

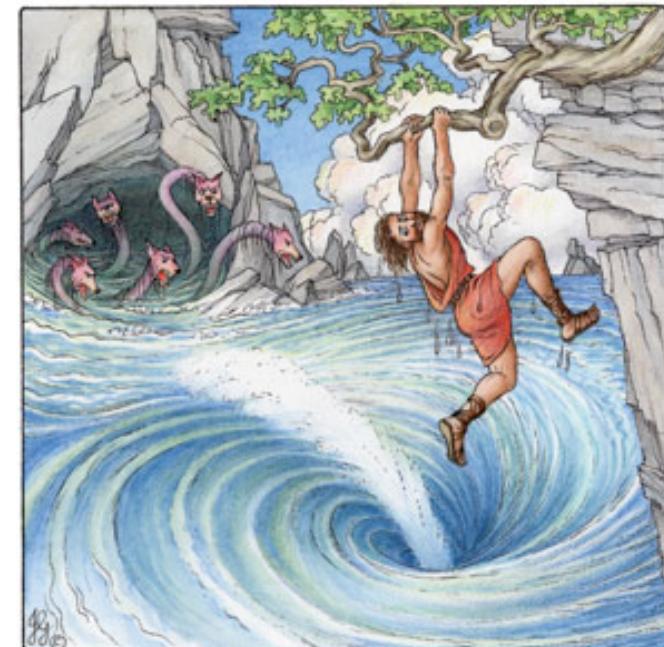
Statistical Rethinking

Week 8: Multilevel Models

Richard McElreath

Ulysses' Compass again

- Why are *varying effects* (partial pooling) more accurate than *fixed effects* (no pooling)?
- Grand mean: maximum underfitting
- Fixed effects: maximum overfitting
- Varying effects: adaptive regularization



```
m12.2 <- map2stan(
  alist(
    surv ~ dbinom( density , p ) ,
    logit(p) <- a_tank[tank] ,
    a_tank[tank] ~ dnorm( a , sigma ) ,
    a ~ dnorm(0,1) ,
    sigma ~ dcauchy(0,1)
  ),
  data=d , iter=4000 , chains=4 )
```

R code
12.3

R code
12.4 compare(m12.1 , m12.2)

	WAIC	pWAIC	dWAIC	weight	SE	dSE
m12.2	1010.2	38.0	0.0	1	37.94	NA
m12.1	1023.3	49.4	13.1	0	43.01	6.54

