Statistical Rethinking
Winter 2019

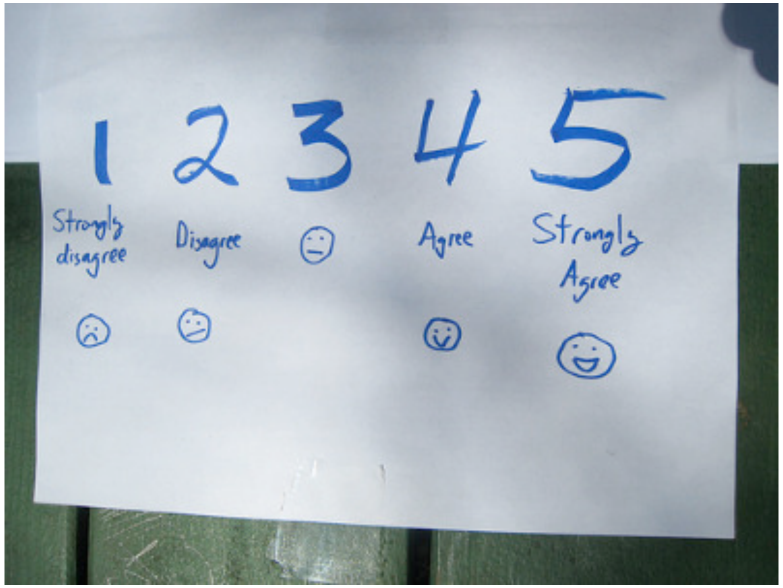
Lecture 14 / Week 7

Ordered Categories, Both Left & Right

Ordered categories

- How much do you like this class? (1–7)
- How important is income of a potential spouse? (1–10)
- How often do you see bats? (never, sometimes, frequently)
- Depth harbor seals dive? (shallow, middle, deep)

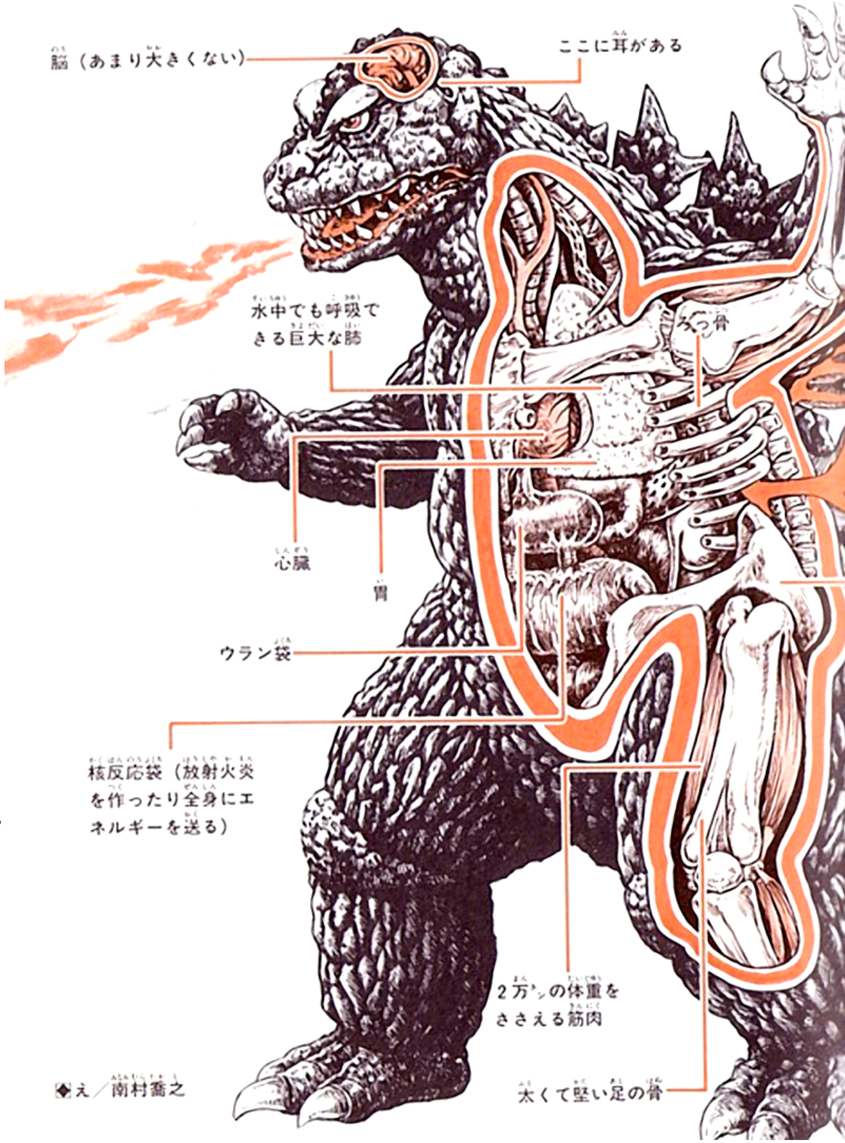
Ordered categories



- Discrete outcomes
- Defined minimum and maximum
- Defined order
- “Distances” between categories unknown

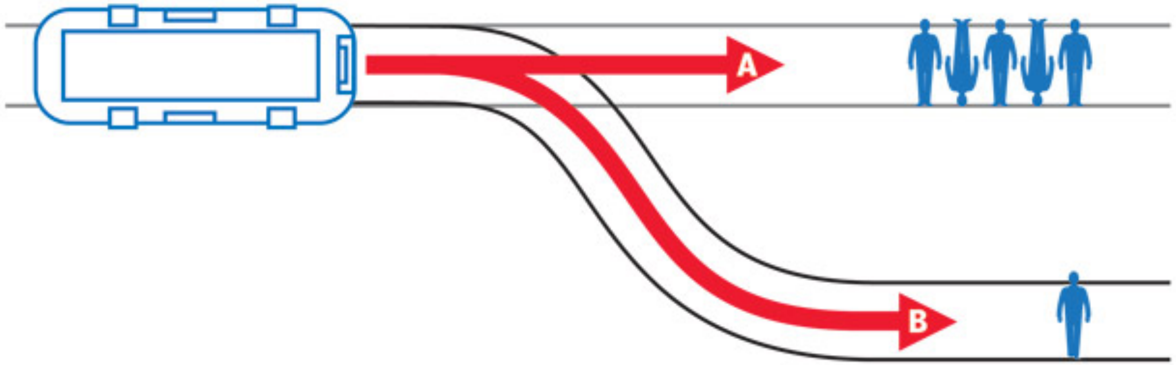
Ordered categories

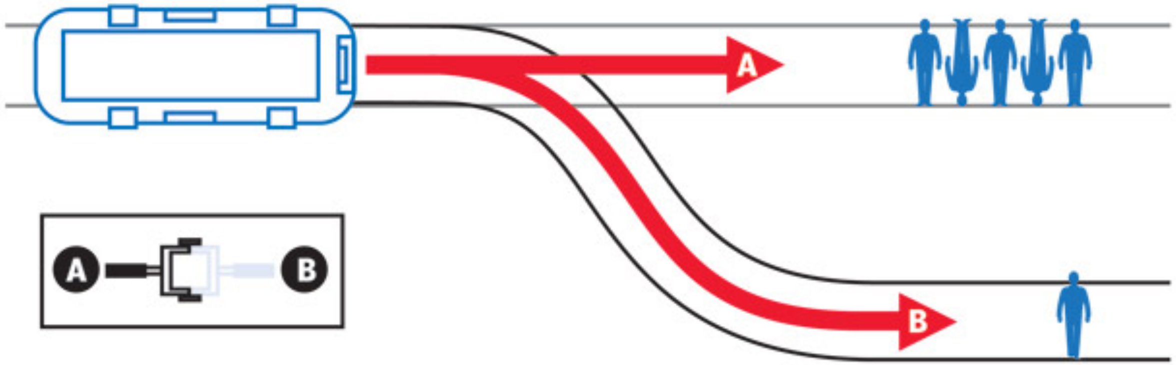
- Hard to model
  - Not continuous
  - Bounded (ceiling & floor)
  - Not counts
- Common solution:
  ordered (aka *ordinal*) logistic
  regression
- Good example of making a monster



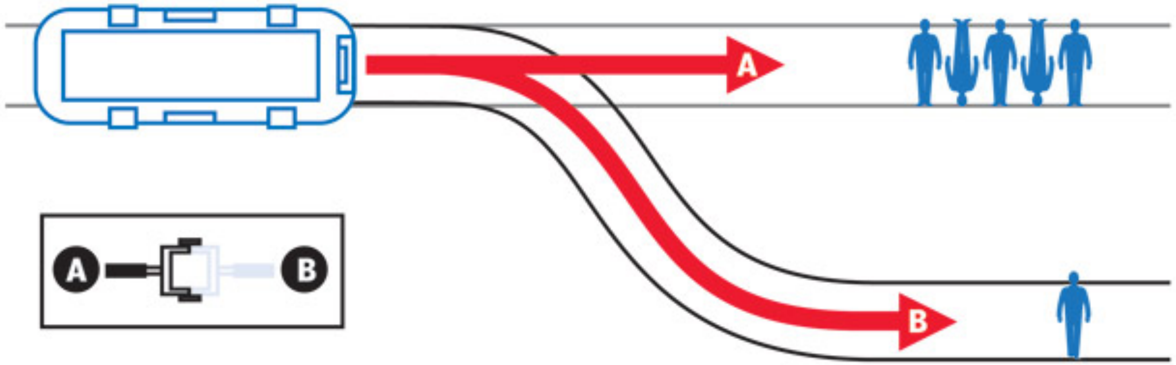

A


A

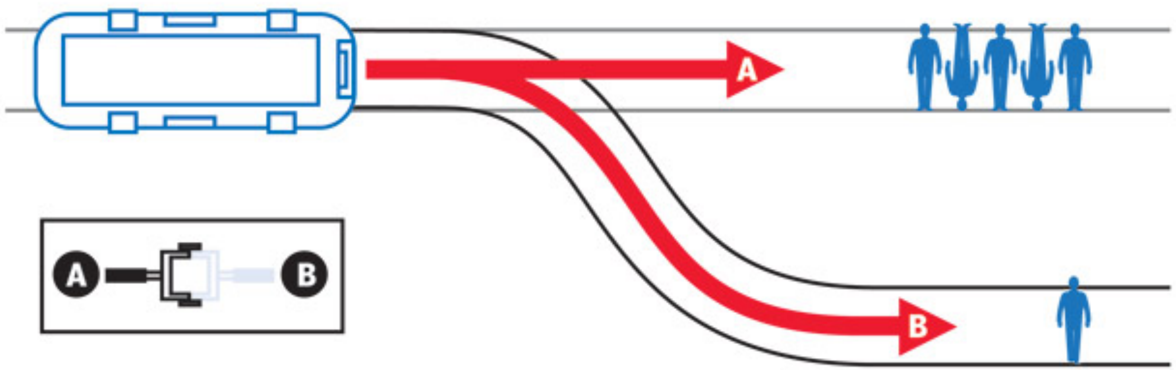

A
B



A
B
A
B

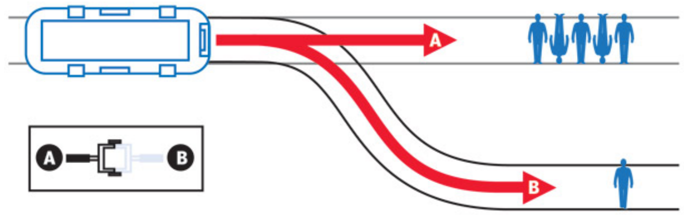


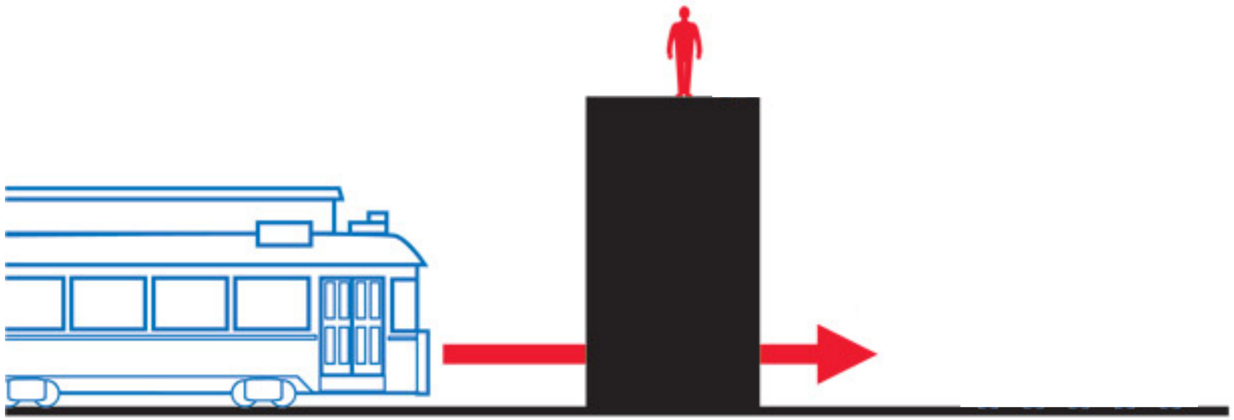
How morally permissible is it
to pull the lever?

