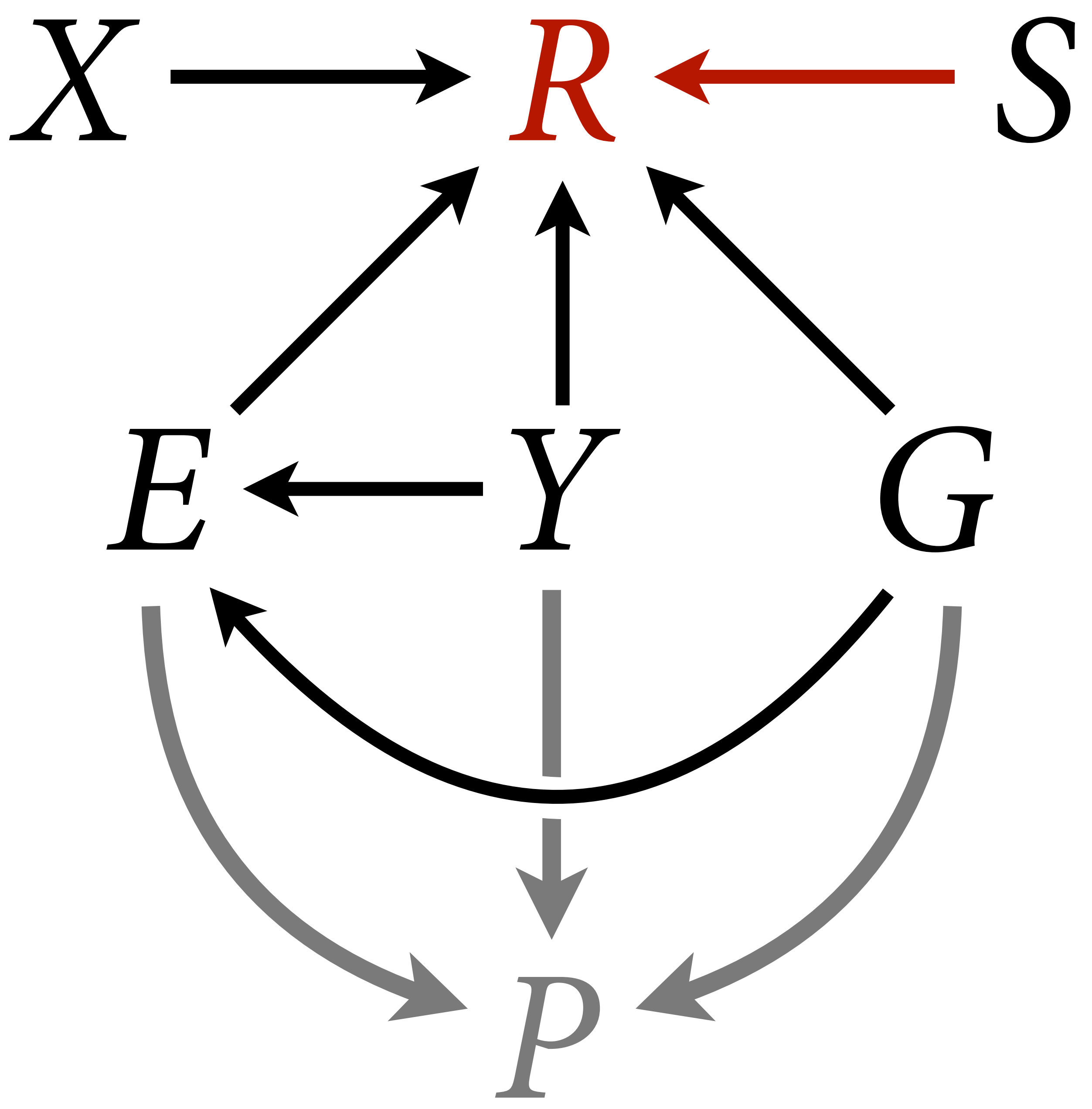


Example: Causal effect of Y requires effect of Y on E, which we cannot estimate (P again!)

Repeat observations

30 stories (S)

```
> table(d$story)

 aqu  boa  box  bur  car  che  pon  rub  sha  shi  spe  swi 
 662  662 1324 1324  662  662  662  662  662  662  993  993 
```