48 tanks + a + sigma => 50 parameters

Ulysses' Compass again

$$s_i \sim \text{Binomial}(n_i, p_i)$$

$$\text{logit}(p_i) = \alpha_{\text{TANK}}[i]$$

$$\alpha_{\text{TANK}} \sim \text{Normal}(\alpha, \sigma)$$

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

$$\sigma \sim \text{HalfCauchy}(0, 1)$$



σ

Ulysses' Compass again

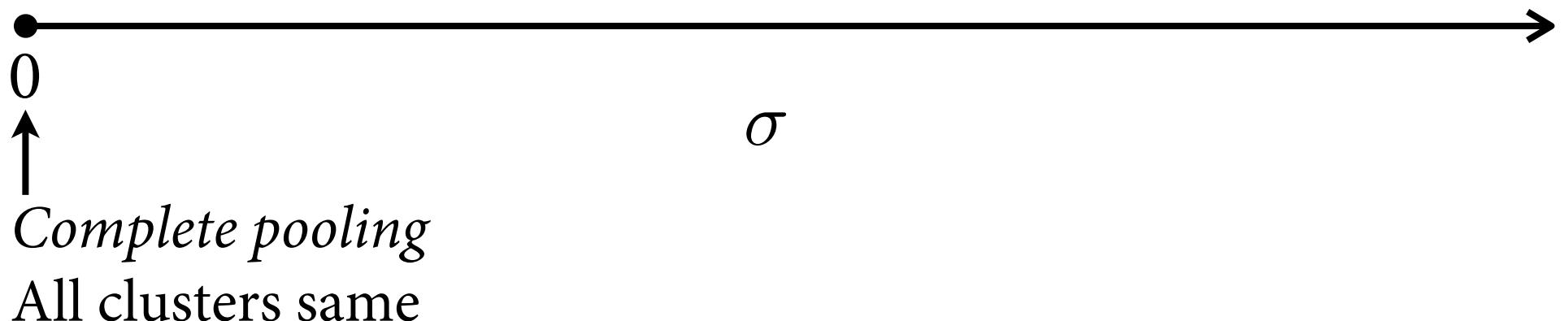
$$s_i \sim \text{Binomial}(n_i, p_i)$$

$$\text{logit}(p_i) = \alpha_{\text{TANK}}[i]$$

$$\alpha_{\text{TANK}} \sim \text{Normal}(\alpha, \sigma)$$

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

$$\sigma \sim \text{HalfCauchy}(0, 1)$$



Ulysses' Compass again

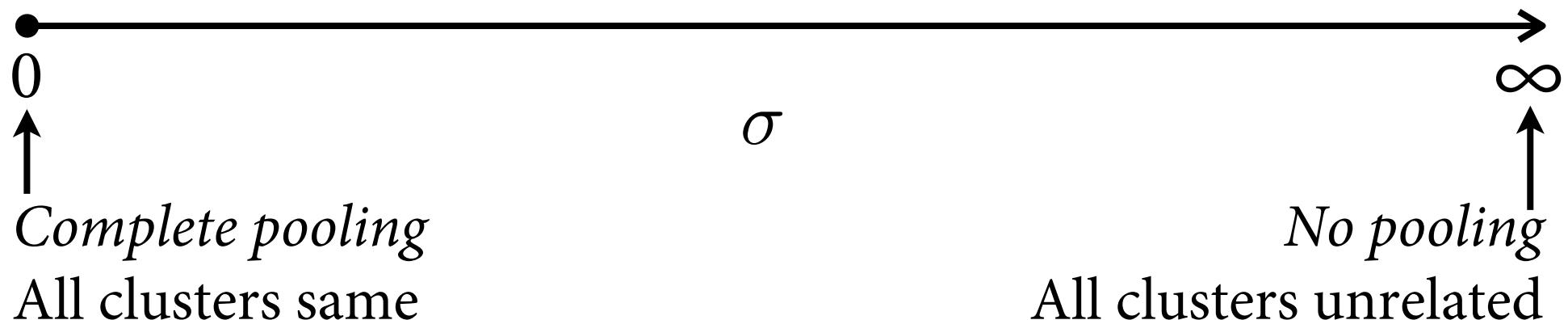
$$s_i \sim \text{Binomial}(n_i, p_i)$$

$$\text{logit}(p_i) = \alpha_{\text{TANK}}[i]$$

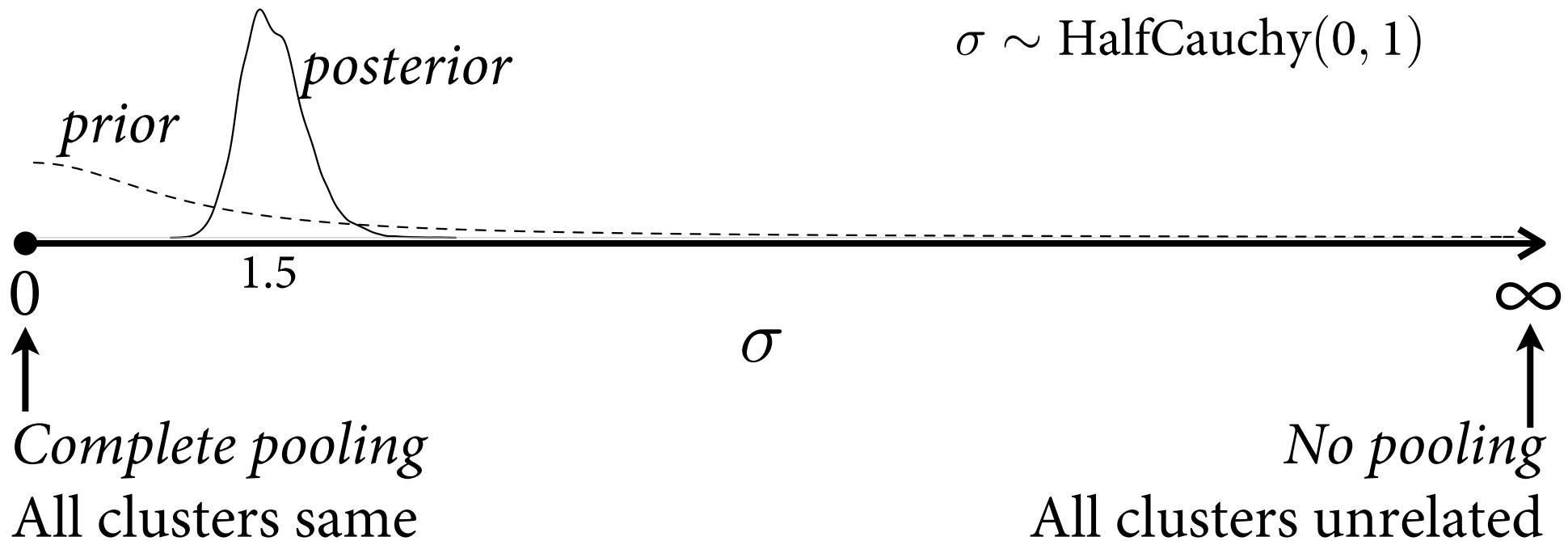
$$\alpha_{\text{TANK}} \sim \text{Normal}(\alpha, \sigma)$$

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

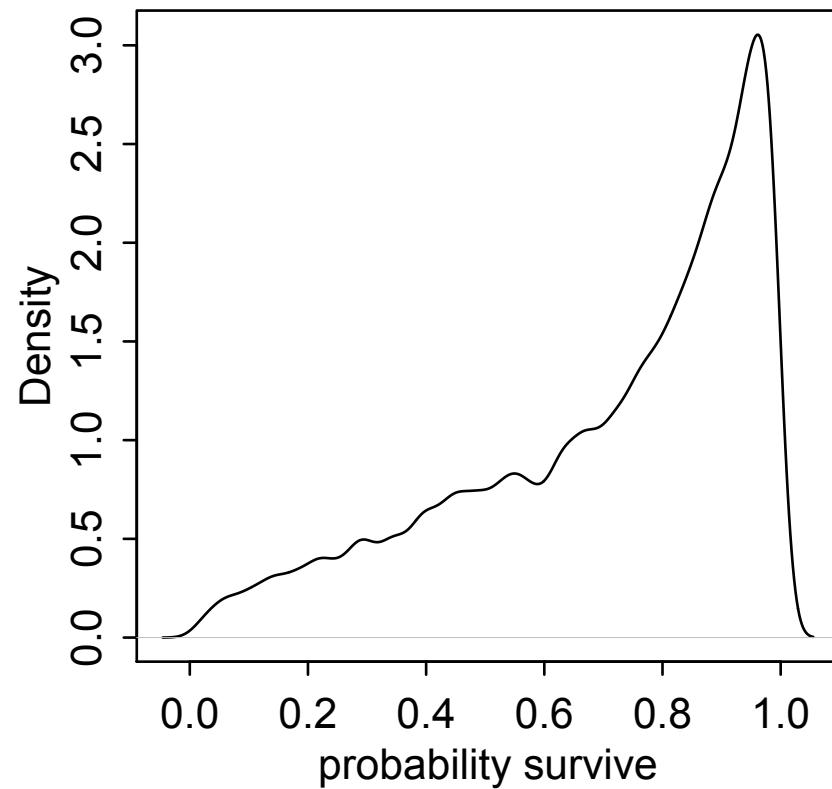
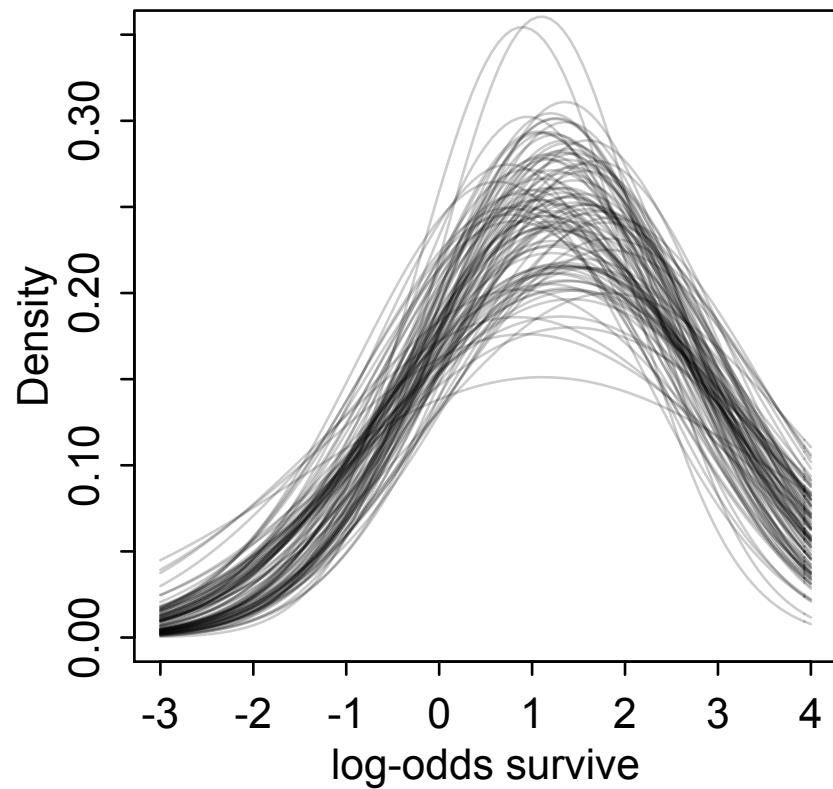
$$\sigma \sim \text{HalfCauchy}(0, 1)$$