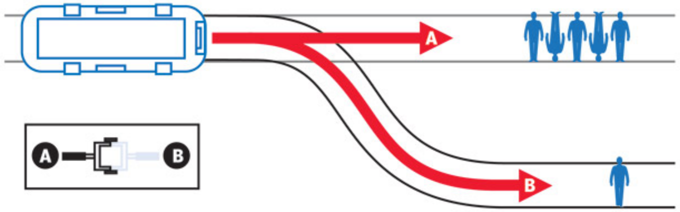


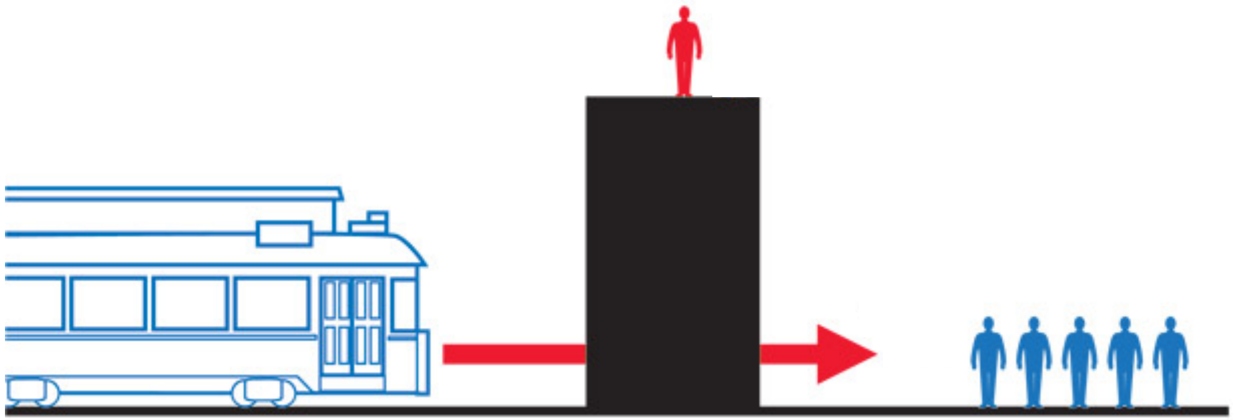
How morally permissible is it
to pull the lever?

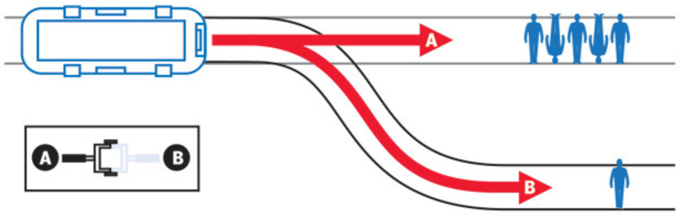
never 1 2 3 4 5 6 7 always

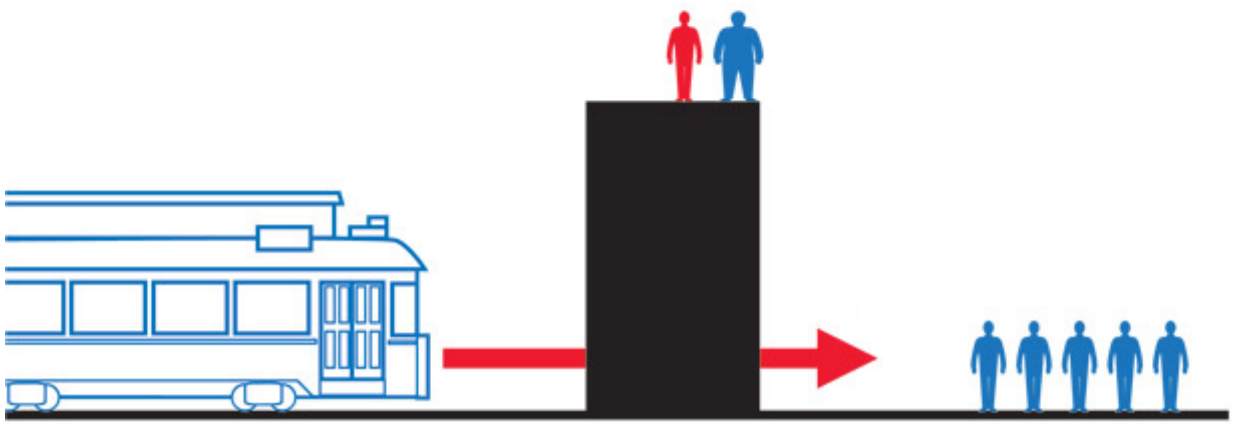

A
B
A
B

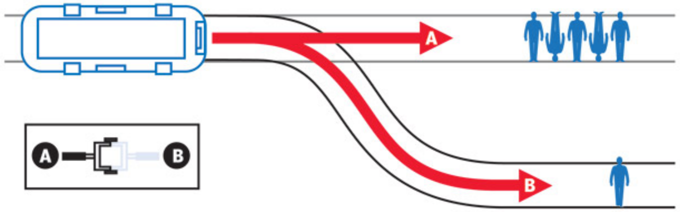


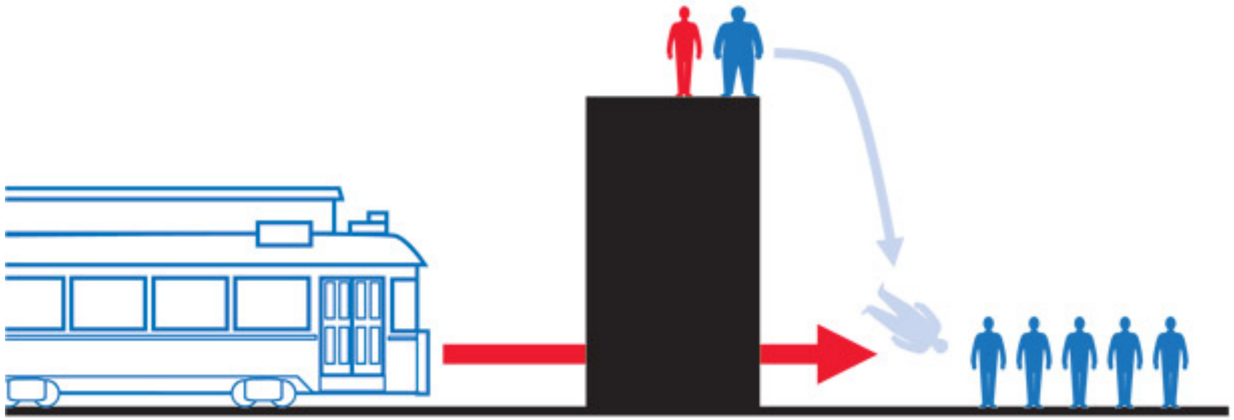

A
A
B
B

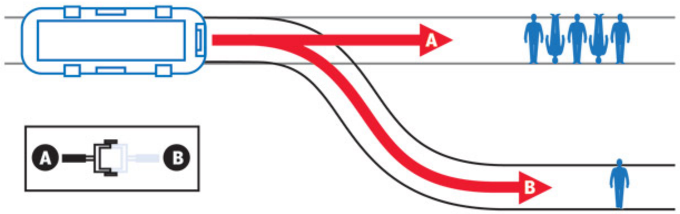


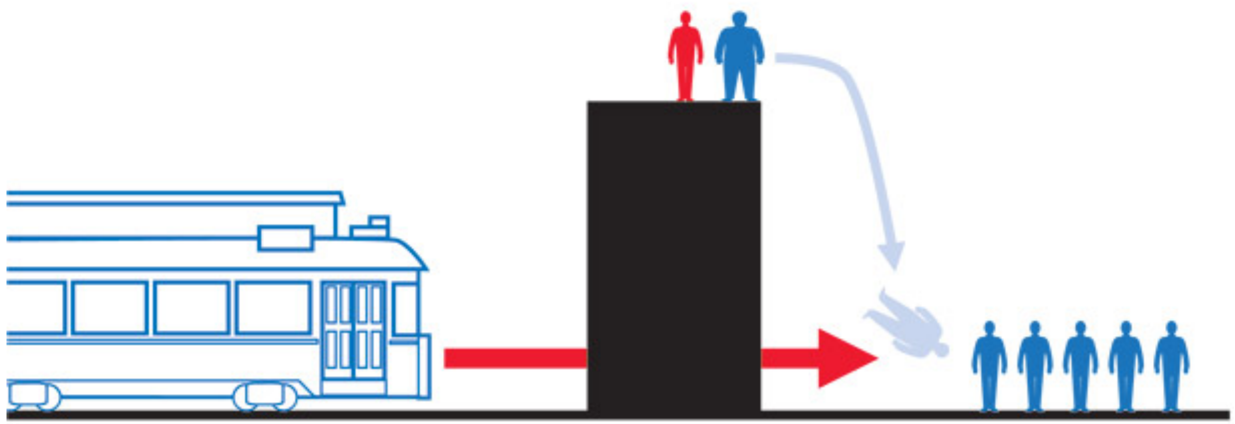

A
B
A
B




A
B
A
B

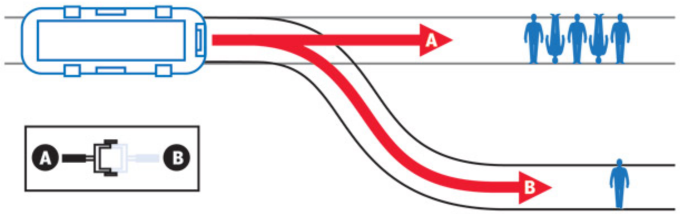


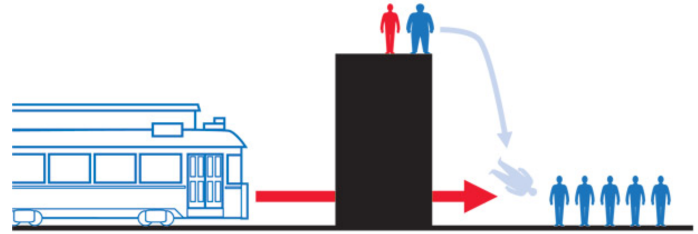


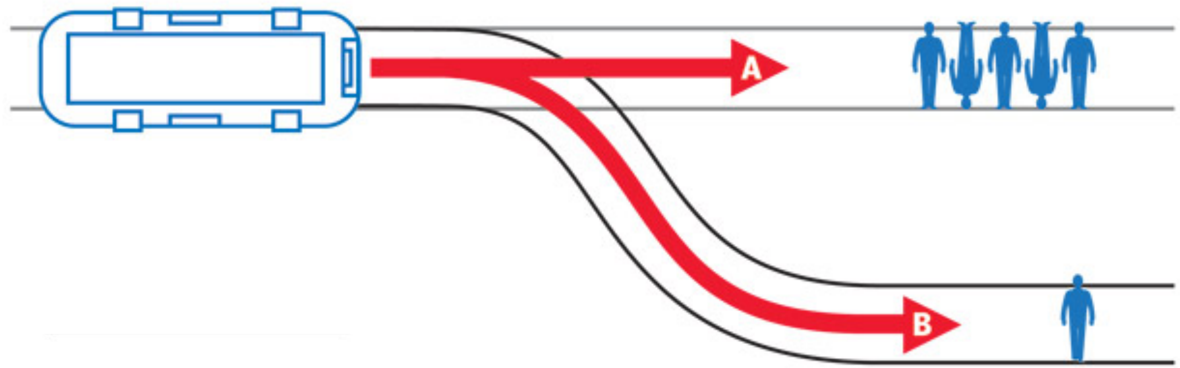


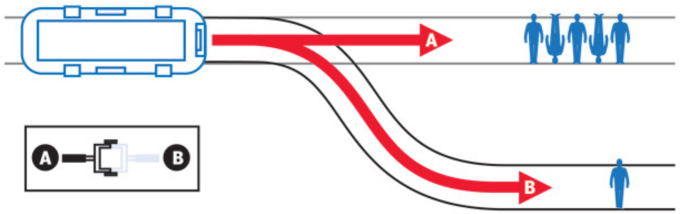
How morally permissible is it
to push the man?