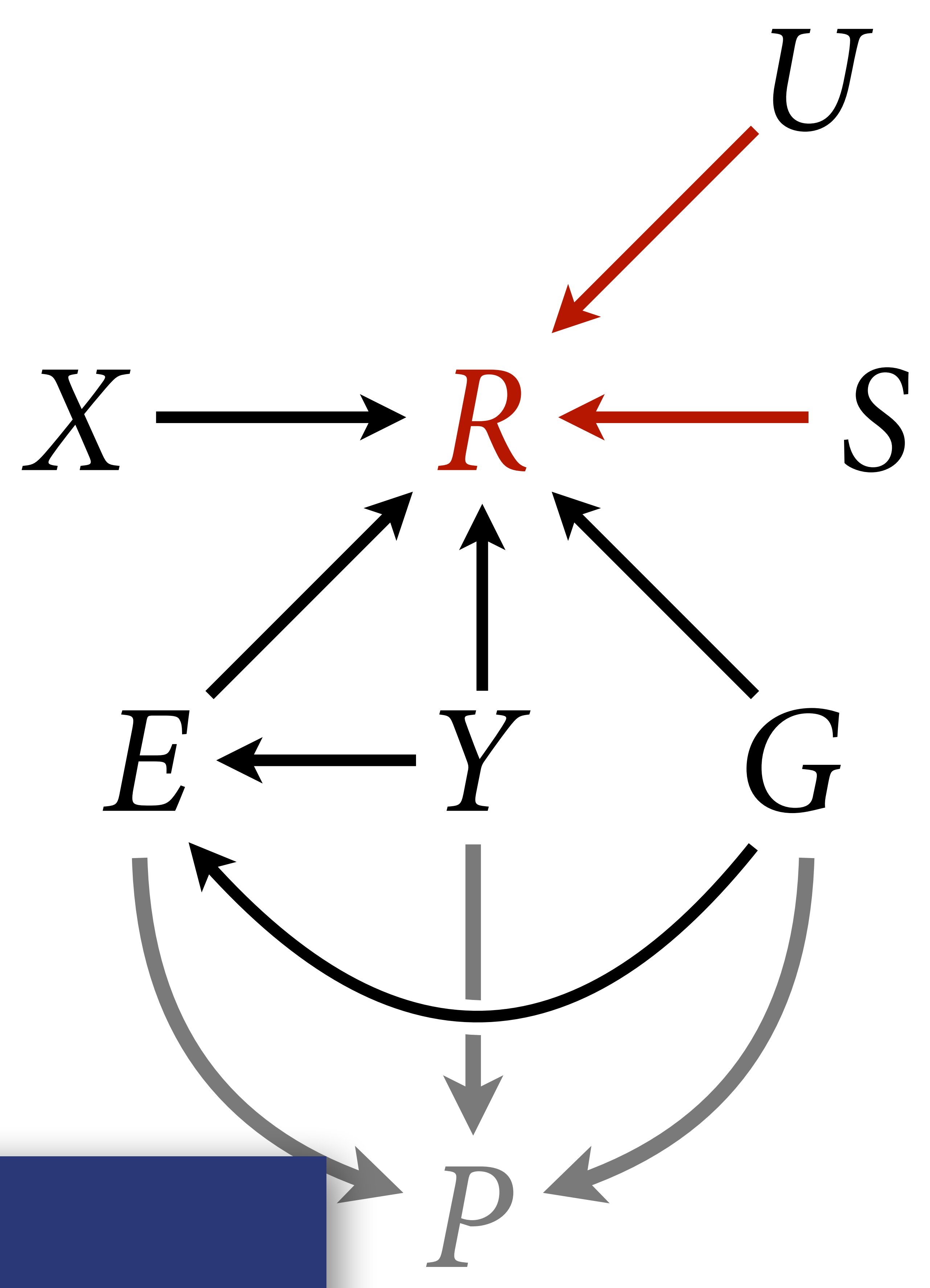
Repeat observations

30 stories (S)

```
> table(d$story)

 aqu  boa  box  bur  car  che  pon  rub  sha  shi  spe  swi 
 662  662 1324 1324  662  662  662  662  662  662  993  993 
```

331 individuals (U)

```
> table(d$id)

96;434 96;445 96;451 96;456 96;458 96;466 96;467 96;474 96;480 96;481 96;497 
    30     30     30     30     30     30     30     30     30     30     30 
96;498 96;502 96;505 96;511 96;512 96;518 96;519 96;531 96;533 96;538 96;547 
    30     30     30     30     30     30     30     30     30     30     30 
96;550 96;553 96;555 96;558 96;560 96;562 96;566 96;570 96;581 96;586 96;591 
    30     30     30     30     30     30     30     30     30     30     30 
```



Course Schedule

Week 1	Bayesian inference	Chapters 1, 2, 3
Week 2	Linear models & Causal Inference	Chapter 4
Week 3	Causes, Confounds & Colliders	Chapters 5 & 6
Week 4	Overfitting / MCMC	Chapters 7, 8, 9
Week 5	Generalized Linear Models	Chapters 10, 11
Week 6	Ordered categories & Multilevel models	Chapters 12 & 13
Week 7	More Multilevel models	Chapters 13 & 14
Week 8	Multilevel models & Gaussian processes	Chapter 14
Week 9	Measurement & Missingness	Chapter 15
Week 10	Generalized Linear Madness	Chapter 16

`https://github.com/rmcelreath/stat_rethinking_2023`

BONUS

	Data Science Task		
	Description	**Prediction**	**Causal inference**
Example of scientific question	How can women aged 60–80 years with stroke history be partitioned in classes defined by their characteristics?	What is the probability of having a stroke next year for women with certain characteristics?	Will starting a statin reduce, on average, the risk of stroke in women with certain characteristics?

Hernán et al. A second chance to get causal inference right

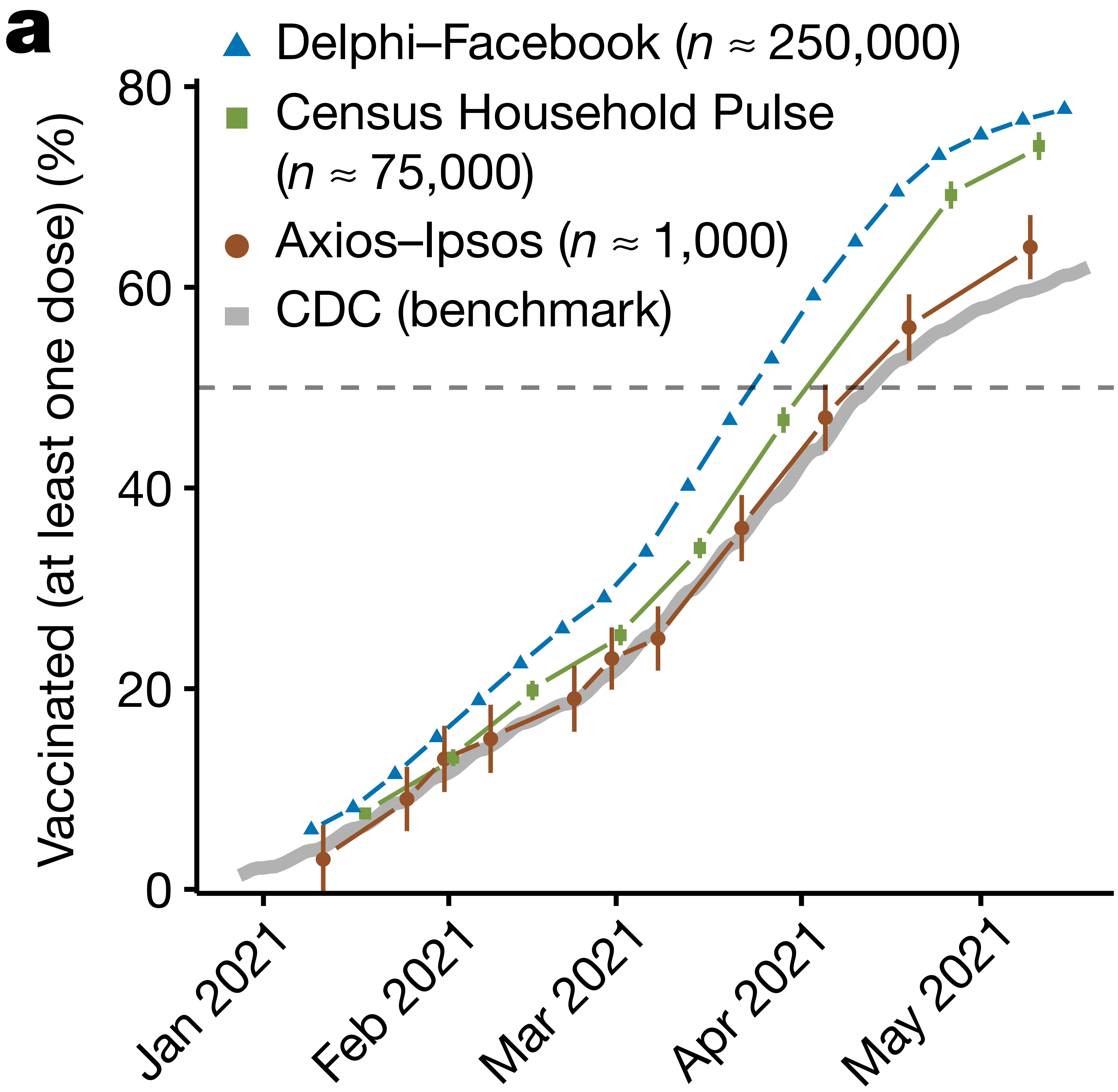
	Data Science Task		
	Description	**Prediction**	**Causal inference**
Example of scientific question	How can women aged 60–80 years with stroke history be partitioned in classes defined by their characteristics?	What is the probability of having a stroke next year for women with certain characteristics?	Will starting a statin reduce, on average, the risk of stroke in women with certain characteristics?
Data	• Eligibility criteria • Features (symptoms, clinical parameters …)	• Eligibility criteria • Output (diagnosis of stroke over the next year) • Inputs (age, blood pressure, history of stroke, diabetes at baseline)	• Eligibility criteria • Outcome (diagnosis of stroke over the next year) • Treatment (initiation of statins at baseline) • Confounders • Effect modifiers (optional)