Ulysses' Compass again



Regularizing distribution

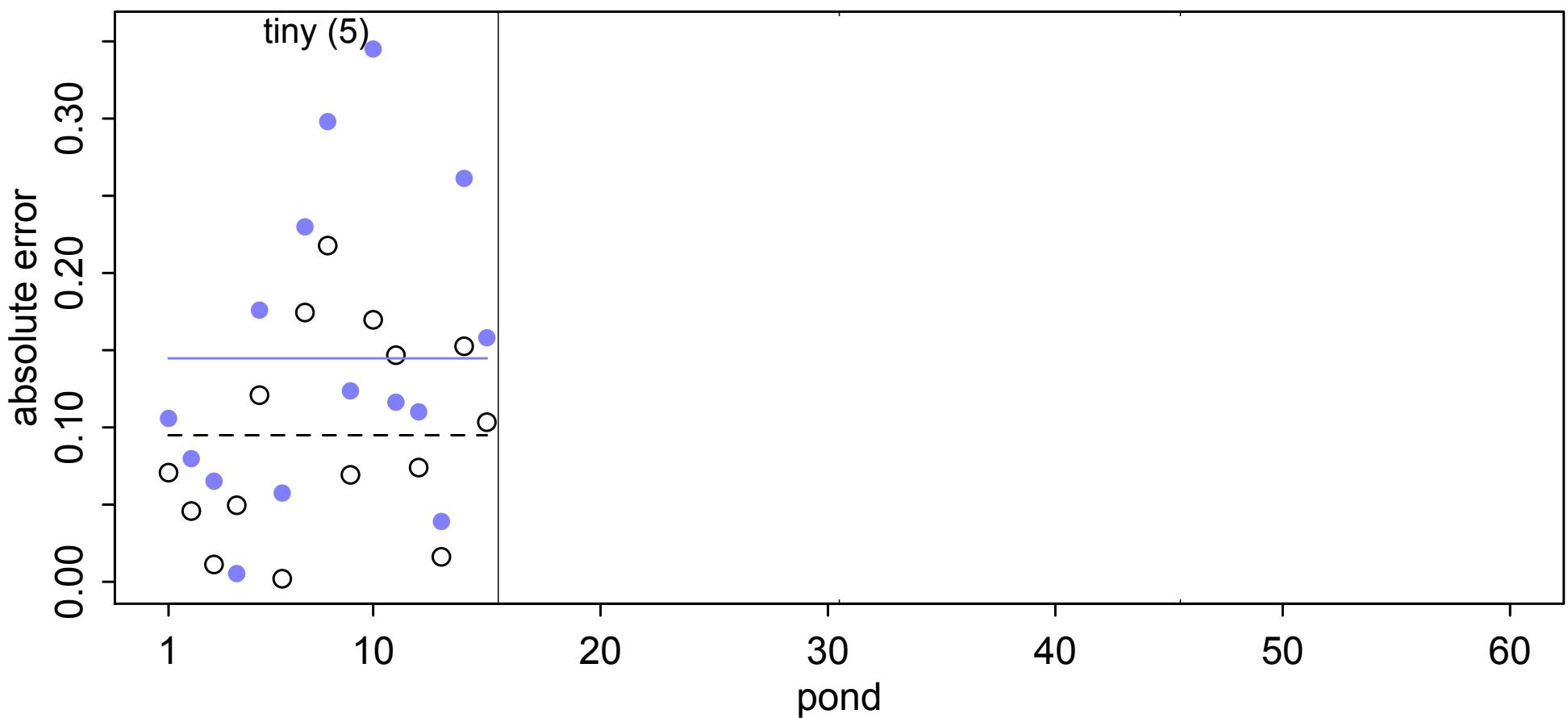


Ulysses' Compass again

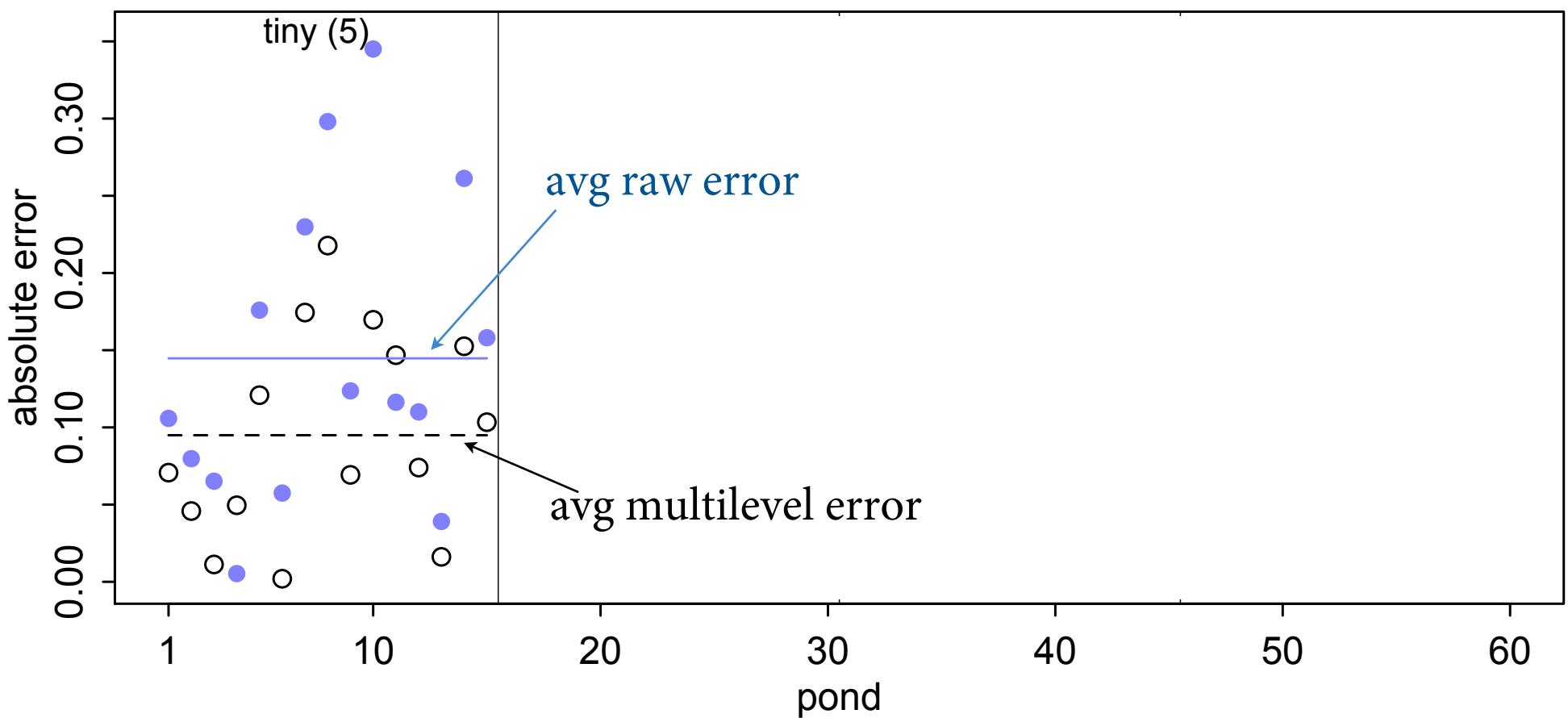
- Simulate to demonstrate accuracy advantage
 - 60 ponds
 - 5, 10, 25, 35 tadpoles each of 15

	pond	n	true.a	s	p.nopool	p.partpool	p.true
1	1	5	-3.089936132	1	0.2000000	0.32173203	0.04352429
2	2	5	0.267290817	5	1.0000000	0.91305884	0.56642768
3	3	5	0.896554101	4	0.8000000	0.79164823	0.71024085
4	4	5	1.934806220	5	1.0000000	0.91276066	0.87378044
5	5	5	-0.758682067	0	0.0000000	0.17692527	0.31893247
6	6	5	3.904836388	5	1.0000000	0.91337140	0.98025353
7	7	5	2.271914139	4	0.8000000	0.79349508	0.90652411
8	8	5	2.886101619	4	0.8000000	0.79557800	0.94715510
9	9	5	1.436457877	3	0.6000000	0.64219989	0.80790553
10	10	5	1.156079068	3	0.6000000	0.64414477	0.76061953