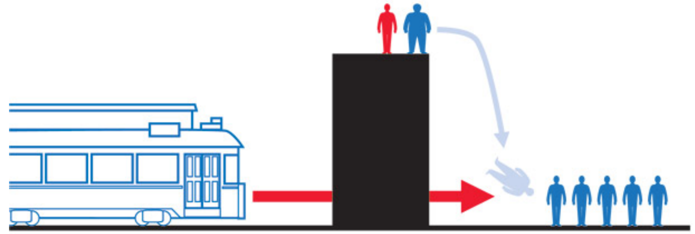
never 1 2 3 4 5 6 7 always

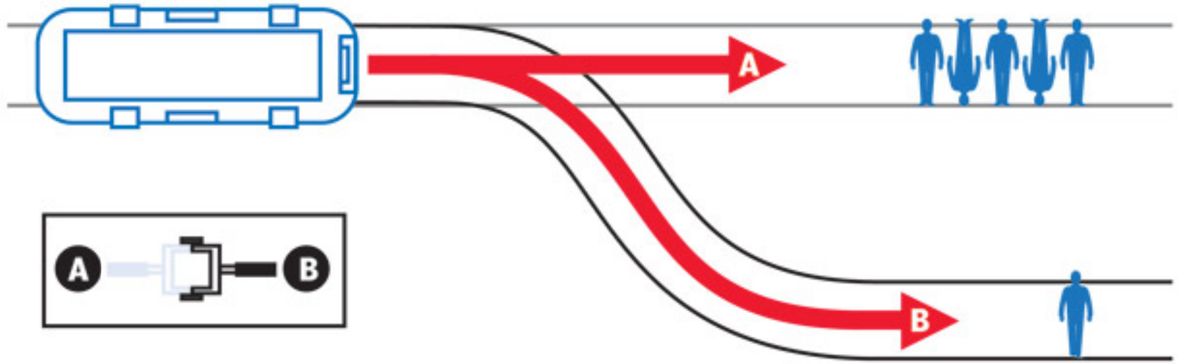

A
B
A
B

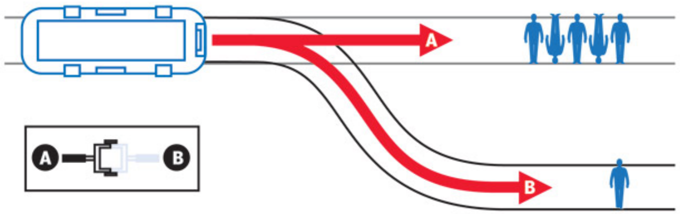


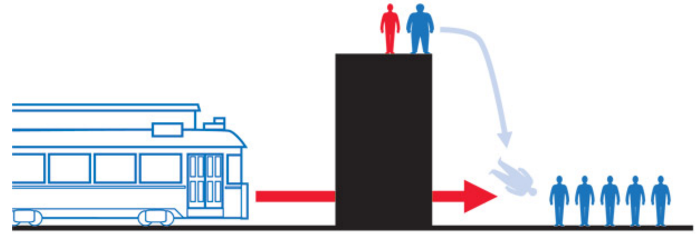

A
B

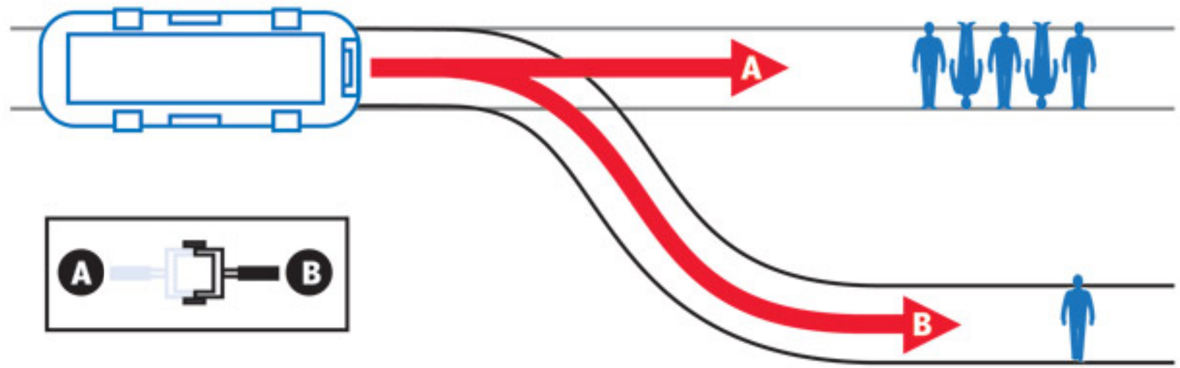

A
B




A
B







How morally permissible is it
to not pull the lever?

never 1 2 3 4 5 6 7 always

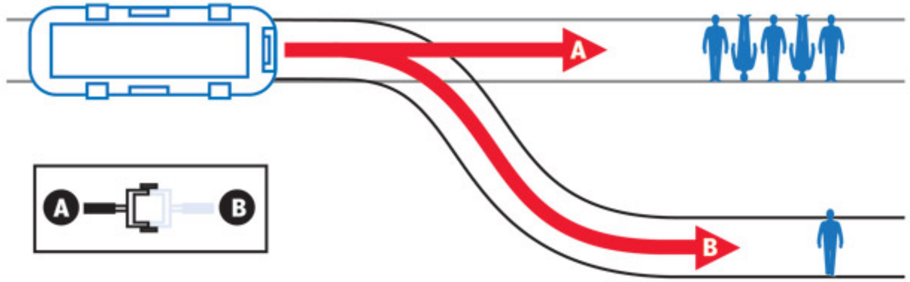
Three principles

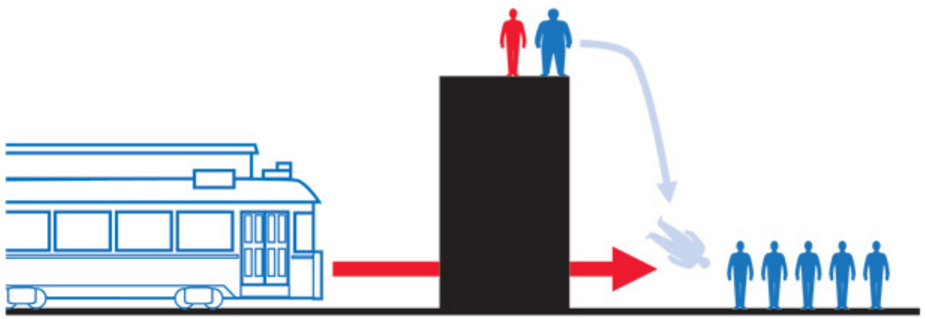
- ***Action***: Harm caused by action is morally worse than same harm caused by inaction.

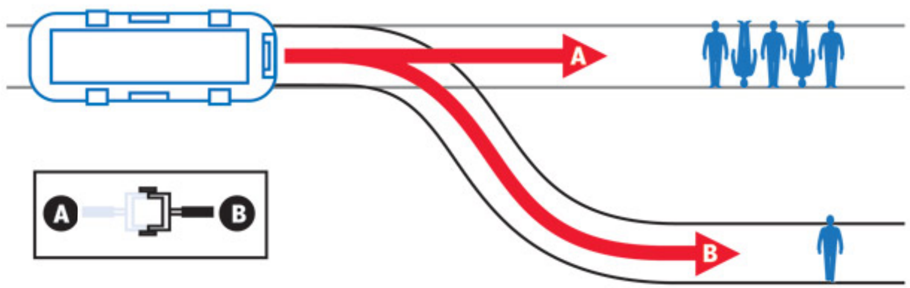
- ***Intention***: Harm intended as means to goal worse than same harm foreseen as a side effect of goal.

- ***Contact***: Harm caused by physical contact worse than same harm without physical contact.

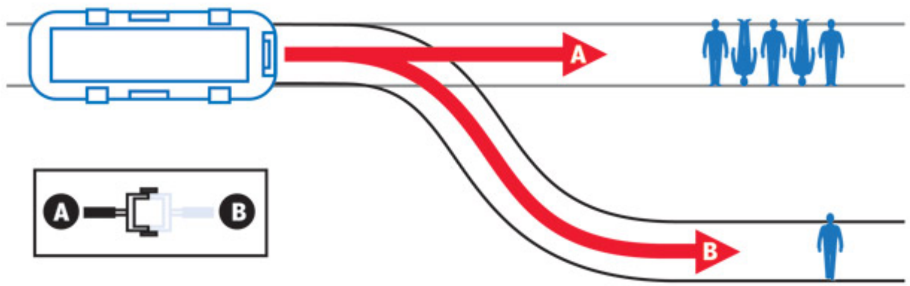
action intention contact



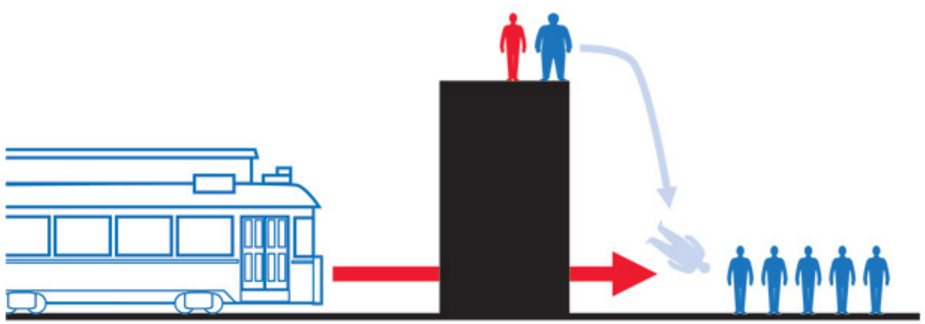


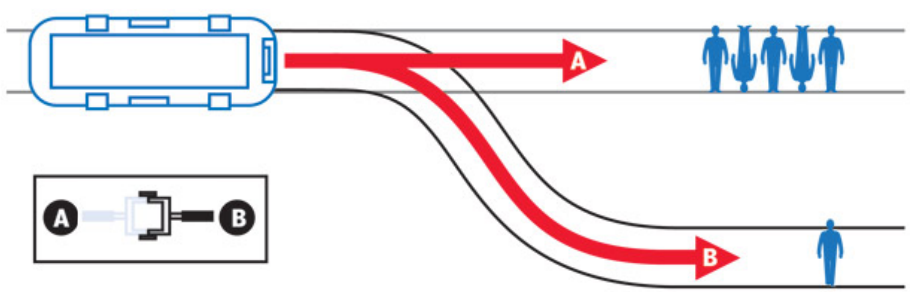


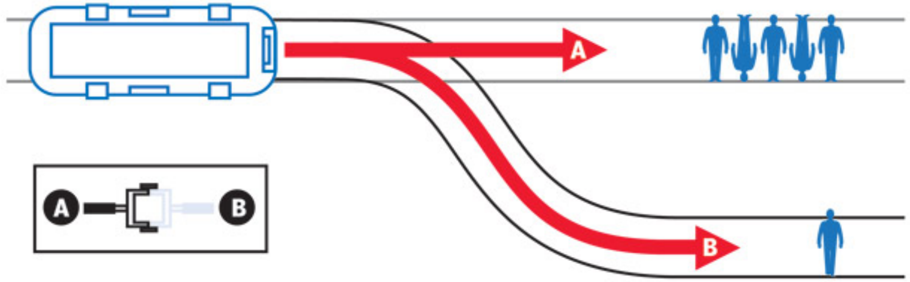
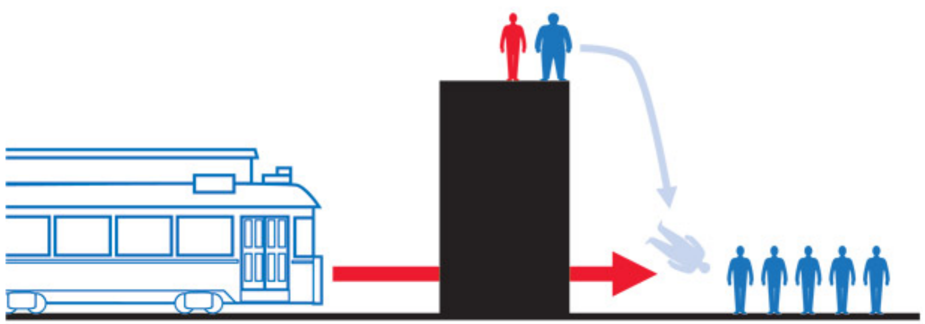
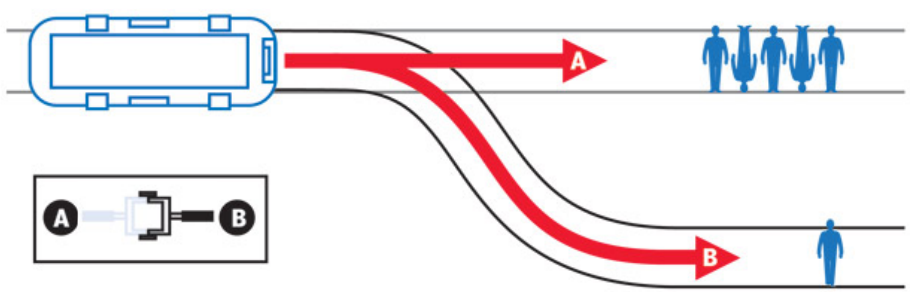
action intention contact