Hernán et al. A second chance to get causal inference right

	Data Science Task		
	Description	**Prediction**	**Causal inference**
Example of scientific question	How can women aged 60–80 years with stroke history be partitioned in classes defined by their characteristics?	What is the probability of having a stroke next year for women with certain characteristics?	Will starting a statin reduce, on average, the risk of stroke in women with certain characteristics?
Data	• Eligibility criteria • Features (symptoms, clinical parameters …)	• Eligibility criteria • Output (diagnosis of stroke over the next year) • Inputs (age, blood pressure, history of stroke, diabetes at baseline)	• Eligibility criteria • Outcome (diagnosis of stroke over the next year) • Treatment (initiation of statins at baseline) • Confounders • Effect modifiers (optional)
Examples of analytics	Cluster analysis …	Regression Decision trees Random forests Support vector machines Neural networks …	Regression Matching Inverse probability weighting G-formula G-estimation Instrumental variable estimation …

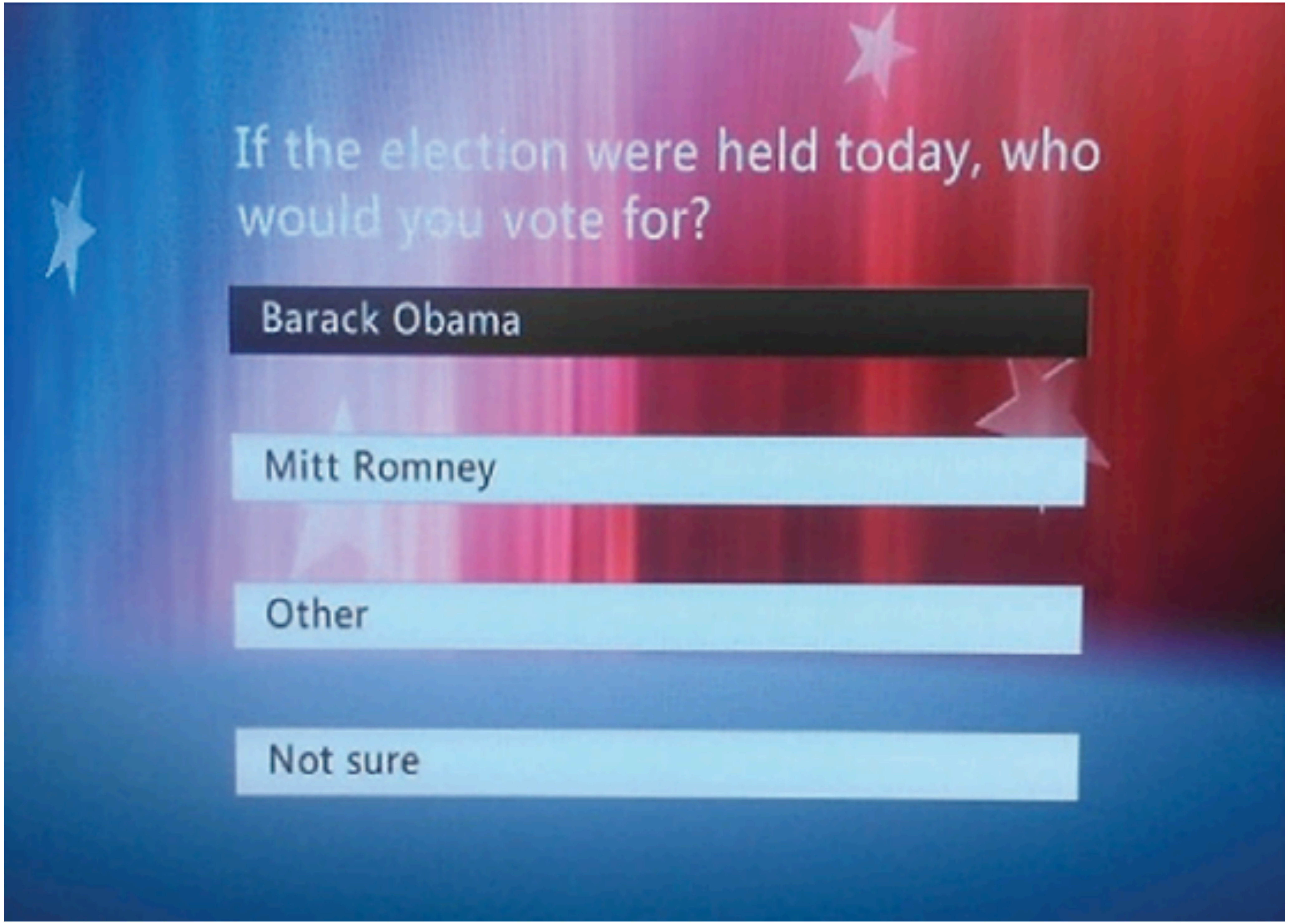
WRONG

Hernán et al. A second chance to get causal inference right

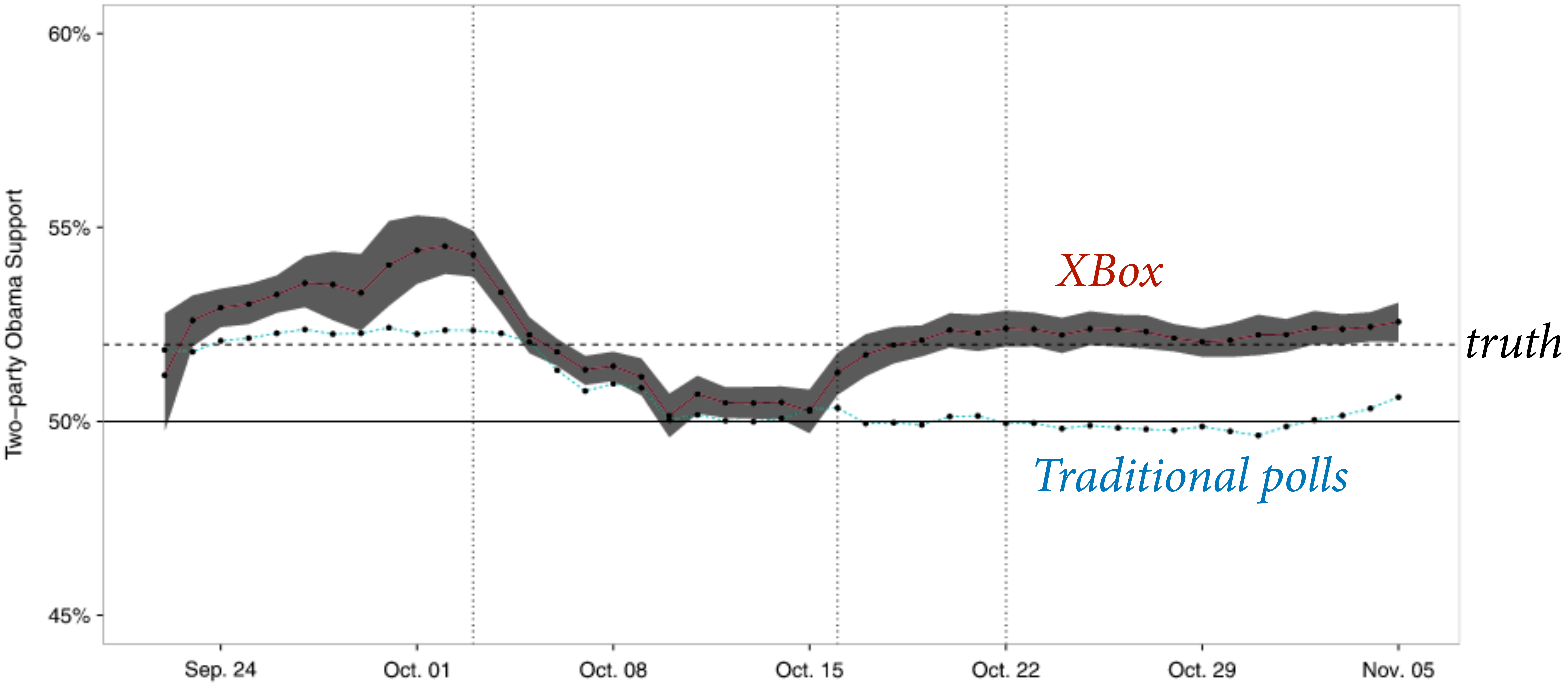


Bradley et al. 2021 Unrepresentative big surveys significantly overestimated US vaccine uptake

XBOX



Wang et al. 2014. Forecasting elections with non-representative polls



Wang et al. 2014. Forecasting elections with non-representative polls

Hitting the Target

Basic problem: **Sample** is not the **target**

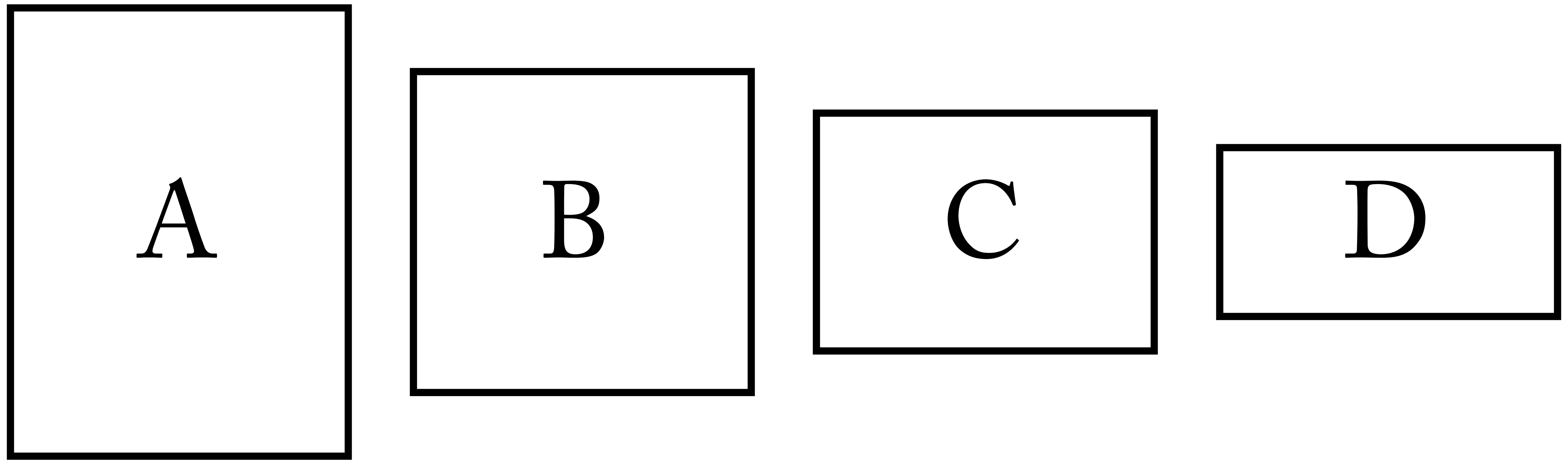
Post-stratification & Transport: Transparent, principled methods for extrapolating from **sample** to **population**

Post-strat requires casual model of reasons sample differs from population

NO CAUSES IN; NO DESCRIPTION OUT

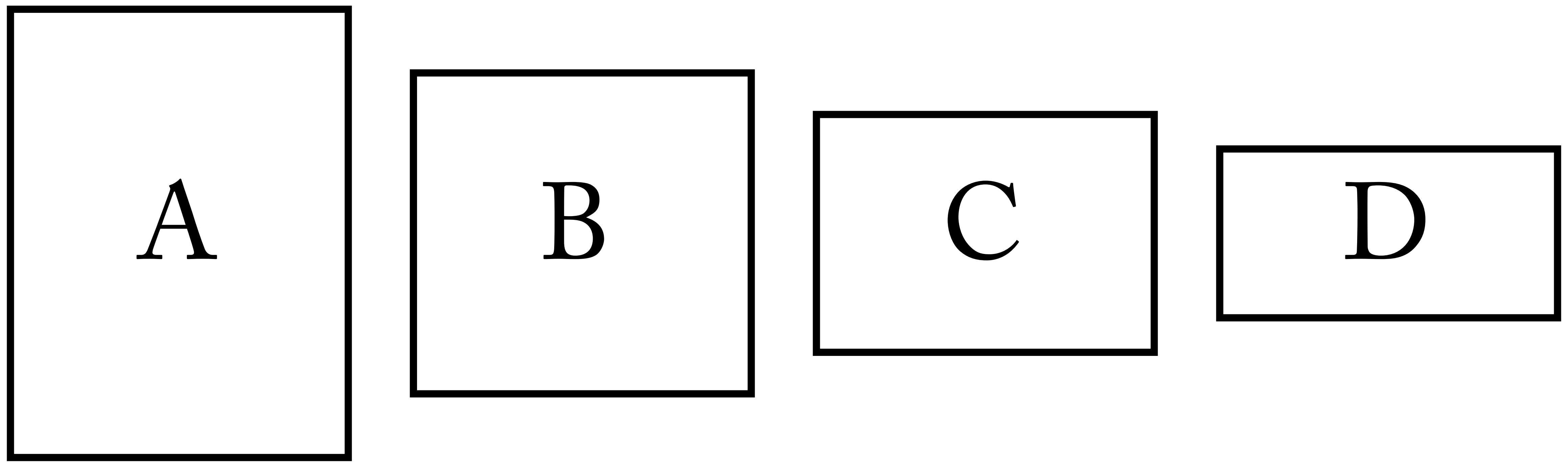
Cartoon example

Four age groups:

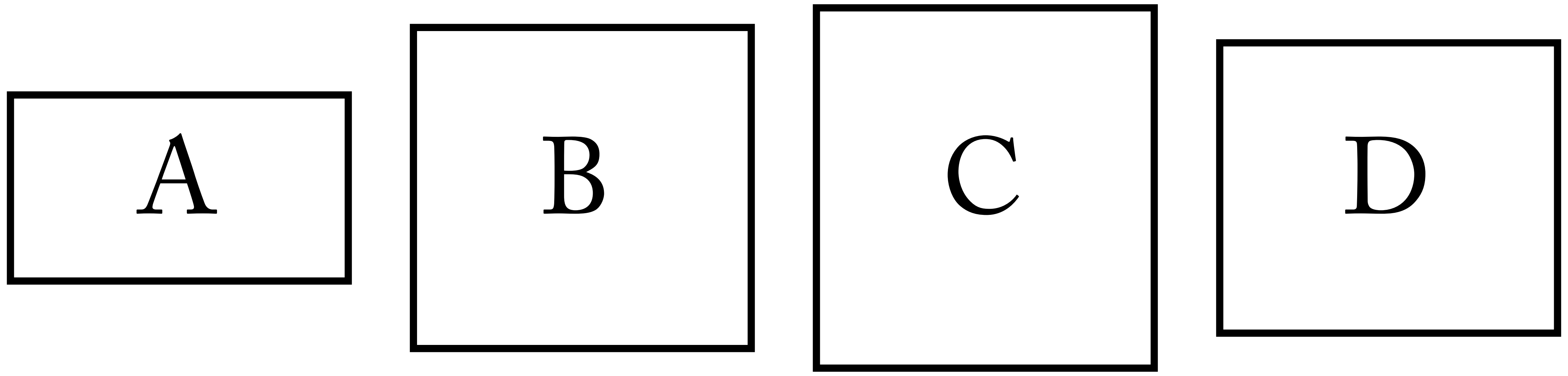


Cartoon example

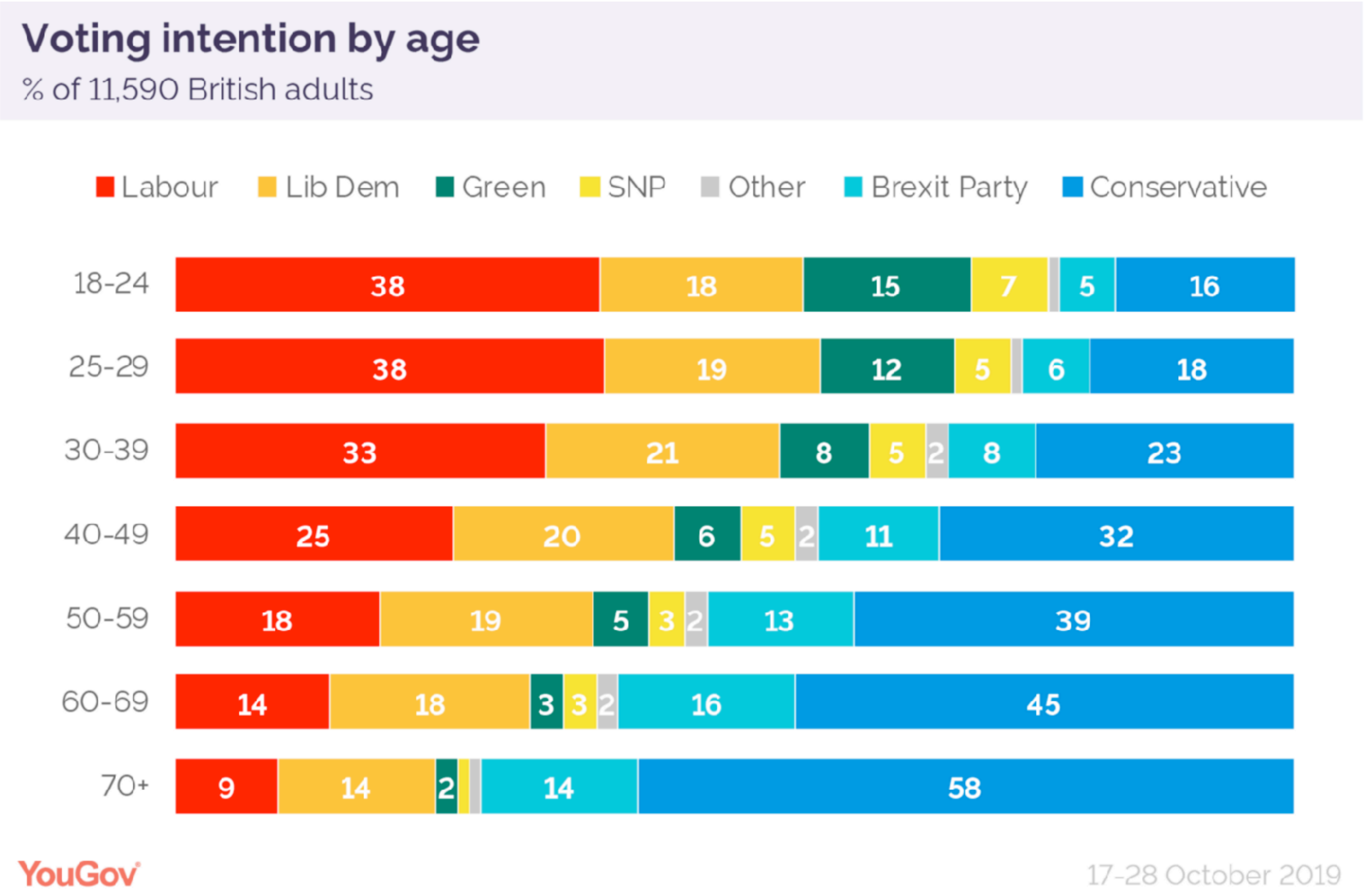
Four age groups:



Proportions of sample:



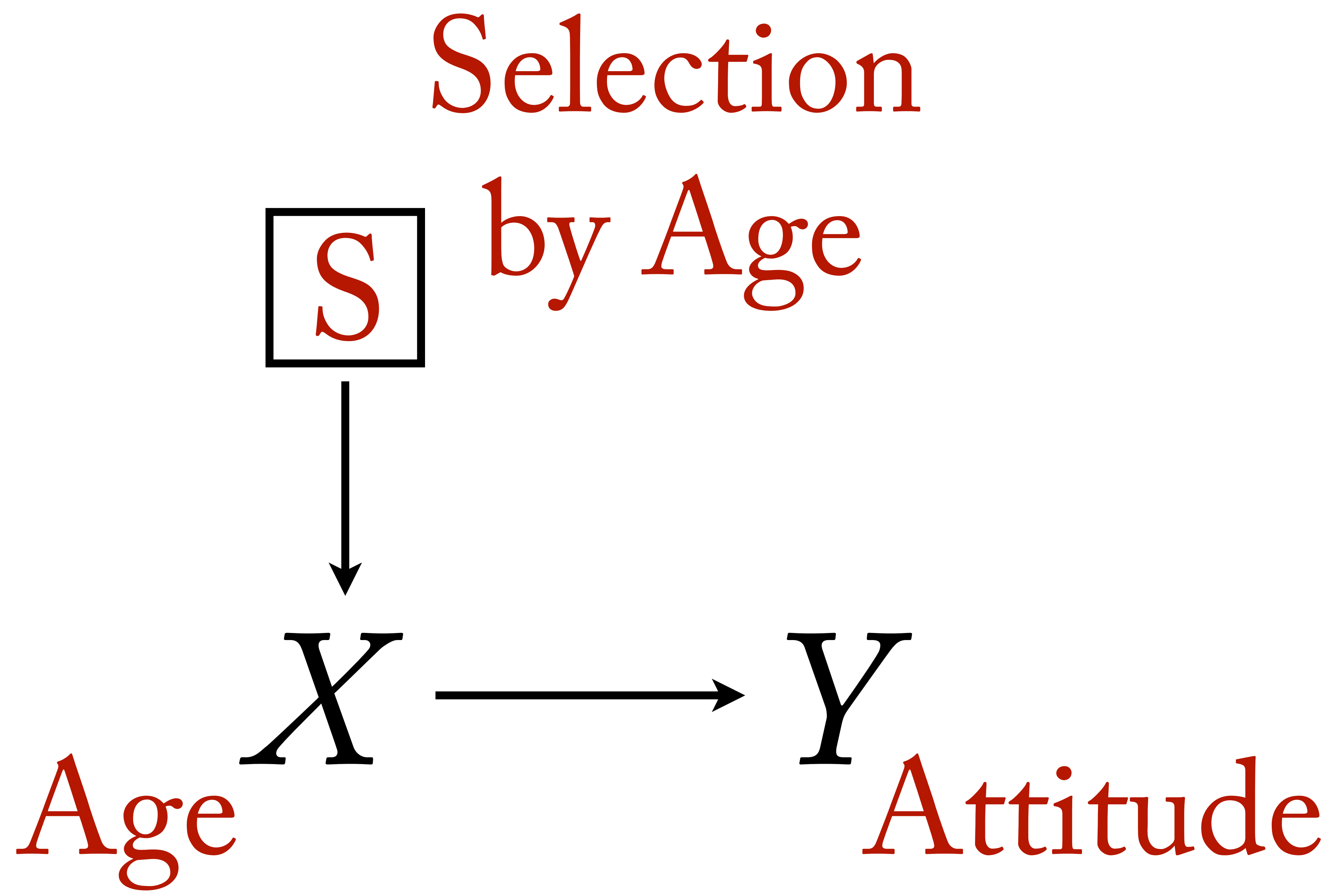
Multi-level regression & post-stratification (MRP)



Selection nodes

$${}_{\text{Age}}X \longrightarrow Y_{\text{Attitude}}$$

Selection nodes



$\boxed{S}$: “Sample differs because of differences in what I point to”

Selection ubiquitous

Many sources of data are already filtered by selection effects

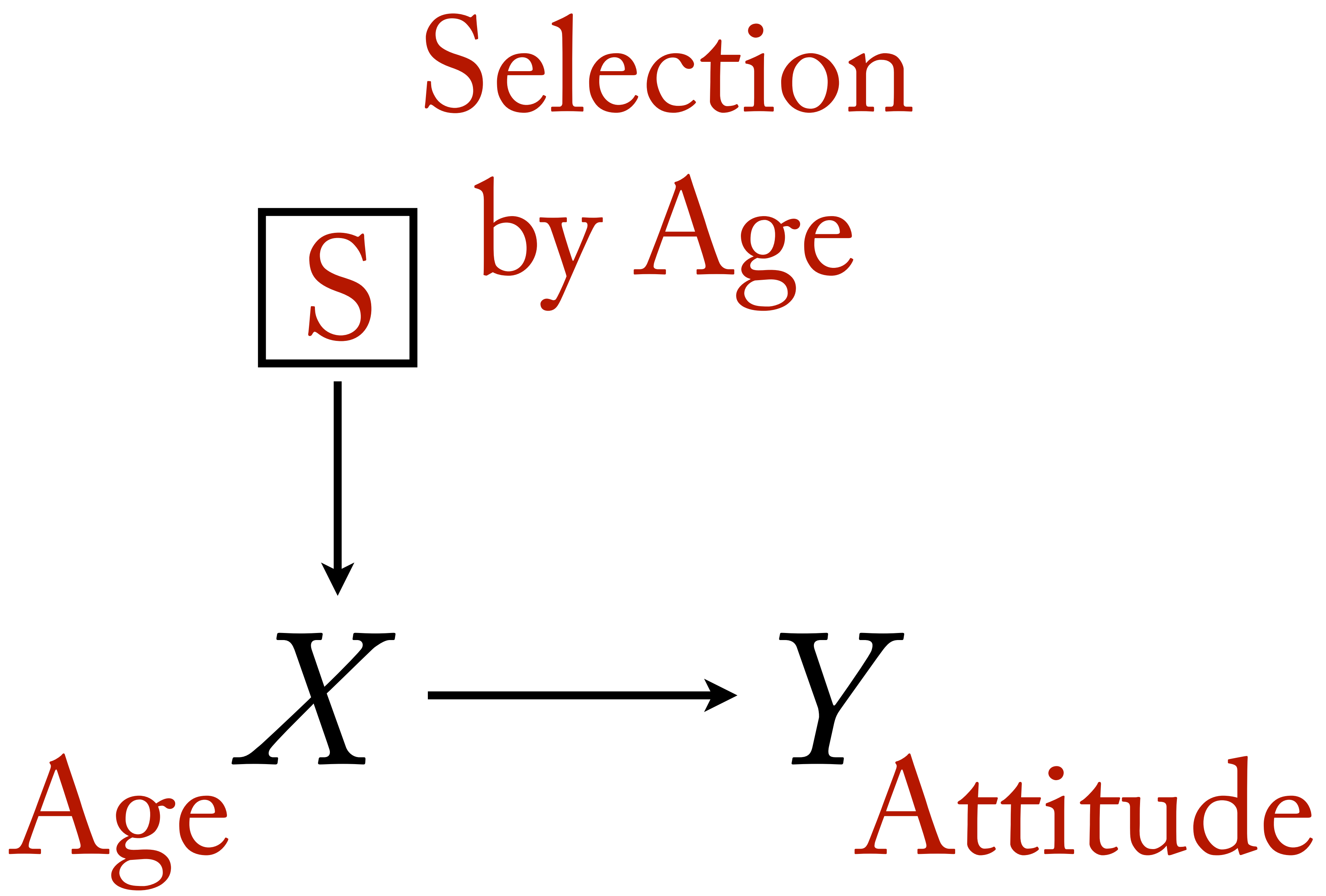
Crime & health statistics

Employment & job performance

Museum collections

Right thing to do depends upon causes of selection

“Young people don’t answer their phones”