● Raw proportion ○ Multilevel estimate

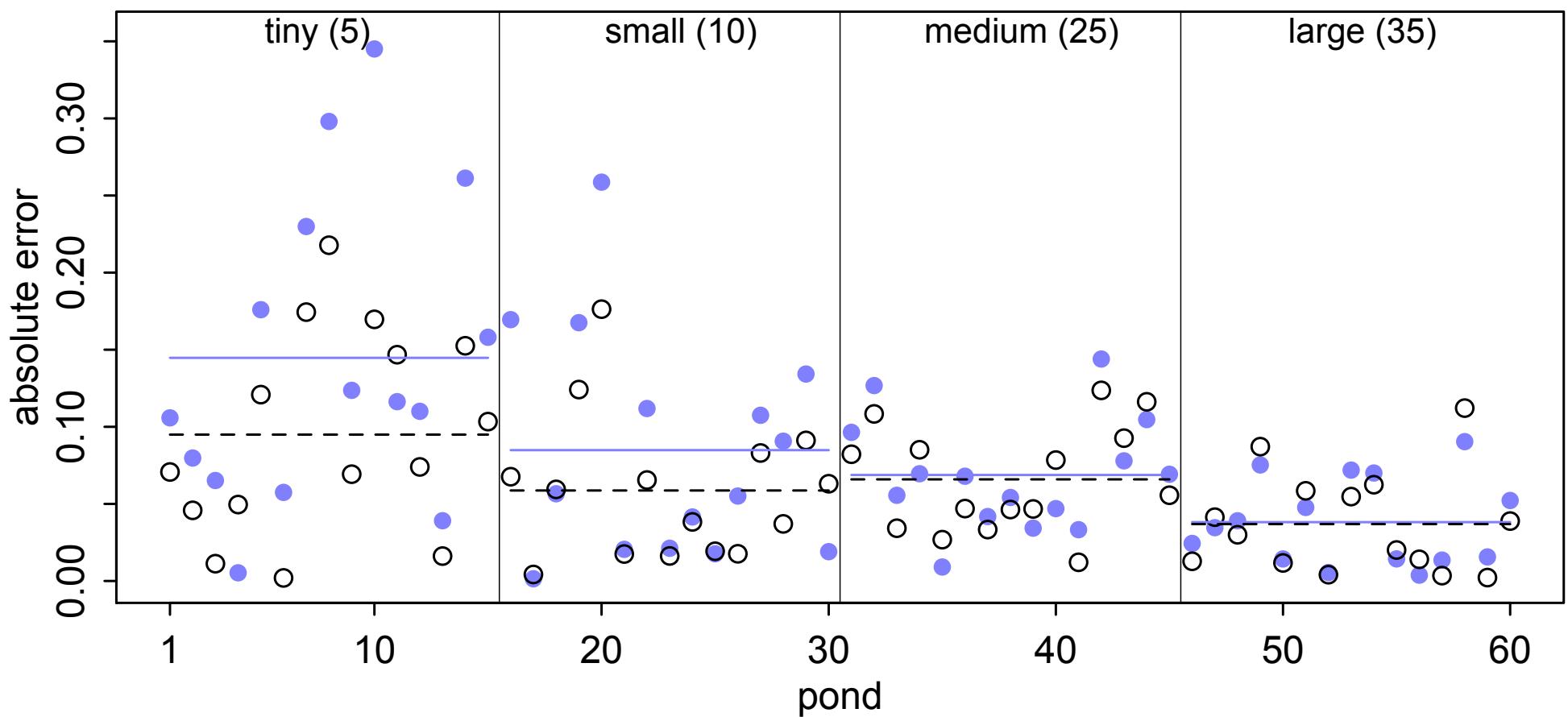


● Raw proportion ○ Multilevel estimate

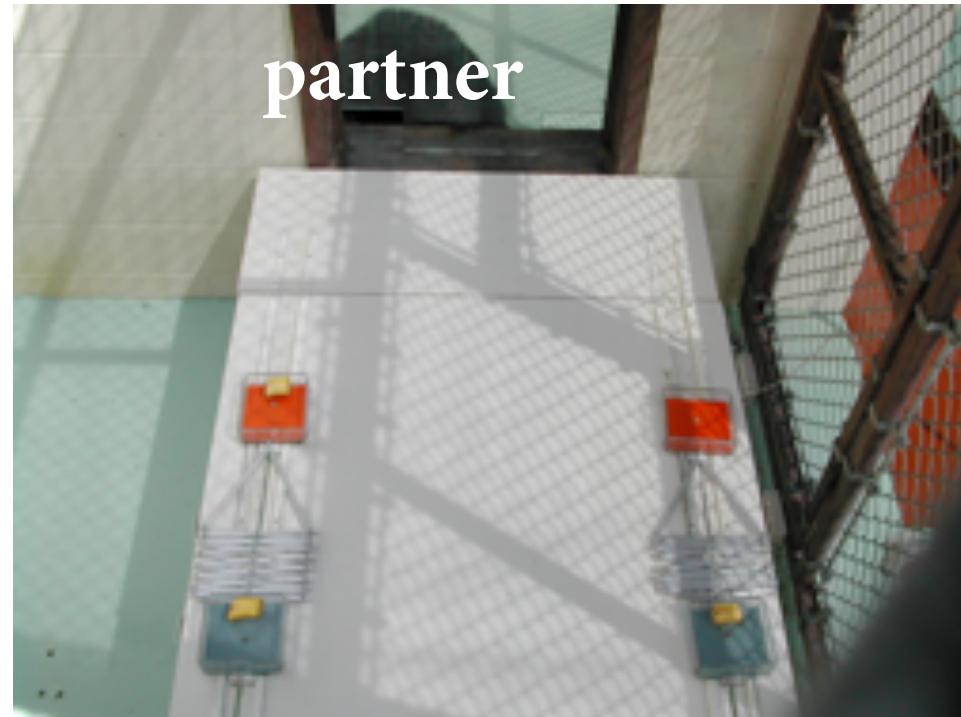
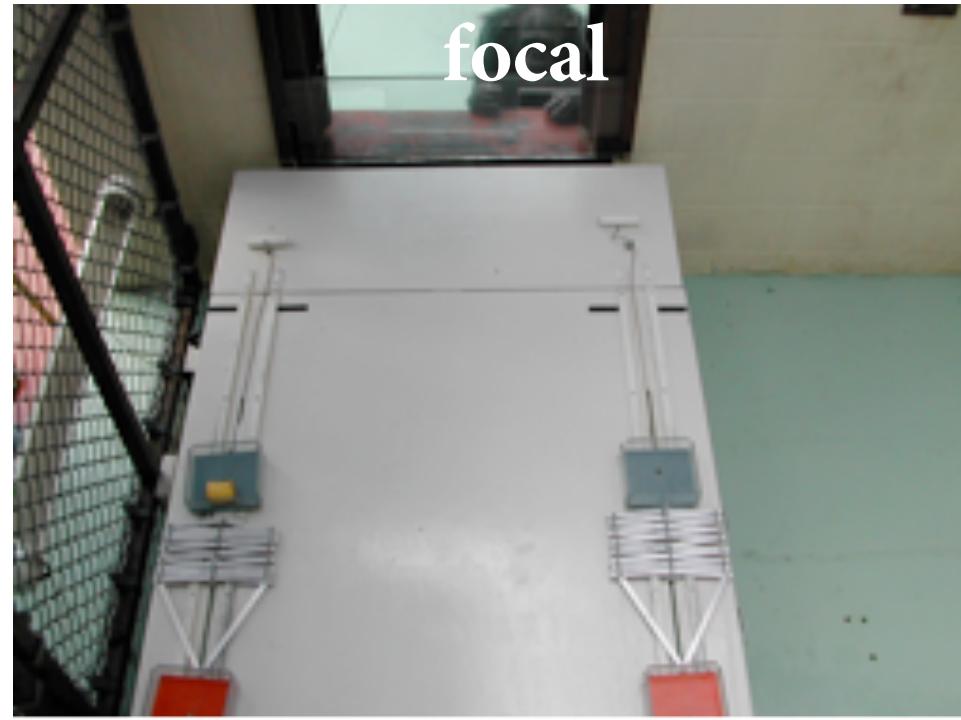
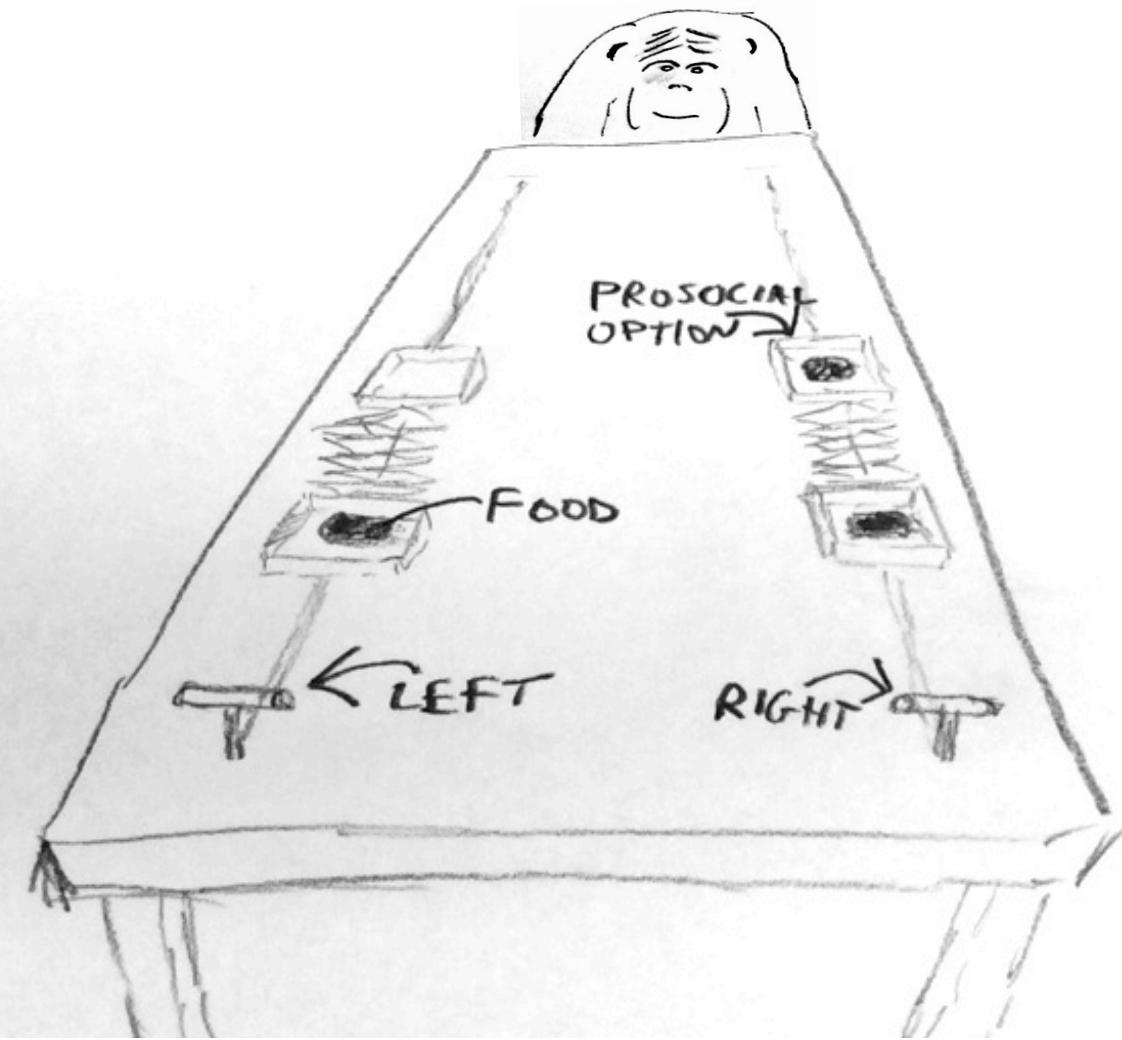