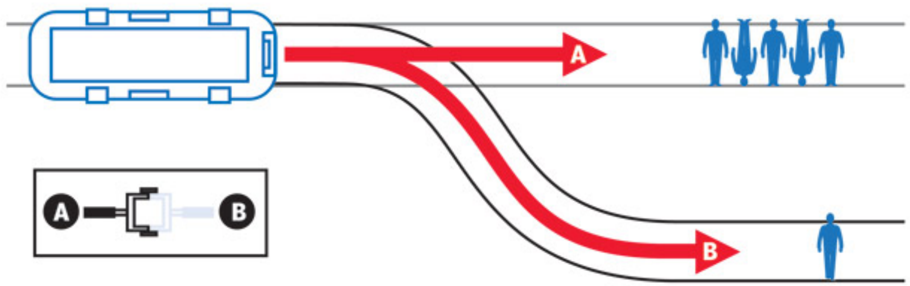
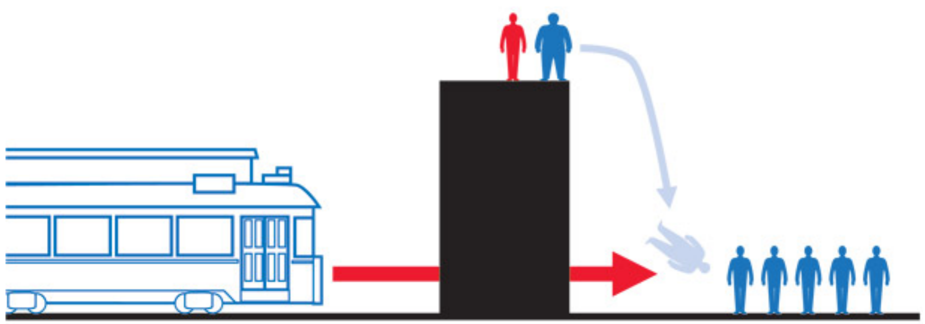
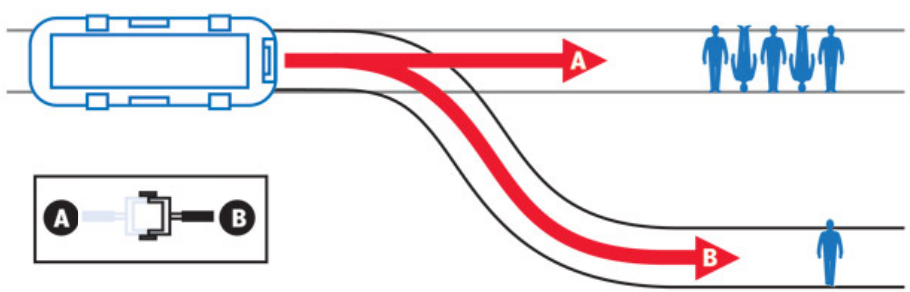


● ○ ○



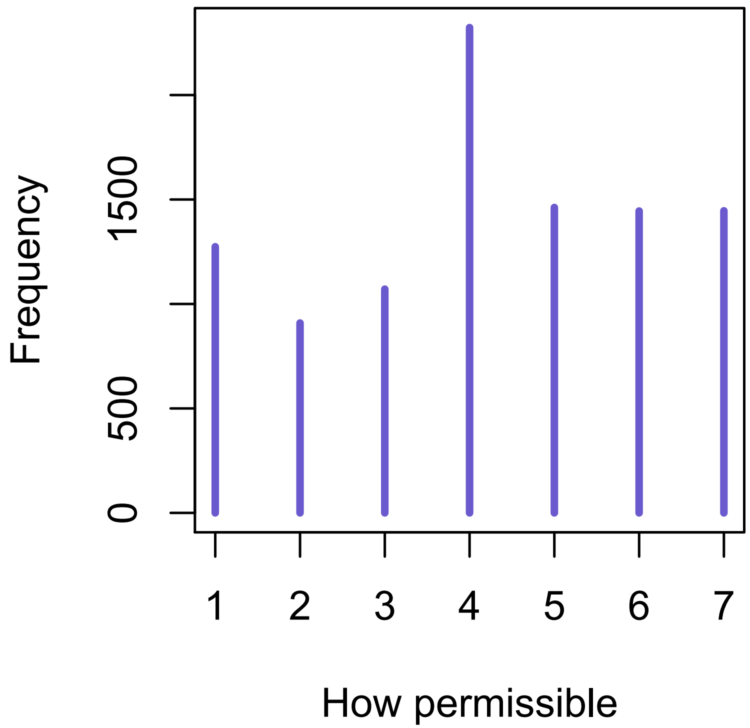


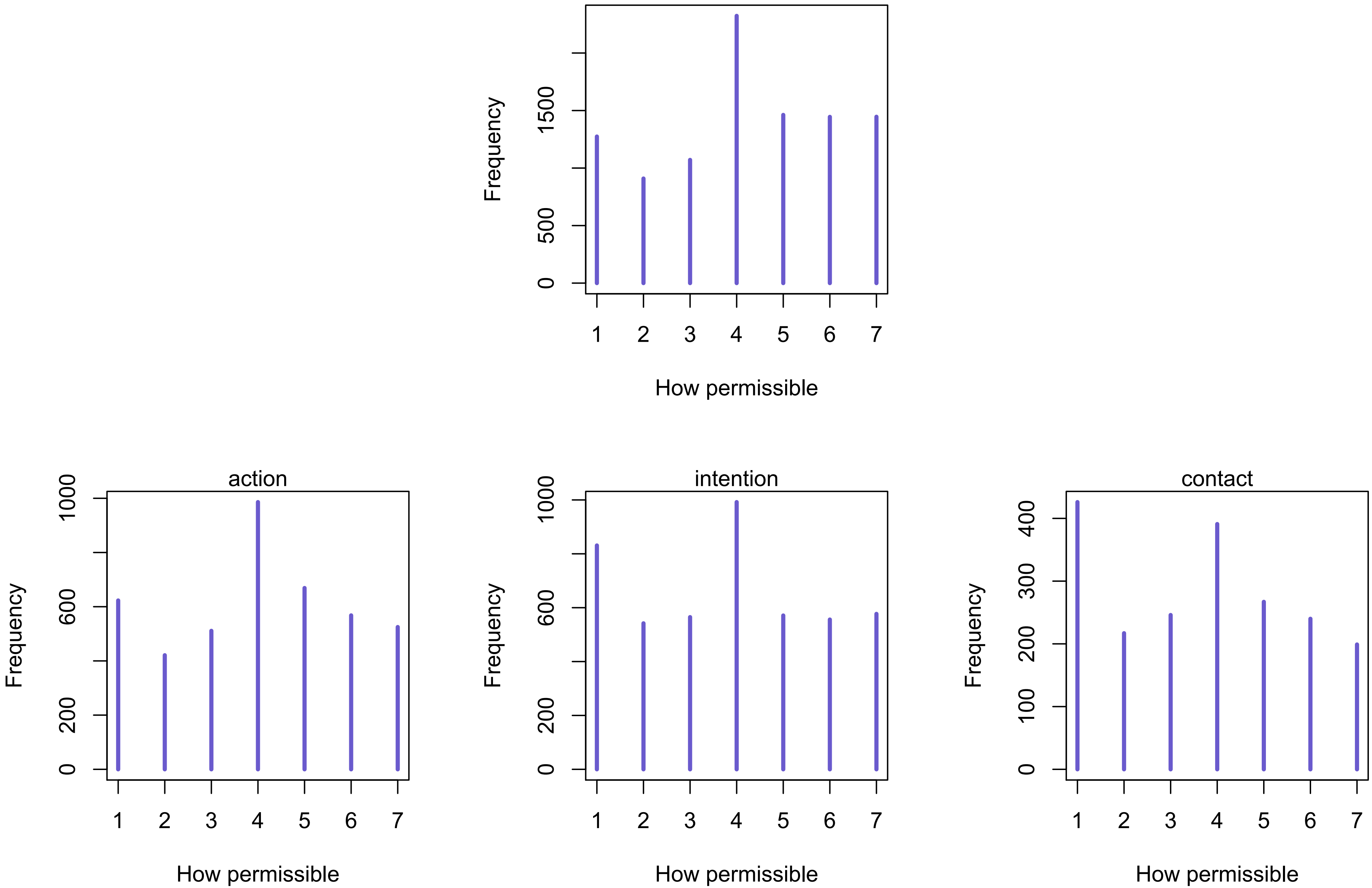
	action	intention	contact
	●	○	○
	●	●	●
			

	action	intention	contact
	●	○	○
	●	●	●
	○	○	○

Moral intuitions

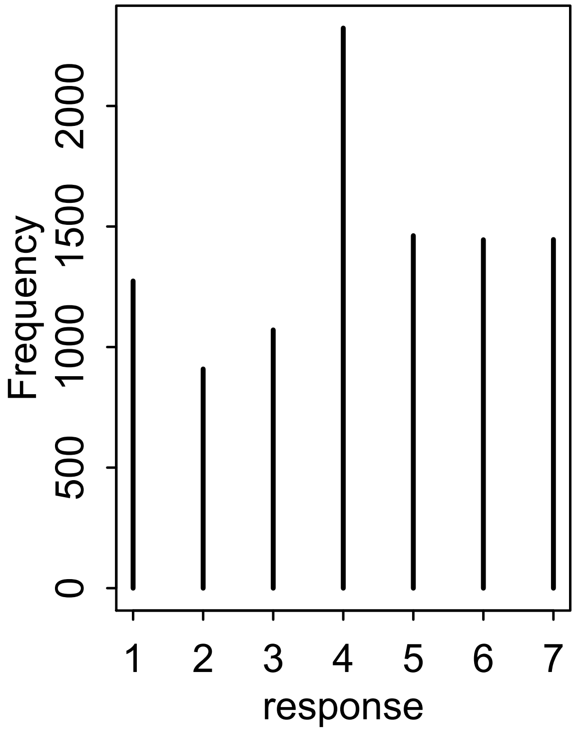
- data(Trolley)
- 331 individuals, 30 scenarios, 9930 responses
- How do responses vary with action, intention, contact?
- Age, gender, individual?




Frequency
0
500
1500
1
2
3
4
5
6
7
How permissible
action
intention
contact
Frequency
0
200
600
1000
How permissible
0
100
200
300
400