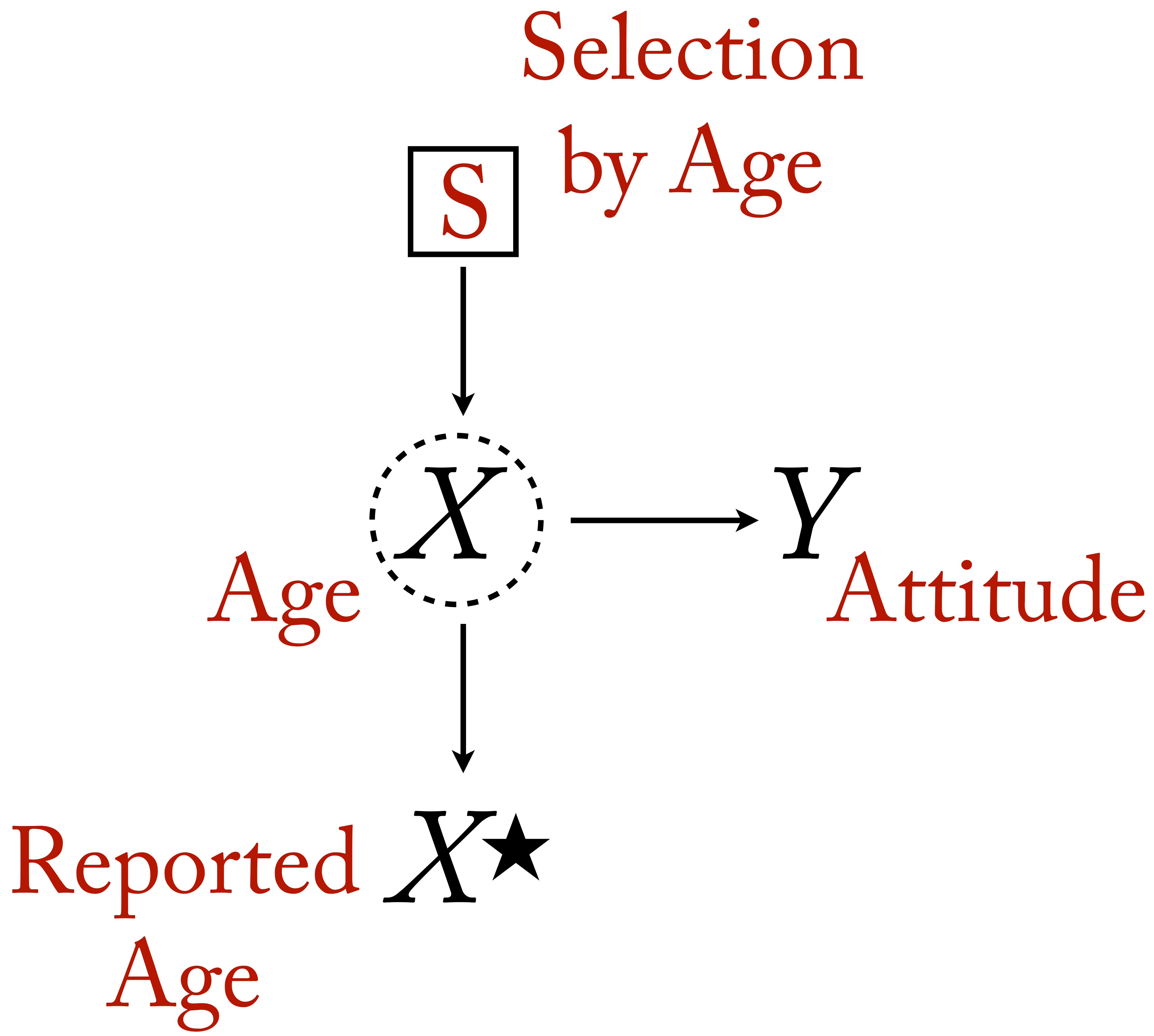
“Young people don’t answer their phones”

$$
\begin{tikzcd}
\boxed{S} \arrow[d] & \\
X \arrow[r] & Y
\end{tikzcd}
$$

“Anarchists don’t answer their phones”