● Raw proportion

○ Multilevel estimate

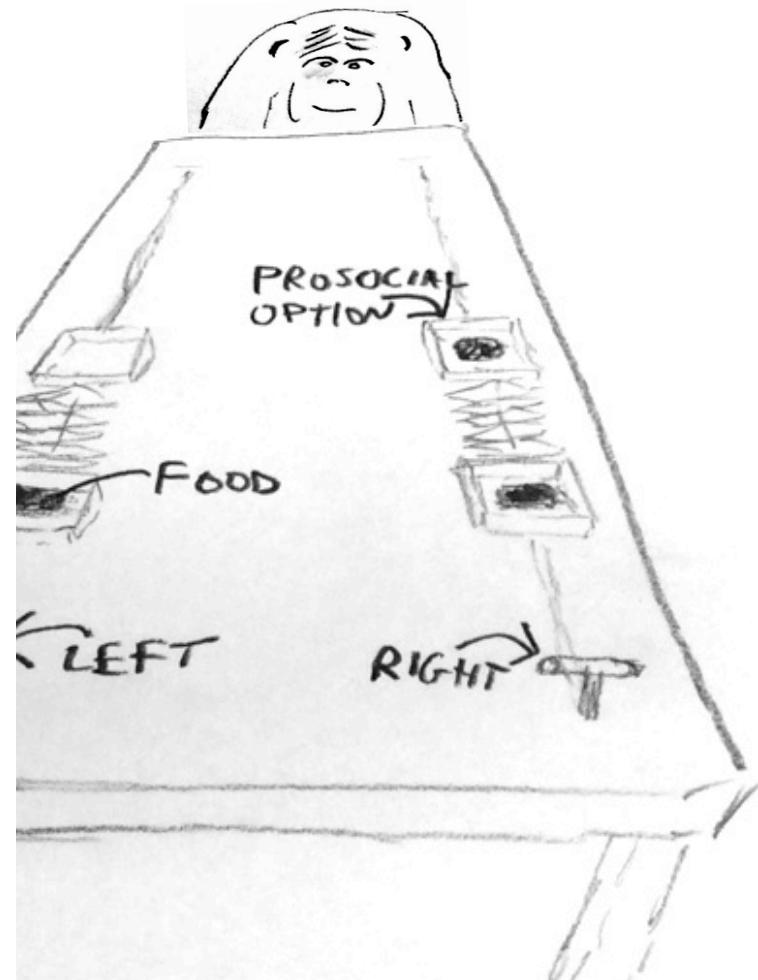


Prosocial chimpanzees



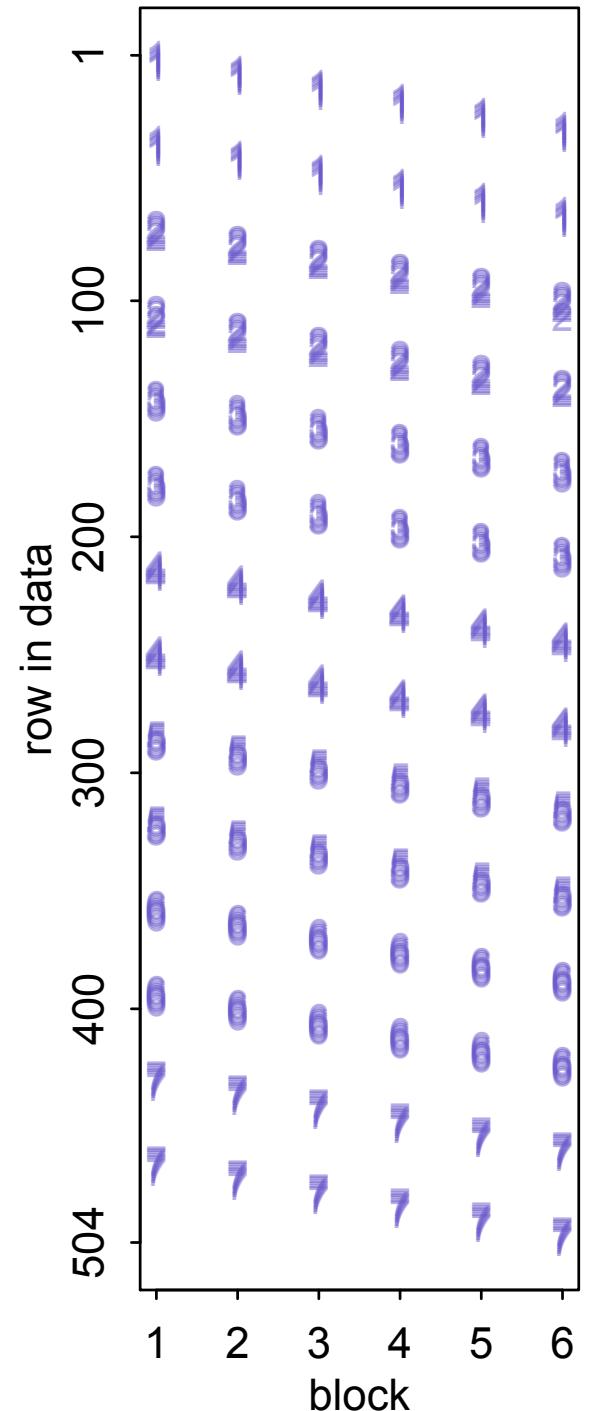
Prosocial chimpanzees

- Two *conditions*: (1) partner, (2) alone
- Two *options*: (1) prosocial, (2) asocial
- Two *outcomes*: (1) left, (2) right
- Six *blocks* (sessions)
- Seven *actors* (individuals)
- Want to predict *outcome* as function of *condition* and where prosocial *option* is
 - Do chimps prefer *left* lever when *partner* present and *prosocial* on *left*?