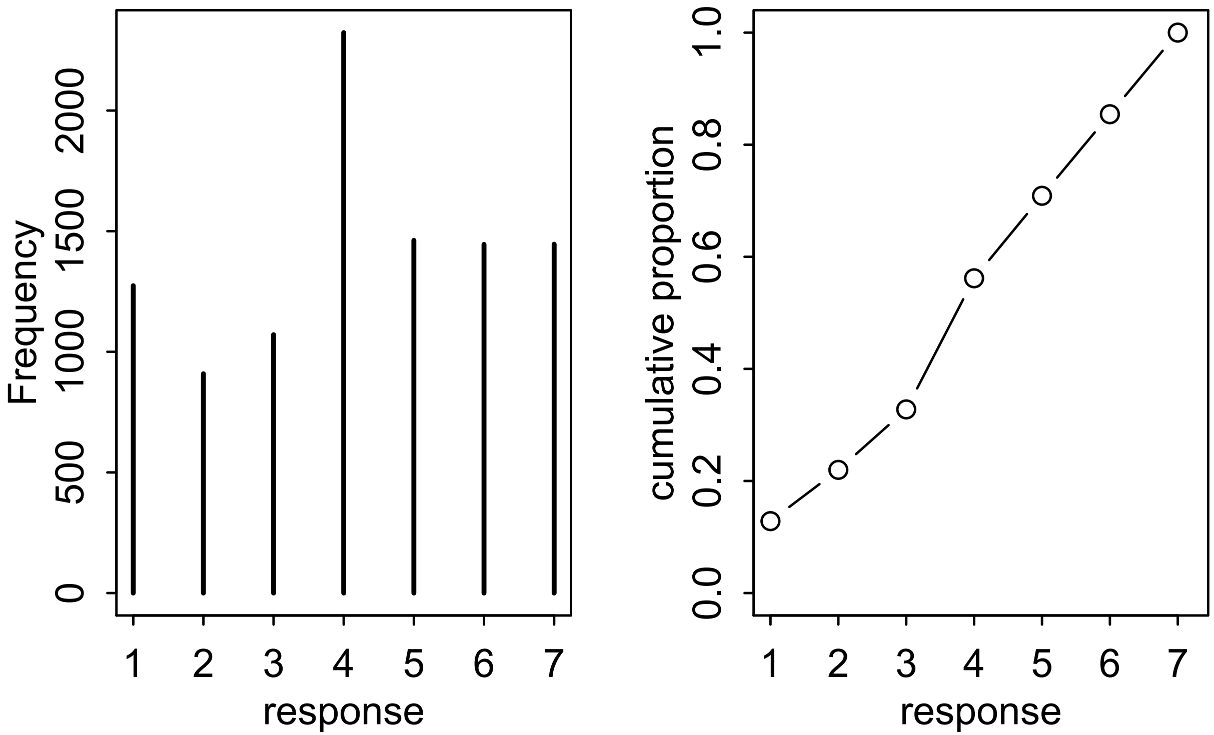
Ordered logit

- A log-cumulative-odds link probability model



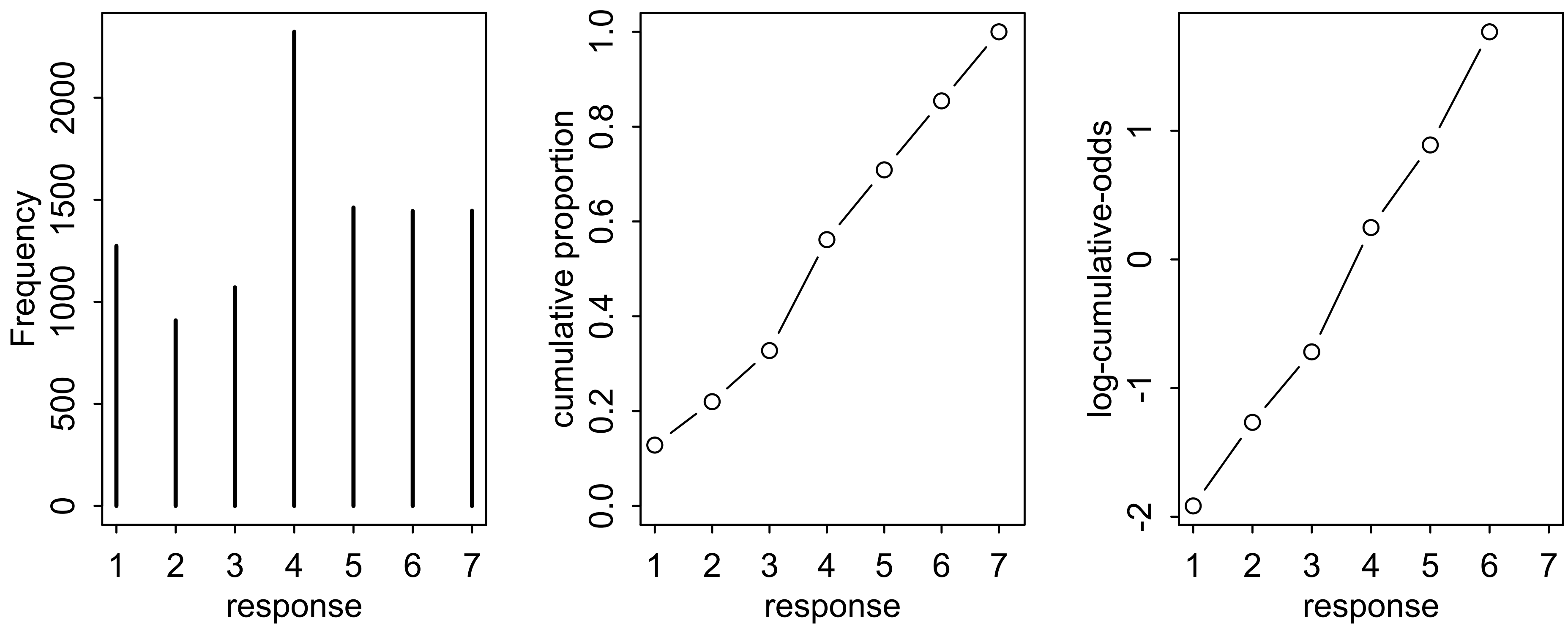
Ordered logit

- A log-cumulative-odds link probability model



Ordered logit

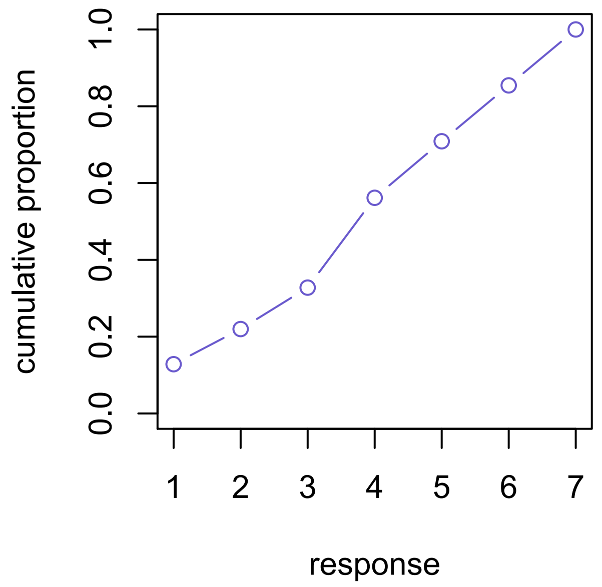
- A log-cumulative-odds link probability model



Ordered logit

- A log-cumulative-odds link probability model

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$

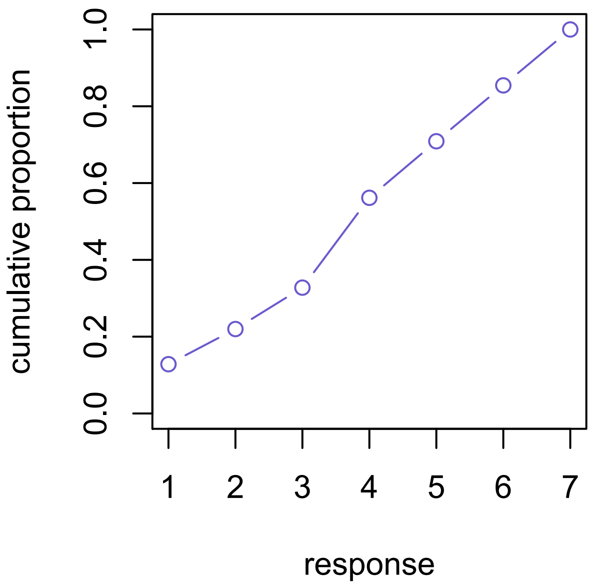


Ordered logit

- A log-cumulative-odds link probability model

cumulative log-odds

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$



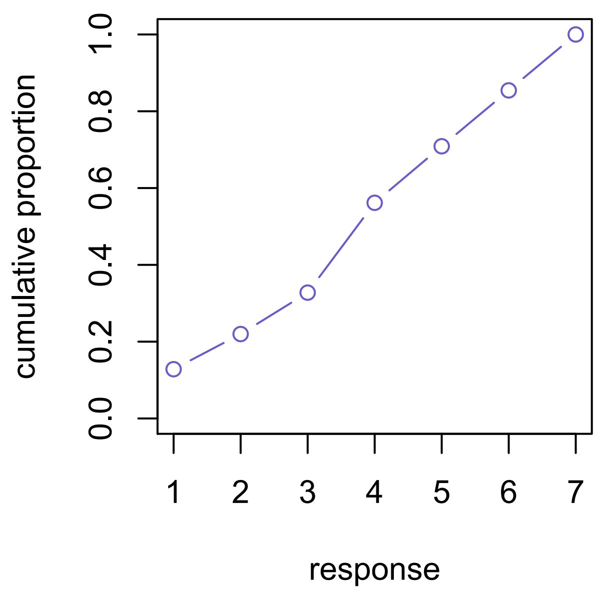
Ordered logit

- A log-cumulative-odds link probability model

response

cumulative log-odds

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$



Ordered logit

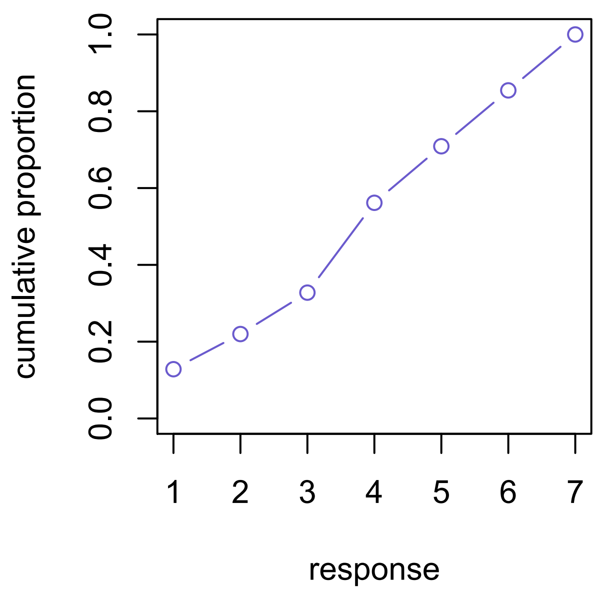
- A log-cumulative-odds link probability model