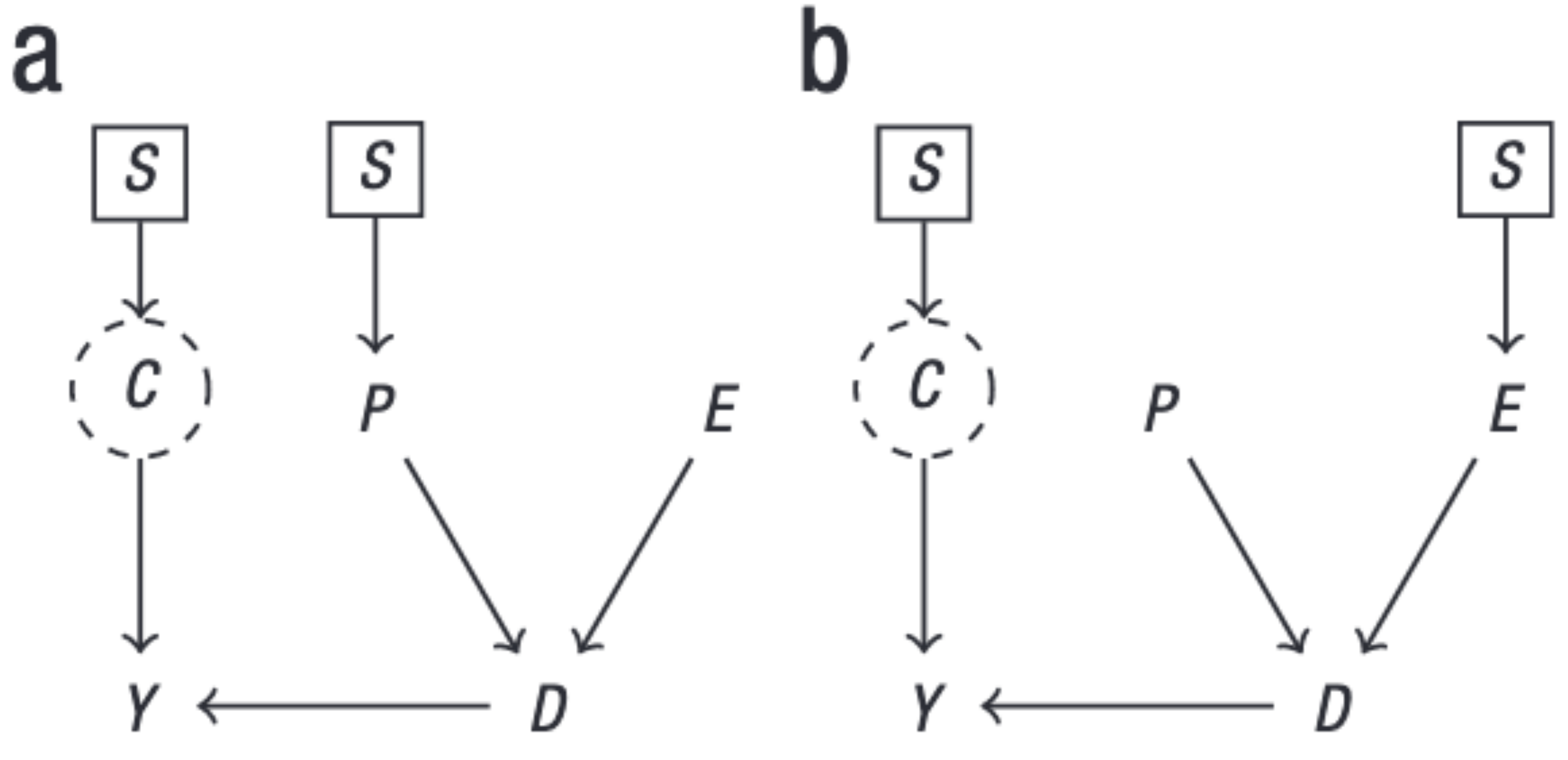
$$
\begin{tikzcd}
& \boxed{S} \arrow[d] \\
X \arrow[r] & Y
\end{tikzcd}
$$

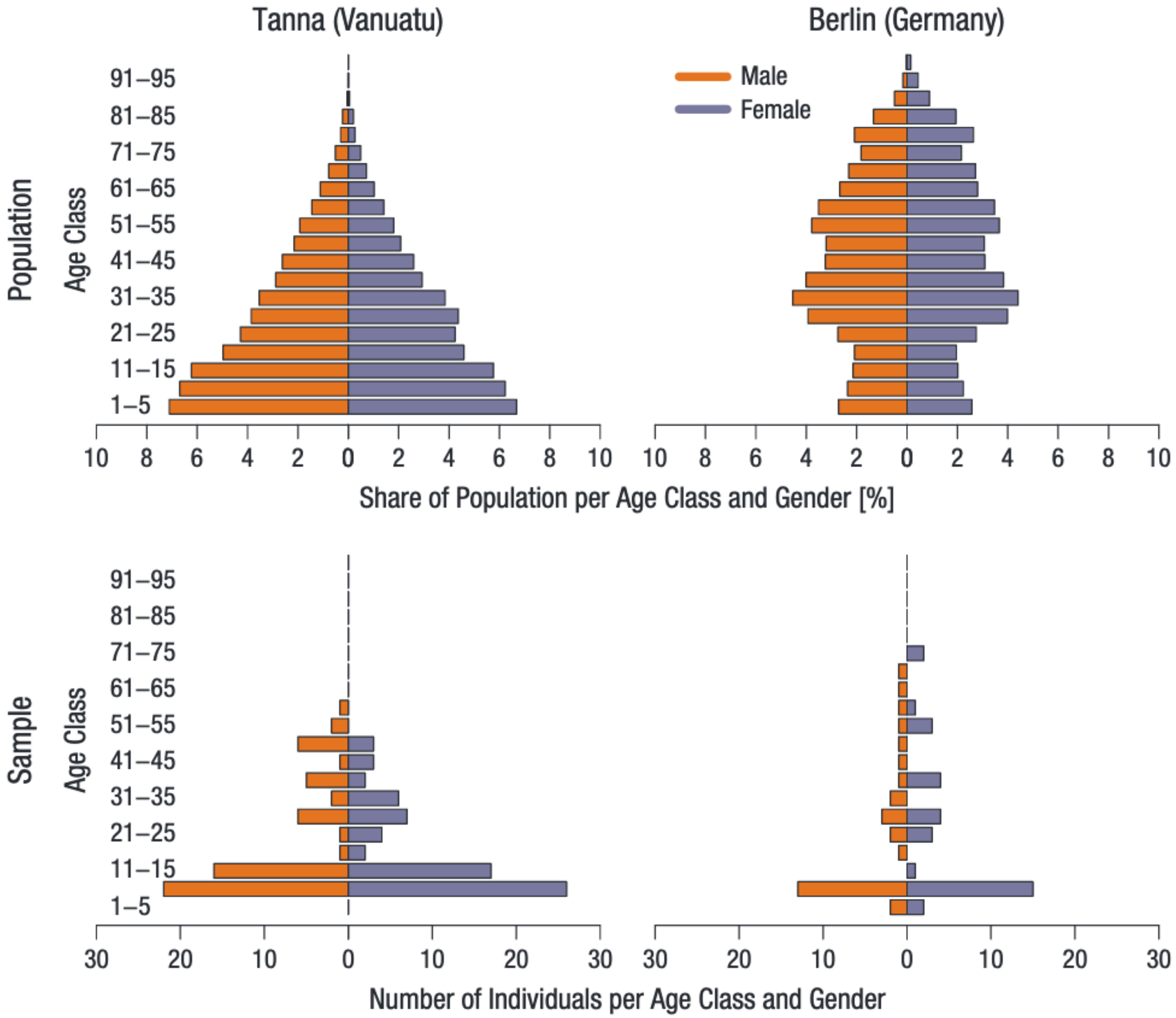
“Young people don’t answer their phones and misreport their age”



A Causal Framework for Cross-Cultural Generalizability

Dominik Deffner[1,2,3], Julia M. Rohrer[4], and Richard McElreath[1]





https://psyarxiv.com/fqukp

Many Qs are really post-strat Qs

Justified descriptions require causal information and post-stratification

Causal effects also, e.g. vaccines

Time trends should account for changes in measurement/population

Comparison is post-stratification from one population to another

Honest Methods for Modest Questions

Surveys Almanacs Collections A
olktales Surveys
raphy Satellites Archives Alma
ecords Scrapi
lections Ethnography Excavations

Simple 4-step plan for honest digital scholarship

(1) What are we trying to describe?

(2) What is the ideal data for doing so?

(3) What data do we actually have?

(4) What causes the differences between (2) and (3)?

(5) [*optional*] Is there a way to use (3) + (4) to do (1)?