Cross-classification

- Can use more than one cluster type
- Chimpanzee experiment data
 - Pulls in chimpanzees
 - Pulls in blocks
 - Each chimp in each block
 - Not nested, but *cross-classified*



Multilevel chimpanzees

$$L_i \sim \text{Binomial}(1, p_i)$$

$$\text{logit}(p_i) = \alpha + \alpha_{\text{ACTOR}[i]} + (\beta_P + \beta_{PC}C_i)P_i$$

varying intercepts on actor $\longrightarrow \alpha_{\text{ACTOR}} \sim \text{Normal}(0, \sigma_{\text{ACTOR}})$

$$\alpha \sim \text{Normal}(0, 10)$$

$$\beta_P \sim \text{Normal}(0, 10)$$

$$\beta_{PC} \sim \text{Normal}(0, 10)$$

$$\sigma_{\text{ACTOR}} \sim \text{HalfCauchy}(0, 1)$$

Mean alpha in linear model now. Is equivalent.

Multilevel chimpanzees

```
m12.4 <- map2stan(  
  alist(  
    pulled_left ~ dbinom( 1 , p ) ,  
    logit(p) <- a + a_actor[actor] +  
      (bp + bpC*condition)*prosoc_left ,  
    a_actor[actor] ~ dnorm( 0 , sigma_actor ) ,  
    a ~ dnorm(0,10) ,  
    bp ~ dnorm(0,10) ,  
    bpC ~ dnorm(0,10) ,  
    sigma_actor ~ dcauchy(0,1)  
  ) ,  
  data=d )
```

$$L_i \sim \text{Binomial}(1, p_i)$$

$$\text{logit}(p_i) = \alpha + \alpha_{\text{ACTOR}[i]} + (\beta_P + \beta_{PC}C_i)P_i$$

$$\alpha_{\text{ACTOR}} \sim \text{Normal}(0, \sigma_{\text{ACTOR}})$$

$$\alpha \sim \text{Normal}(0, 10)$$

$$\beta_P \sim \text{Normal}(0, 10)$$

$$\beta_{PC} \sim \text{Normal}(0, 10)$$

$$\sigma_{\text{ACTOR}} \sim \text{HalfCauchy}(0, 1)$$

Cross-classified chimpanzees

$$L_i \sim \text{Binomial}(1, p_i)$$

$$\text{logit}(p_i) = \alpha + \alpha_{\text{ACTOR}[i]} + \alpha_{\text{BLOCK}[i]} + (\beta_P + \beta_{PC}C_i)P_i$$

varying intercepts on actor $\longrightarrow \alpha_{\text{ACTOR}} \sim \text{Normal}(0, \sigma_{\text{ACTOR}})$

varying intercepts on block $\longrightarrow \alpha_{\text{BLOCK}} \sim \text{Normal}(0, \sigma_{\text{BLOCK}})$

joint mean $\longrightarrow \alpha \sim \text{Normal}(0, 10)$

$$\beta_P \sim \text{Normal}(0, 10)$$

$$\beta_{PC} \sim \text{Normal}(0, 10)$$

$$\sigma_{\text{ACTOR}} \sim \text{HalfCauchy}(0, 1)$$

$$\sigma_{\text{BLOCK}} \sim \text{HalfCauchy}(0, 1)$$

Just one “alpha” for both cluster types.
Otherwise unidentified parameters.

Cross-classified chimpanzees

```
m12.5 <- map2stan(  
  alist(  
    pulled_left ~ dbinom( 1 , p ),  
    logit(p) <- a + a_actor[actor] +  
      a_block[block_num] +  
      (bp + bpc*condition)*prosoc_left,  
    a_actor[actor] ~ dnorm( 0 , sigma_actor ),  
    a_block[block_num] ~ dnorm( 0 , sigma_block ),  
    c(a,bp,bpc) ~ dnorm(0,10),  
    sigma_actor ~ dcauchy(0,1),  
    sigma_block ~ dcauchy(0,1)  
  ) ,  
  data=d )
```

$$L_i \sim \text{Binomial}(1, p_i)$$

$$\logit(p_i) = \alpha + \alpha_{\text{ACTOR}[i]} + \alpha_{\text{BLOCK}[i]} + (\beta_P + \beta_{PC} C_i)$$

$$\alpha_{\text{ACTOR}} \sim \text{Normal}(0, \sigma_{\text{ACTOR}})$$

$$\alpha_{\text{BLOCK}} \sim \text{Normal}(0, \sigma_{\text{BLOCK}})$$

$$\alpha \sim \text{Normal}(0, 10)$$

$$\beta_P \sim \text{Normal}(0, 10)$$

$$\beta_{PC} \sim \text{Normal}(0, 10)$$

$$\sigma_{\text{ACTOR}} \sim \text{HalfCauchy}(0, 1)$$

$$\sigma_{\text{BLOCK}} \sim \text{HalfCauchy}(0, 1)$$

Cross-classified chimpanzees

```
m12.5 <- map2stan(  
  alist(  
    pulled_left ~ dbinom( 1 , p ),  
    logit(p) <- a + a_actor[actor] +  
      a_block[block_num] +  
      (bp + bpc*condition)*prosoc_left,  
    a_actor[actor] ~ dnorm( 0 , sigma_actor ),  
    a_block[block_num] ~ dnorm( 0 , sigma_block ),  
    c(a,bp,bpc) ~ dnorm(0,10),  
    sigma_actor ~ dcauchy(0,1),  
    sigma_block ~ dcauchy(0,1)  
  ) ,  
  data=d )
```

$$L_i \sim \text{Binomial}(1, p_i)$$

$$\logit(p_i) = \alpha + \alpha_{\text{ACTOR}[i]} + \alpha_{\text{BLOCK}[i]} + (\beta_P + \beta_{PC} C_i)$$

$$\alpha_{\text{ACTOR}} \sim \text{Normal}(0, \sigma_{\text{ACTOR}})$$

$$\alpha_{\text{BLOCK}} \sim \text{Normal}(0, \sigma_{\text{BLOCK}})$$

$$\alpha \sim \text{Normal}(0, 10)$$

$$\beta_P \sim \text{Normal}(0, 10)$$

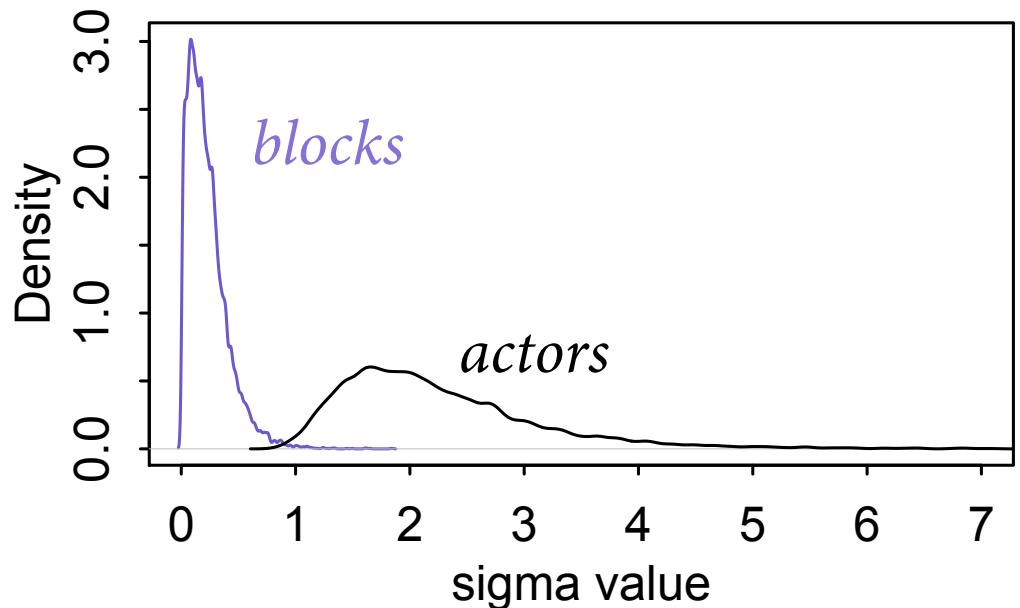
$$\beta_{PC} \sim \text{Normal}(0, 10)$$