category

response

cumulative log-odds

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$



Ordered logit

- A log-cumulative-odds link probability model

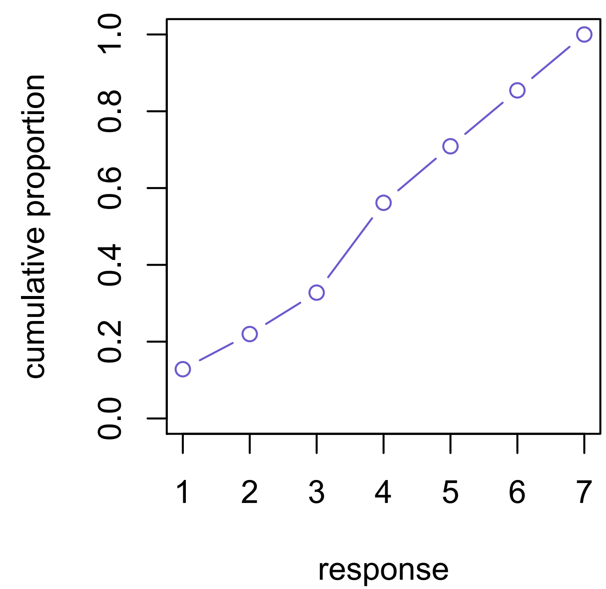
category

response

cumulative log-odds

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$

linear model

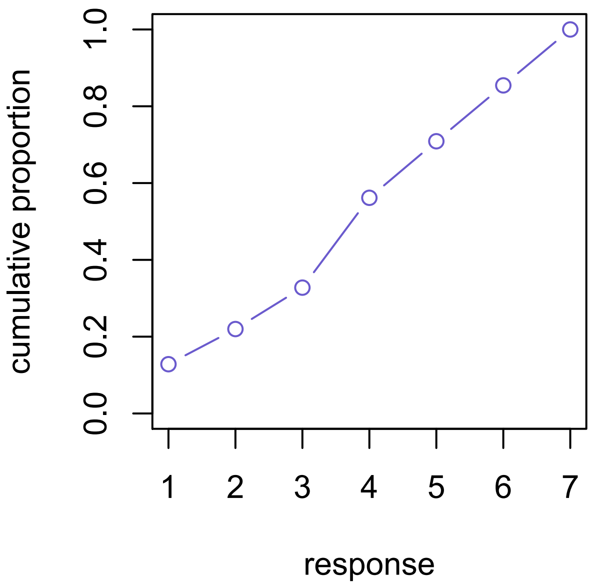


Ordered logit

- A log-cumulative-odds link probability model

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$

$$\Pr(y_i \leq k) = \frac{\exp(\phi_k)}{1 + \exp(\phi_k)}$$

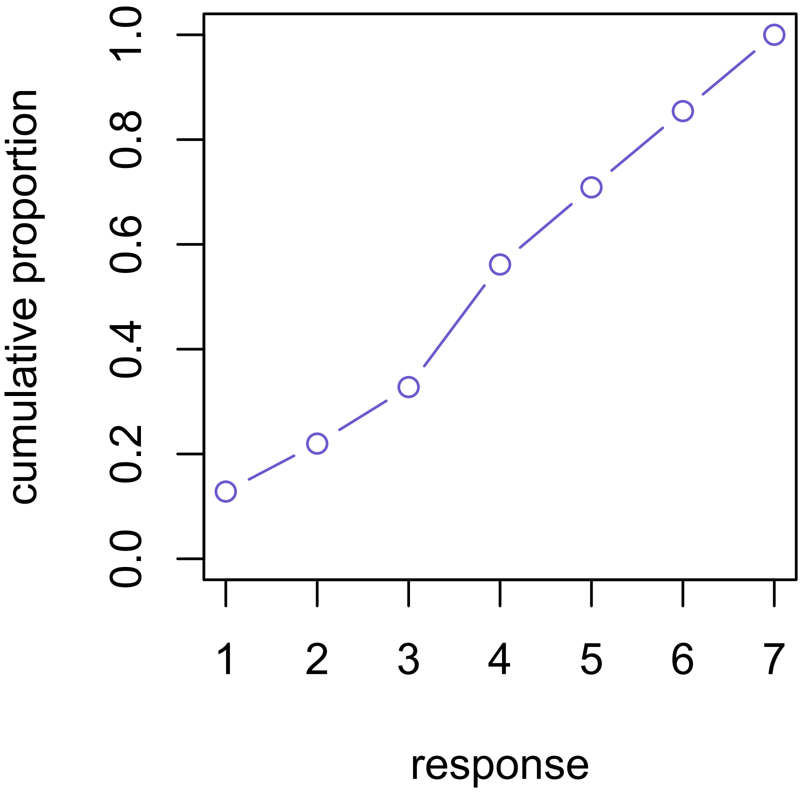


Ordered logit

- A log-cumulative-odds link probability model

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$

$$\Pr(y_i \leq k) = \frac{\exp(\phi_k)}{1 + \exp(\phi_k)}$$

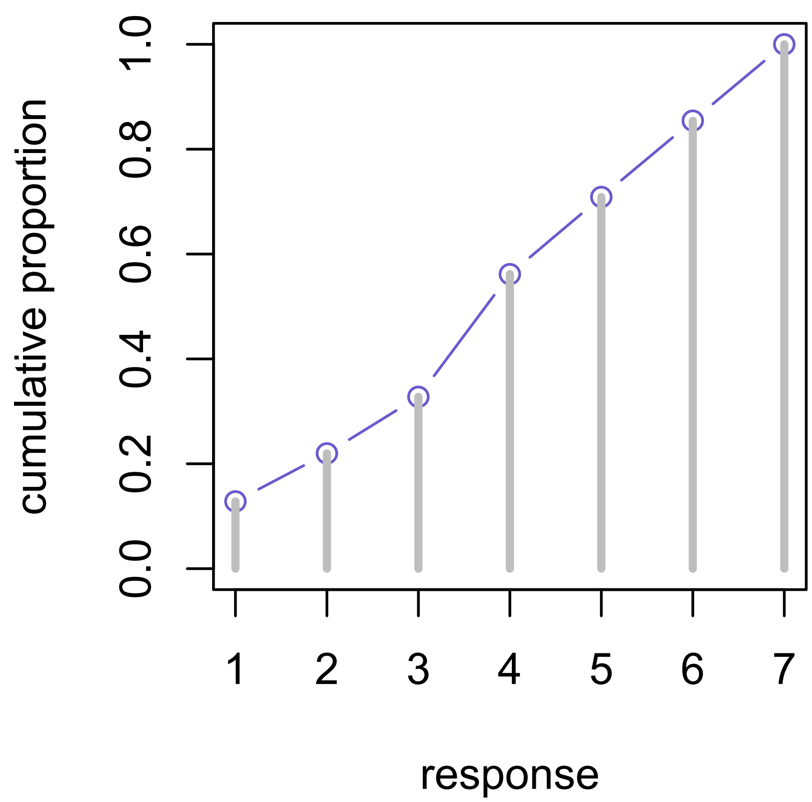


Ordered logit

- A log-cumulative-odds link probability model

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$

$$\boxed{\Pr(y_i \leq k)} = \frac{\exp(\phi_k)}{1 + \exp(\phi_k)}$$



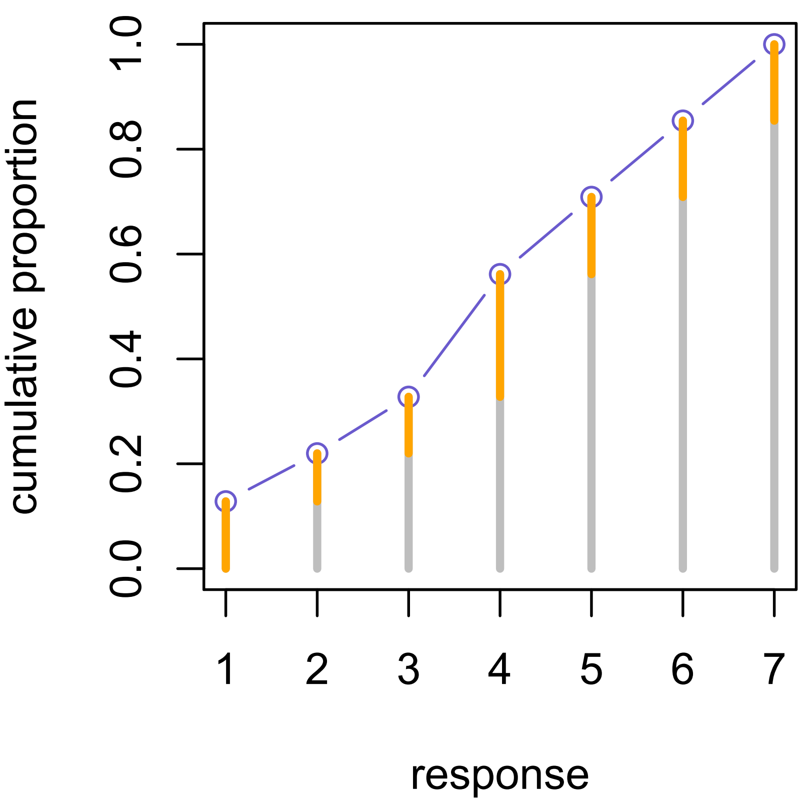
Ordered logit

- A log-cumulative-odds link probability model

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \phi_k$$

$$\boxed{\Pr(y_i \leq k)} = \frac{\exp(\phi_k)}{1 + \exp(\phi_k)}$$

$$\boxed{\Pr(y_i = k)} = \Pr(y_i \leq k) - \Pr(y_i \leq k - 1)$$



Ordered logit

- Cutpoints: vector of intercepts
- Linear model influences every category

$R_i \sim \text{Categorical}(\mathbf{p})$	[probability of data]
$p_1 = q_1$	[probabilities of each value k]
$p_k = q_k - q_{k-1} \quad \text{for } K > k > 1$	
$p_K = 1 - q_{k-1}$	
$\text{logit}(q_k) = \kappa_k - \phi_i$	[cumulative logit link]
$\phi_i = \text{terms of linear model}$	[linear model]
$\kappa_k \sim \text{Normal}(0, 1.5)$	[common prior for each intercept]

Ordered logit

```
m12.5 <- ulam(
    alist(
        R ~ dordlogit( 0 , cutpoints ),
        cutpoints ~ dnorm( 0 , 1.5 )
    ) ,
    data=list( R=d$response ), chains=4 , cores=3 )
```

$$
\begin{aligned}
R_i &\sim \text{Categorical}(\mathbf{p}) && \text{[probability of data]} \\
p_1 &= q_1 && \text{[probabilities of each value } k\text{]} \\
p_k &= q_k - q_{k-1} \quad \text{for } K > k > 1 \\
p_K &= 1 - q_{k-1} \\
\text{logit}(q_k) &= \kappa_k - \phi_i && \text{[cumulative logit link]} \\
\phi_i &= \text{terms of linear model} && \text{[linear model]} \\
\kappa_k &\sim \text{Normal}(0, 1.5) && \text{[common prior for each intercept]}
\end{aligned}
$$

Ordered logit

```
m12.5 <- ulam(
    alist(
        R ~ dordlogit( 0 , cutpoints ),
        cutpoints ~ dnorm( 0 , 1.5 )
    ) ,
    data=list( R=d$response ), chains=4 , cores=3 )
```

```
precis( m12.5 , depth=2 )
```

```
               mean   sd  5.5% 94.5% n_eff Rhat
cutpoints[1] -1.92 0.03 -1.96 -1.87  1460    1
cutpoints[2] -1.27 0.02 -1.31 -1.23  2091    1
cutpoints[3] -0.72 0.02 -0.75 -0.68  2480    1
cutpoints[4]  0.25 0.02  0.22  0.28  2701    1
cutpoints[5]  0.89 0.02  0.85  0.92  2373    1
cutpoints[6]  1.77 0.03  1.72  1.81  2345    1
```

Back to probability scale

```
precis( m12.5 , depth=2 )
```

R code 12.18

```
              mean   sd  5.5% 94.5% n_eff Rhat
cutpoints[1] -1.92 0.03 -1.96 -1.87  1460    1
cutpoints[2] -1.27 0.02 -1.31 -1.23  2091    1
cutpoints[3] -0.72 0.02 -0.75 -0.68  2480    1
cutpoints[4]  0.25 0.02  0.22  0.28  2701    1
cutpoints[5]  0.89 0.02  0.85  0.92  2373    1
cutpoints[6]  1.77 0.03  1.72  1.81  2345    1
```

```
inv_logit(coef(m12.5))
```

R code 12.19

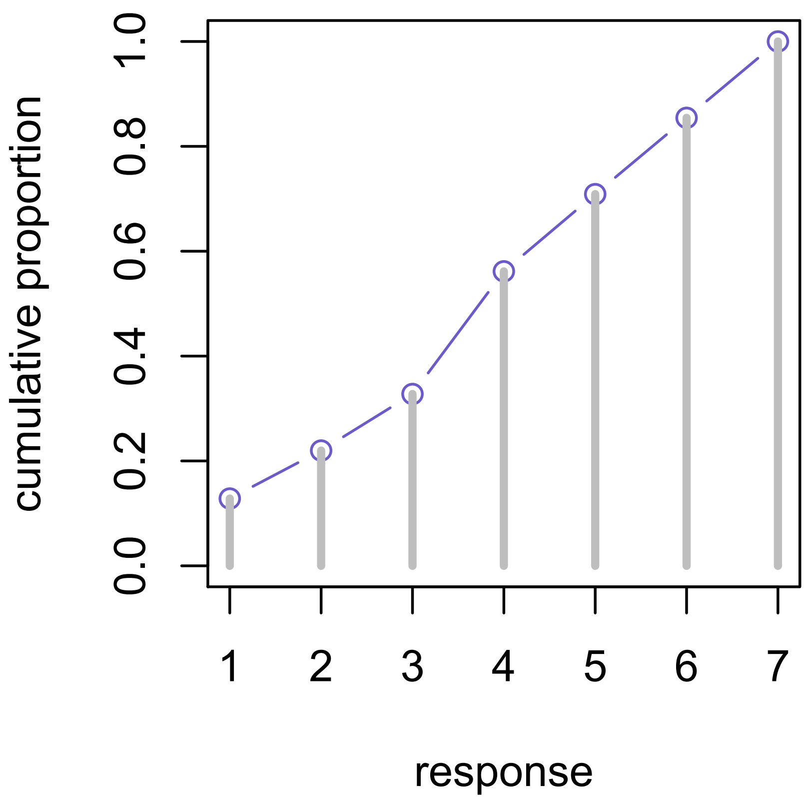
```
cutpoints[1] cutpoints[2] cutpoints[3] cutpoints[4] cutpoints[5] cutpoints[6]
   0.1283325    0.2198127    0.3276819    0.5615751    0.7087310    0.8543559
```

Back to probability scale

```
inv_logit(coef(m12.5))
```

R code 12.19

```
cutpoints[1] cutpoints[2] cutpoints[3] cutpoints[4] cutpoints[5] cutpoints[6]
   0.1283325    0.2198127    0.3276819    0.5615751    0.7087310    0.8543559
```



Adding predictor variables

In general:

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \alpha_k - \phi_i$$
$$\phi_i = \beta x_i$$

Trolley data:

$$\log \frac{\Pr(y_i \leq k)}{1 - \Pr(y_i \leq k)} = \alpha_k - \phi_i$$
$$\phi_i = \beta_A A_i + \beta_C C_i + \mathrm{B}_{I,i} I_i$$
$$\mathrm{B}_{I,i} = \beta_I + \beta_{IA} A_i + \beta_{IC} C_i$$

Adding predictor variables

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

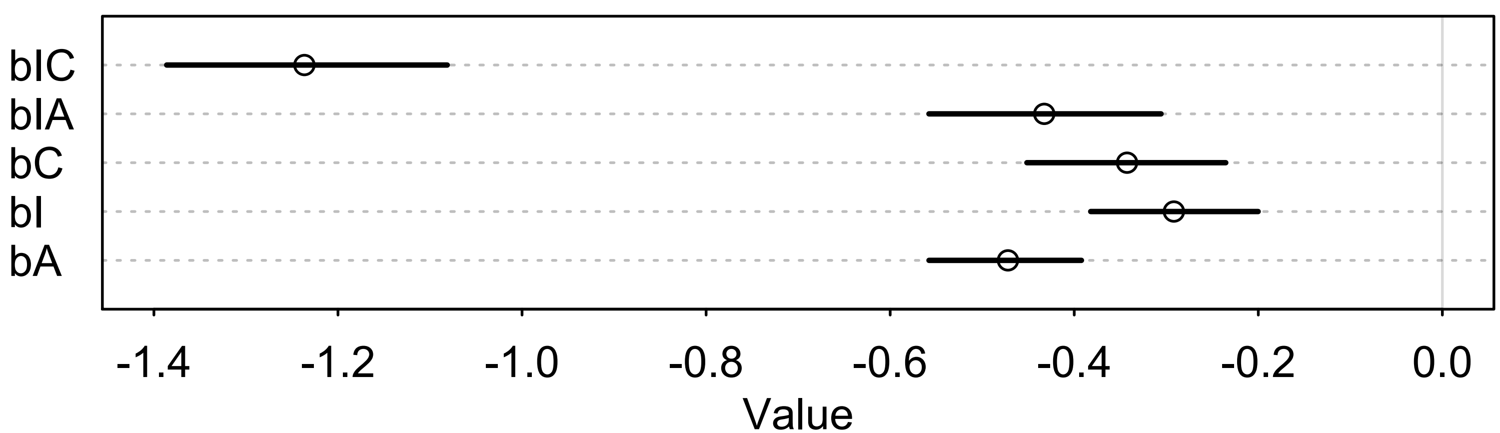
$$\phi_i = \beta_A A_i + \beta_C C_i + \mathrm{B}_{I,i} I_i$$

$$\mathrm{B}_{I,i} = \beta_I + \beta_{IA} A_i + \beta_{IC} C_i$$

```
m12.6 <- ulam(
    alist(
        R ~ dordlogit( phi , cutpoints ),
        phi <- bA*A + bC*C + BI*I ,
        BI <- bI + bIA*A + bIC*C ,
        c(bA,bI,bC,bIA,bIC) ~ dnorm( 0 , 0.5 ),
        cutpoints ~ dnorm( 0 , 1.5 )
    ) , data=dat , chains=4 , cores=4 )
```