$$\sigma_{\text{ACTOR}} \sim \text{HalfCauchy}(0, 1)$$

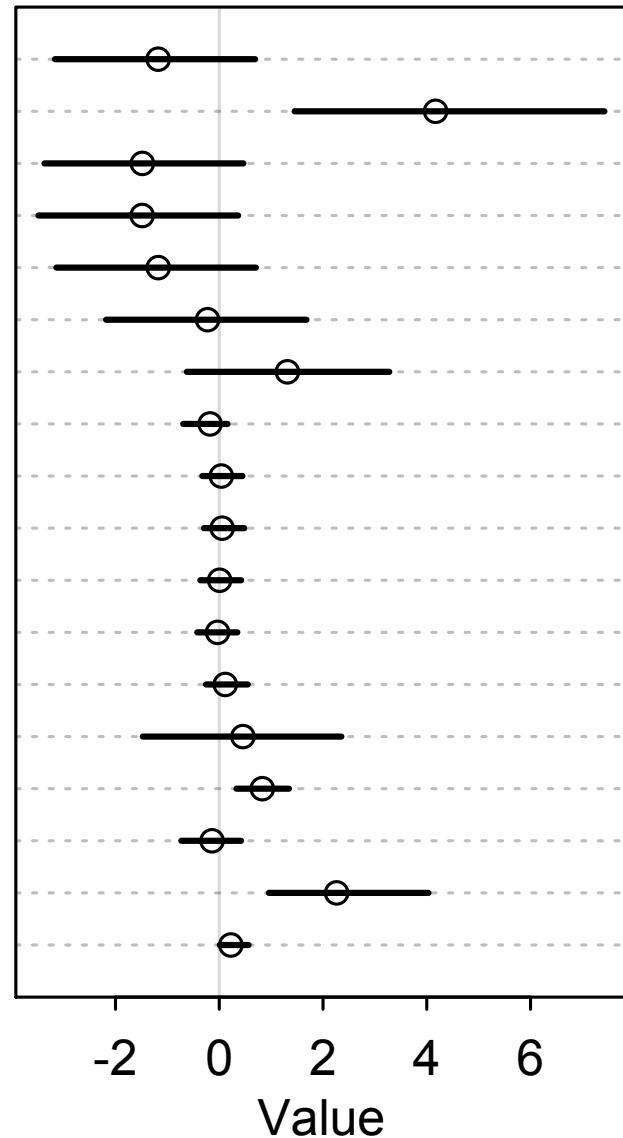
$$\sigma_{\text{BLOCK}} \sim \text{HalfCauchy}(0, 1)$$

Cross-classified chimpanzees

- Lots of variation among actors
- Little variation among blocks
- `a_actor`'s vary a lot
- `a_block`'s vary hardly at all



`a_actor[1]`
`a_actor[2]`
`a_actor[3]`
`a_actor[4]`
`a_actor[5]`
`a_actor[6]`
`a_actor[7]`
`a_block[1]`
`a_block[2]`
`a_block[3]`
`a_block[4]`
`a_block[5]`
`a_block[6]`
`a`
`bp`
`bpc`
`sigma_actor`
`sigma_block`



Cross-classified chimpanzees

- Incorporating block: no anticipated benefits; little cost

Vary	Parameters	Effective parameters	WAIC	weight
actor, block	18	11	533	0.35
actor	11	8	532	0.65

```
compare(m12.4,m12.5)
```

R code
12.25

	WAIC	pWAIC	dWAIC	weight	SE	dSE
m12.4	531.5	8.1	0.0	0.65	19.50	NA
m12.5	532.7	10.5	1.2	0.35	19.74	1.94

Posterior predictions

- Predictions more subtle: Same clusters or new clusters?
- Same clusters: proceed as usual
- New clusters: should average over distribution of varying effects
- In this case:
 - Same clusters: Predictions for these chimpanzees
 - New clusters: Prediction for a new chimpanzee or rather for population of chimpanzees

Same clusters, new clusters

- Same actors:
 - Really same as before: varying effects are just parameters; you know the model; push samples back through the model
 - `link()` and `sim()` obey this rule
- New actors (counterfactual):
 - which actor (cluster) to use for counterfactual predictions?
 - average actor
 - marginal of actor
 - show sample of actors from posterior

Average actor

- “average actor” means actor with population average intercept, “alpha”
- Strategy:
 - replace varying intercept samples with zeros => all actors have average intercept now
 - compute predictions as usual

```
# replace varying intercept samples with zeros
# 1000 samples by 7 actors
a_actor_zeros <- matrix(0,1000,7)
```

R code
12.32

```
# fire up link
# note use of replace list
link.m12.4 <- link( m12.4 , n=1000 , data=d.pred ,
replace=list(a_actor=a_actor_zeros) )
```

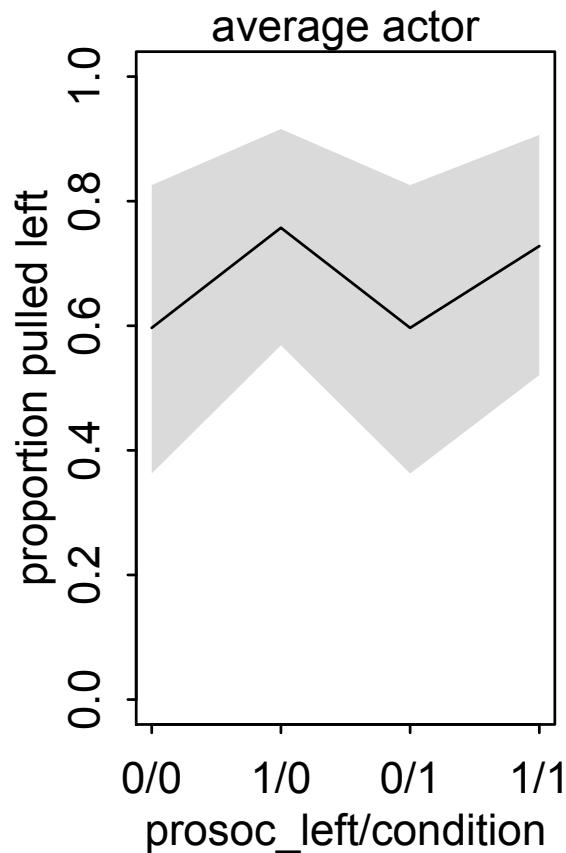
R code
12.33

```
# replace varying intercept samples with zeros  
# 1000 samples by 7 actors  
a_actor_zeros <- matrix(0,1000,7)
```

R code
12.32

```
# fire up link  
# note use of replace list  
link.m12.4 <- link( m12.4 , n=1000 , data=d.pred ,  
    replace=list(a_actor=a_actor_zeros) )
```

R code
12.33



Marginal of actor

- “Marginal of” means “averaging over variation in actors” => shows variation arising from variation across actors