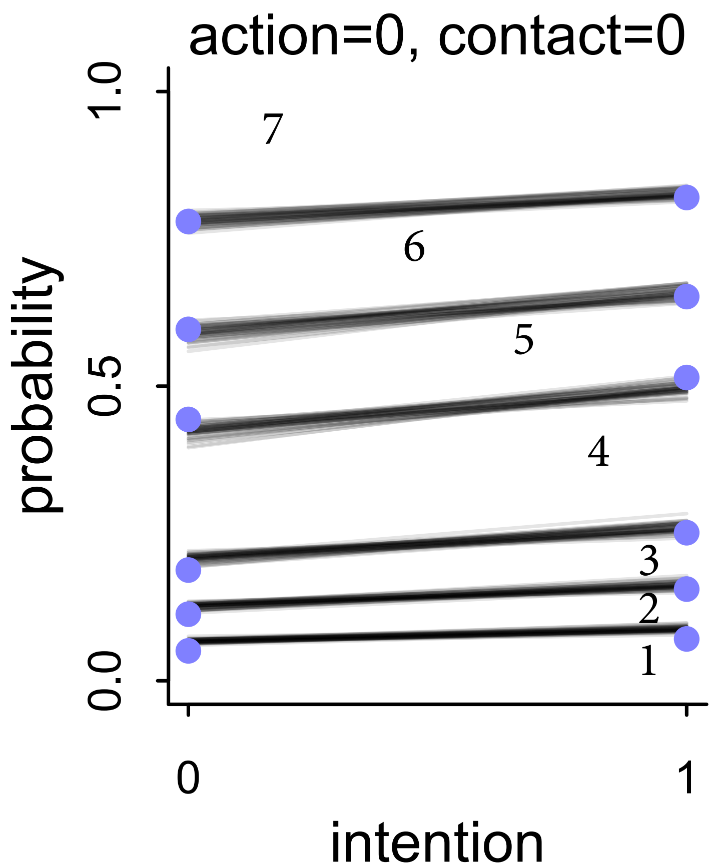
```
m12.6 <- ulam(
    alist(
        R ~ dordlogit( phi , cutpoints ),
        phi <- bA*A + bC*C + BI*I ,
        BI <- bI + bIA*A + bIC*C ,
        c(bA,bI,bC,bIA,bIC) ~ dnorm( 0 , 0.5 ),
        cutpoints ~ dnorm( 0 , 1.5 )
    ) , data=dat , chains=4 , cores=4 )
precis( m12.6 )
```

```
6 vector or matrix parameters omitted in display. Use depth=2 to show them.
     mean   sd  5.5% 94.5% n_eff Rhat
bIC -1.24 0.09 -1.39 -1.08  1038    1
bIA -0.43 0.08 -0.56 -0.31   927    1
bC  -0.34 0.07 -0.45 -0.24  1168    1
bI  -0.29 0.06 -0.38 -0.20   806    1
bA  -0.47 0.05 -0.56 -0.39  1042    1
```



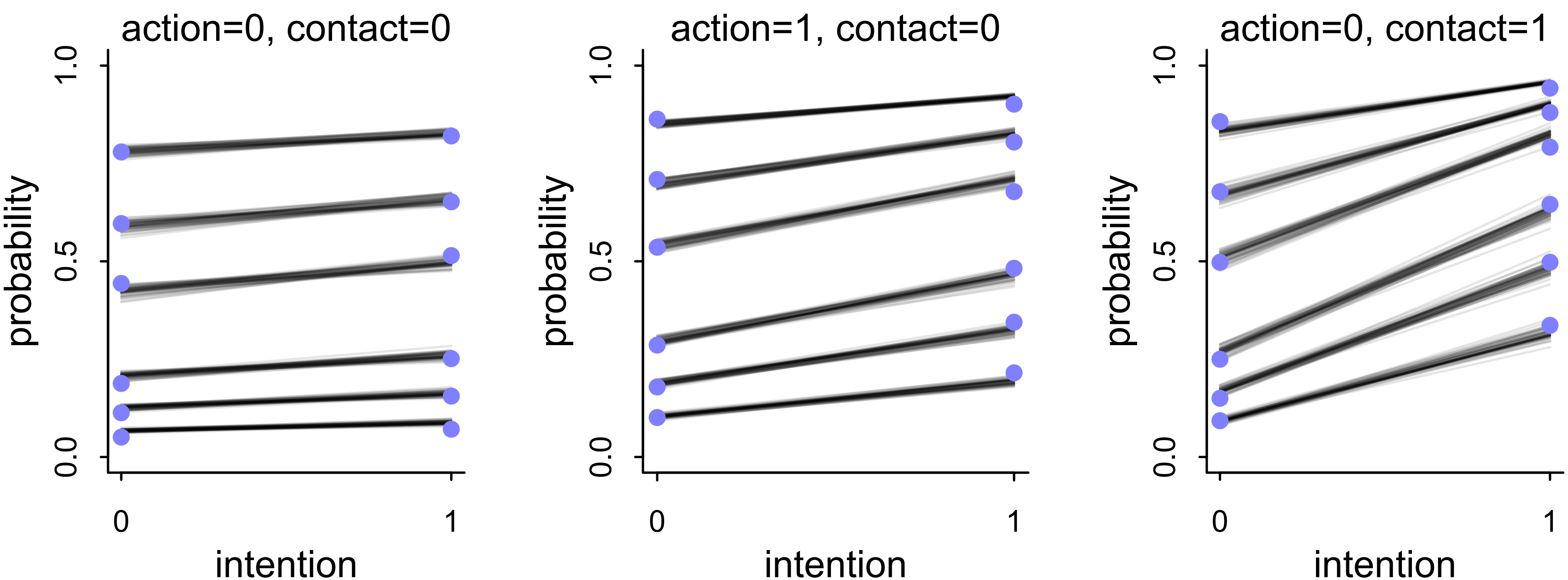
Plotting ordered logits

- Oh, bother: Posterior prediction a *vector* of probabilities, one for each level of outcome
- How to plot this?
- Several useful options



action=0, contact=0
1.0
0.5
0.0
probability
7
6
5
4
3
2
1
0
1
intention

Figure 12.6



action=0, contact=0
action=1, contact=0
action=0, contact=1
probability
1.0
0.5
0.0
0
1
intention

Figure 12.6

Ordered categorical predictors

- Not just outcomes, also predictors
- Ordinary metric predictor assumes each unit change same change (on linear scale)
- Ordered category predictor: Each level could have different size of influence, but all effects in same direction (monotonic)
  - Example: Education

```
library(rethinking)
data(Trolley)
d <- Trolley
levels(d$edu)
```

R code 12.30

```
[1] "Bachelor's Degree"    "Elementary School"    "Graduate Degree"
[4] "High School Graduate" "Master's Degree"      "Middle School"
[7] "Some College"         "Some High School"
```

Ordered categorical predictors

```
library(rethinking)
data(Trolley)
d <- Trolley
levels(d$edu)
```

R code 12.30

```
[1] "Bachelor's Degree"    "Elementary School"    "Graduate Degree"
[4] "High School Graduate" "Master's Degree"      "Middle School"
[7] "Some College"         "Some High School"
```

- Strategy
  - Each level gets a unique parameter
  - Parameters sum to total maximum effect of education
  - Assign prior to total effect and to *proportions* of total effect

Ordered categorical predictors

- Linear predictor equation:

$$\phi_i = \delta_1 + \text{other stuff}$$

effect of first increment of education

Ordered categorical predictors

- Linear predictor equation:

$$\phi_i = \delta_1 + \delta_2 + \text{other stuff}$$

effect of second increment of education

Ordered categorical predictors

- Linear predictor equation:

$$\phi_i = \sum_{j=1}^{7} \delta_j + \text{other stuff}$$

all 7 increments of education

Ordered categorical predictors

- Make all deltas sum to 1, and use leading coefficient for maximum effect:

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

maximum effect of education

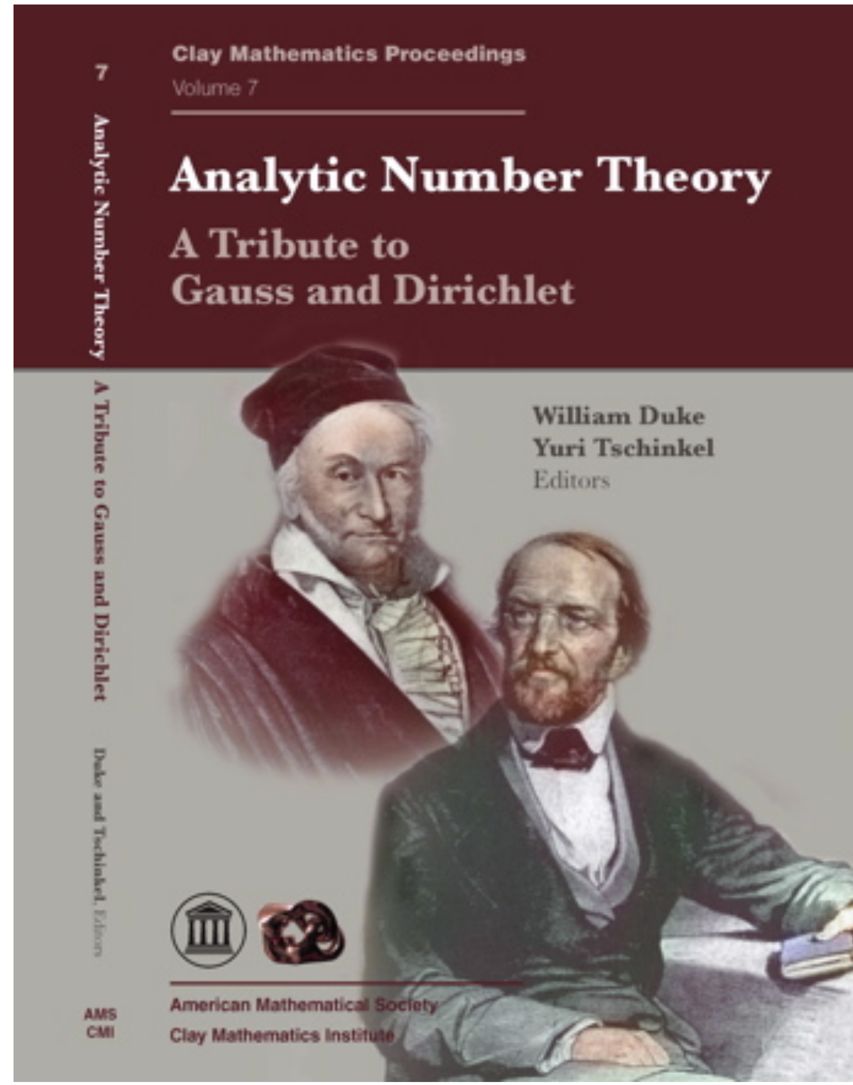
Ordered categorical predictors

- Make all deltas sum to 1, and use leading coefficient for maximum effect:

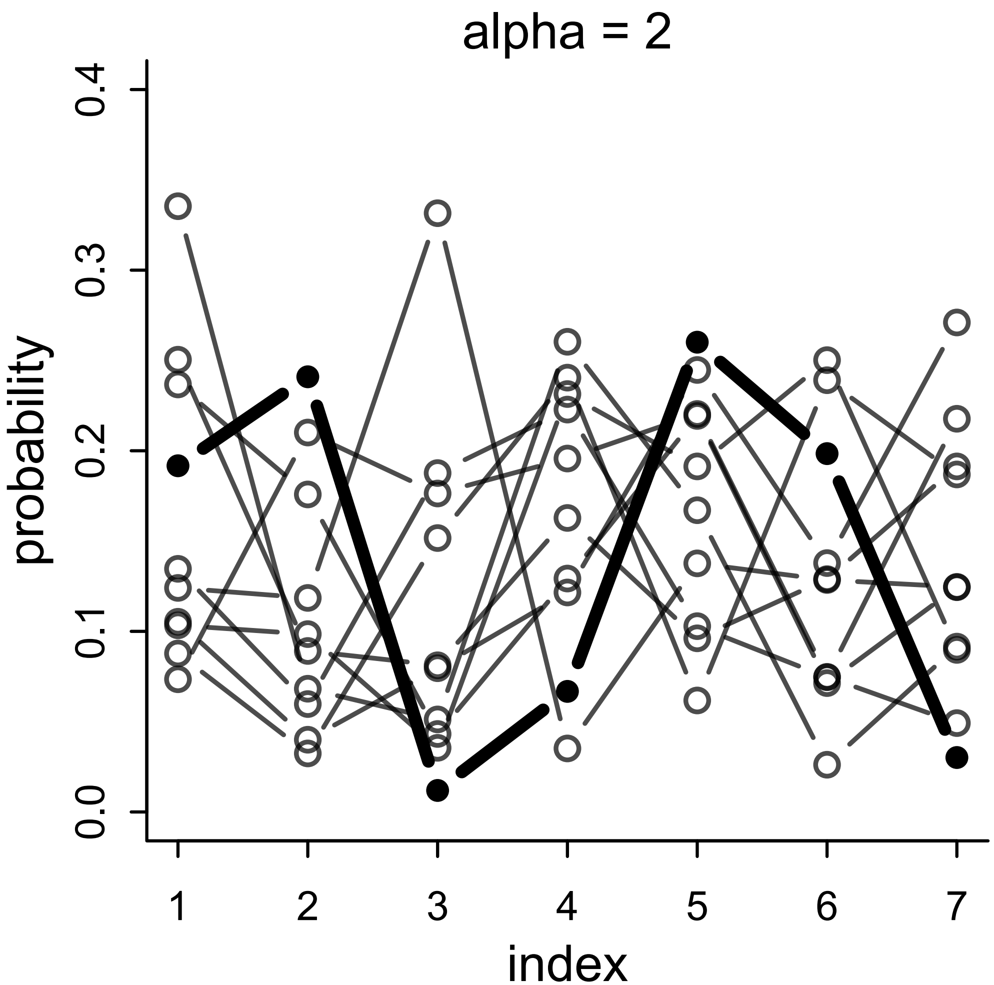
$$
\begin{aligned}
R_i &\sim \text{Ordered-logit}(\phi_i, \kappa) \\
\phi_i &= \beta_E \sum_{j=0}^{E_i-1} \delta_j + \beta_A A_I + \beta_I I_i + \beta_C C_i \\
\kappa_k &\sim \text{Normal}(0, 1.5) \\
\beta_A, \beta_I, \beta_C, \beta_E &\sim \text{Normal}(0, 1) \\
\delta &\sim \text{Dirichlet}(\alpha)
\end{aligned}
$$

Dirichlet (dee-ree-klay)

- Dirichlet: Distribution of N probabilities
- A distribution of distributions
- Generalization of beta distribution
- Shape determined by vector of N parameters
- Each parameter is a *pseudo-count*
  - Large value means that category more probable



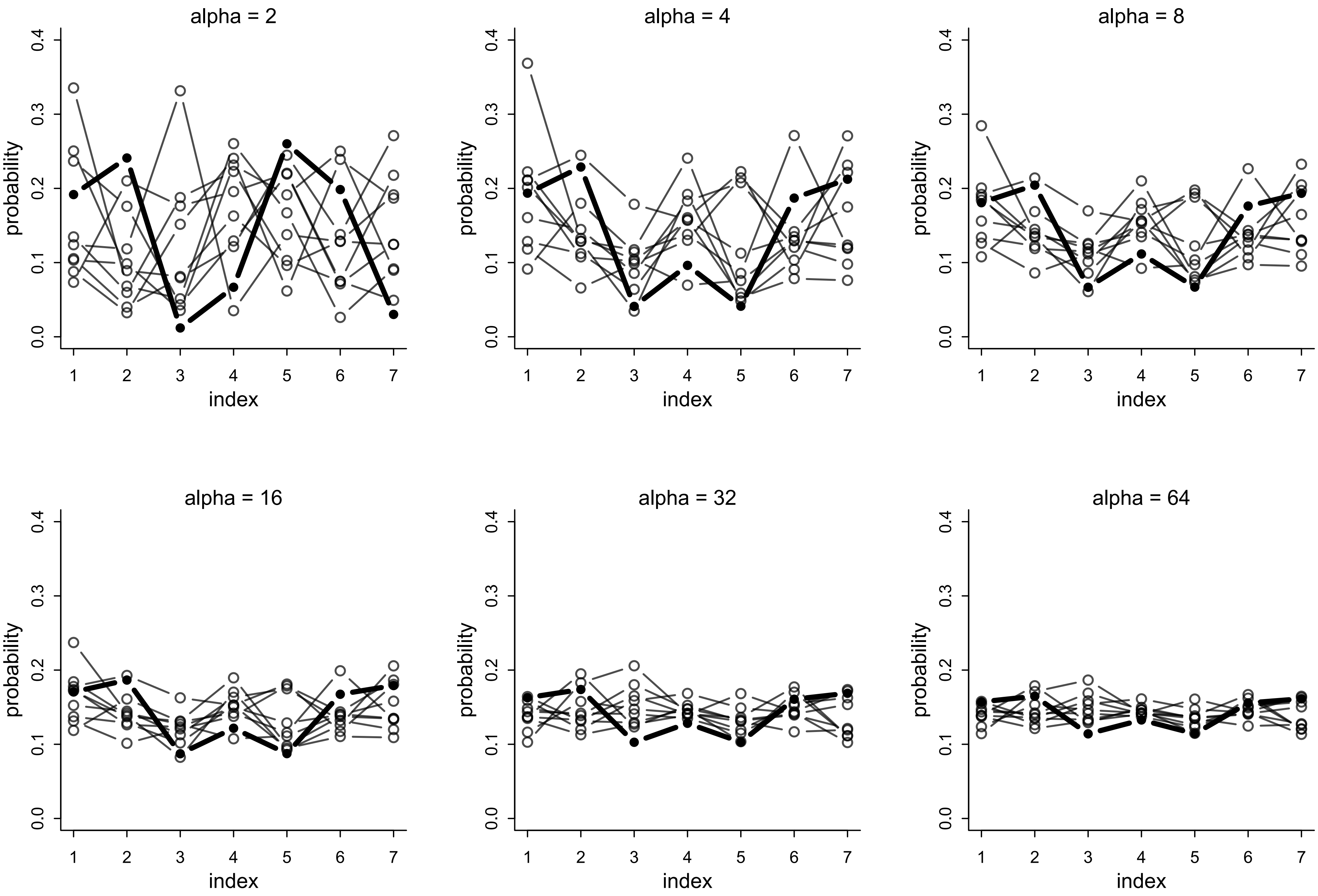
Johann Peter Gustav **Lejeune Dirichlet** (1805–1859)



R code 12.32

```
library(gtools)
set.seed(1805)
delta <- rdirichlet( 10 , alpha=rep(2,7) )
str(delta)
```

```
num [1:10, 1:7] 0.1053 0.2504 0.1917 0.1241 0.0877 ...
```



alpha = 2
alpha = 4
alpha = 8
alpha = 16
alpha = 32
alpha = 64
probability
index
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
7

Ordered categorical predictors

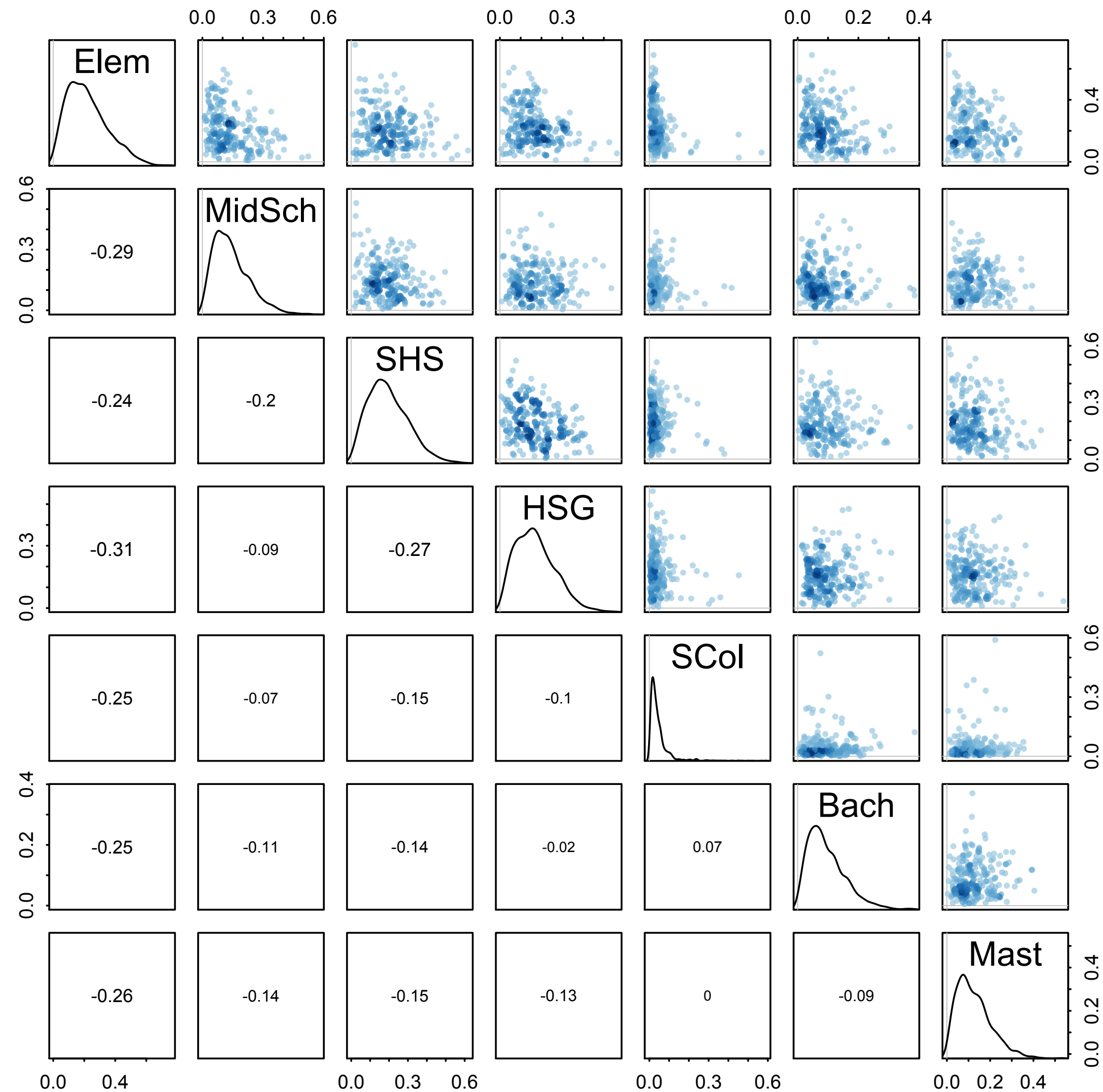
- Use some advanced ulam features
  - Explicit typing of variables (simplex)
  - Vector construction (append_row)

```
m12.5 <- ulam(
    alist(
        R ~ ordered_logistic( phi , kappa ),
        phi <- bE*sum( delta_j[1:E] ) + bA*action + bI*intention + bC*contact,
        kappa ~ normal( 0 , 1.5 ),
        c(bA,bI,bC,bE) ~ normal( 0 , 1 ),
        vector[8]: delta_j <<- append_row( 0 , delta ),
        simplex[7]: delta ~ dirichlet( alpha )
    ),
    data=dat , chains=3 , cores=3 )
```

Ordered categorical predictors

```
precis( m12.5 , depth=2 , omit="cutpoints" )
```

	mean	sd	5.5%	94.5%	n_eff	Rhat
bE	-0.31	0.17	-0.57	-0.05	761	1
bC	-0.96	0.05	-1.03	-0.88	1558	1
bI	-0.72	0.04	-0.77	-0.66	1676	1
bA	-0.71	0.04	-0.77	-0.64	1437	1
delta[1]	0.22	0.13	0.05	0.47	1128	1
delta[2]	0.14	0.09	0.03	0.31	1847	1
delta[3]	0.20	0.11	0.05	0.38	1696	1
delta[4]	0.17	0.10	0.04	0.34	1963	1
delta[5]	0.05	0.06	0.01	0.12	625	1
delta[6]	0.10	0.06	0.02	0.21	1745	1
delta[7]	0.13	0.08	0.03	0.27	2092	1



Ordered categorical predictors

R code
12.35

```
precis( m12.5 , depth=2 , omit="cutpoints" )
```

```
          mean   sd  5.5% 94.5% n_eff Rhat
bE       -0.31 0.17 -0.57 -0.05   761    1
bC       -0.96 0.05 -1.03 -0.88  1558    1
bI       -0.72 0.04 -0.77 -0.66  1676    1
bA       -0.71 0.04 -0.77 -0.64  1437    1
```

```
dat$edu_norm <- normalize( d$edu_new )
m12.6 <- ulam(
    alist(
        y ~ ordered_logistic( mu , cutpoints ),
        mu <- bE*edu_norm + bA*action + bI*intention + bC*contact,
        c(bA,bI,bC,bE) ~ normal( 0 , 1 ),
        cutpoints ~ normal( 0 , 1.5 )
    ), data=dat , chains=3 , cores=3 )
precis( m12.6 )
```

```
              mean   sd  5.5% 94.5% n_eff Rhat
bE           -0.10 0.09 -0.25  0.05   991    1
bC           -0.96 0.05 -1.04 -0.88  1754    1
bI           -0.72 0.04 -0.78 -0.66  1575    1
bA           -0.71 0.04 -0.77 -0.65  1313    1
```

Homeward & onward

- Homework: Online later
- Next week: Multilevel models