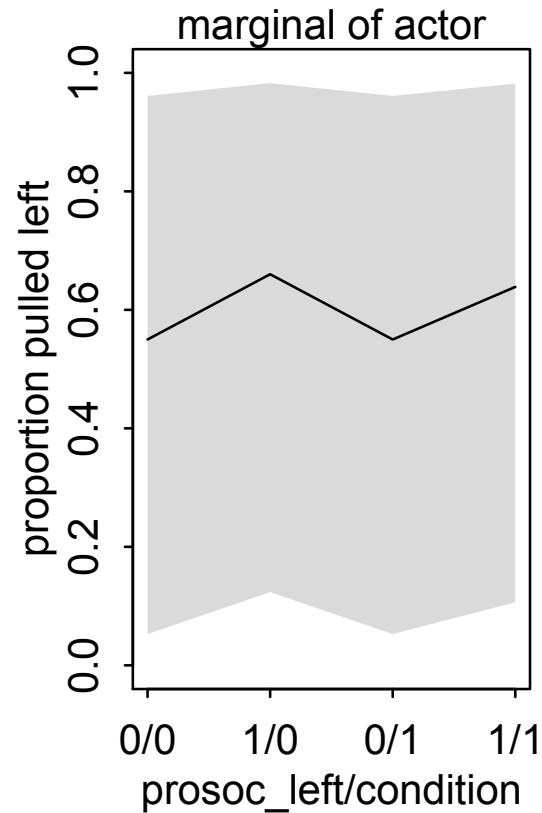
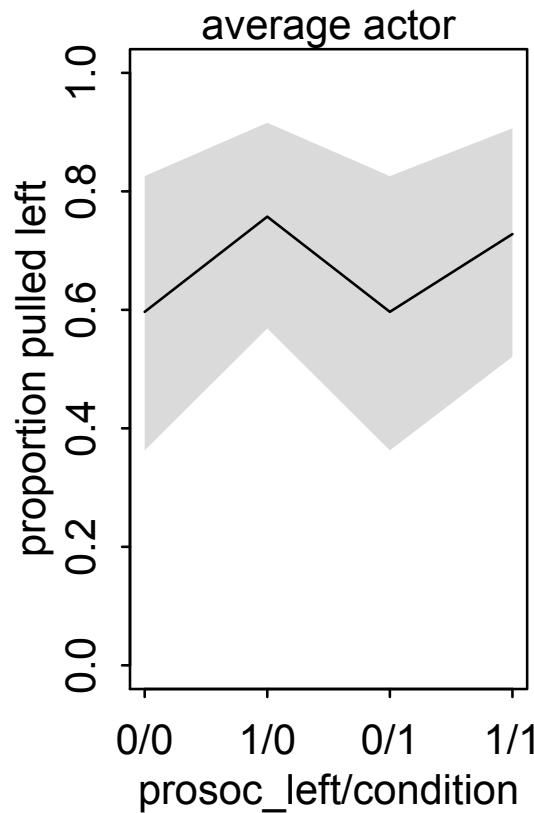
$$\alpha_{\text{ACTOR}} \sim \text{Normal}(0, \sigma_{\text{ACTOR}})$$

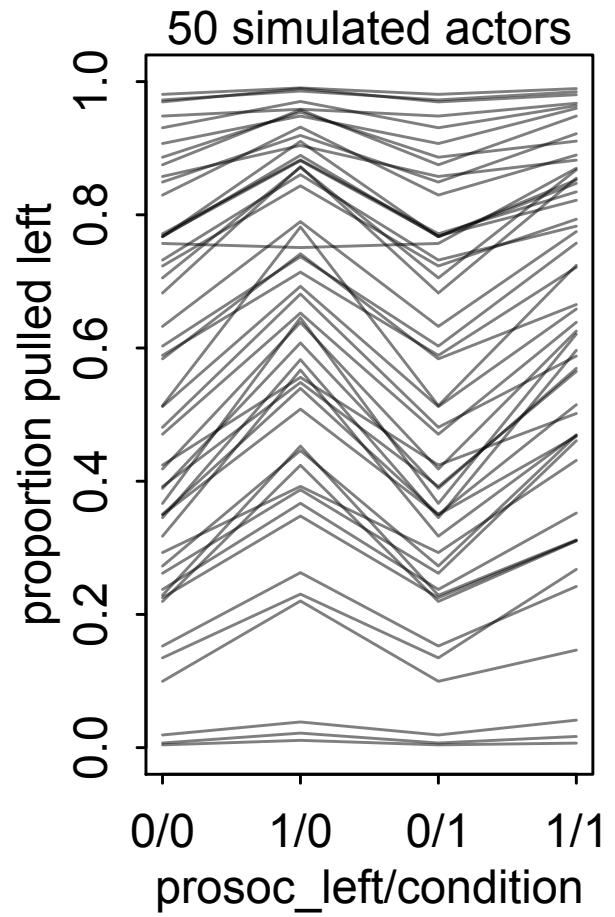
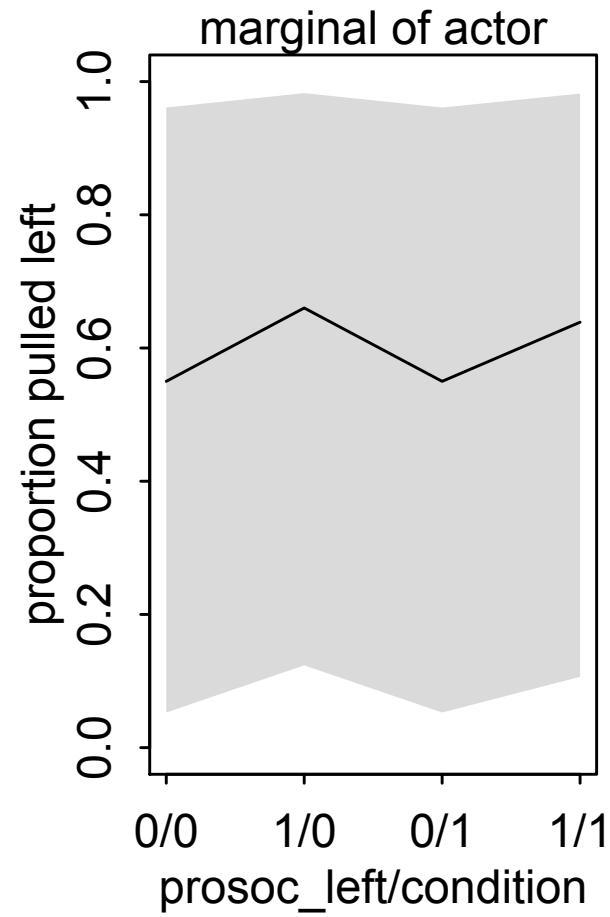
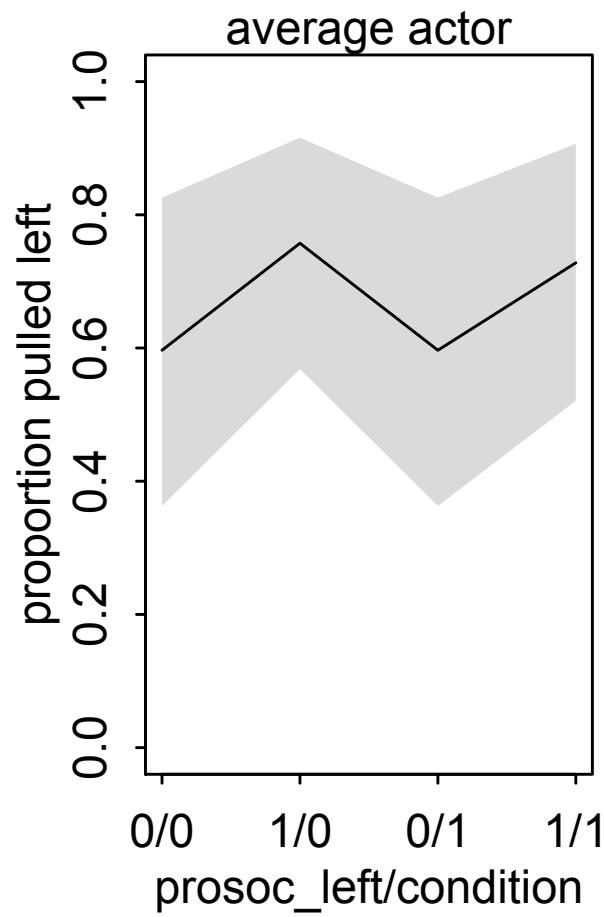
- Strategy:
 - Extract samples for sigma_actor
 - Simulate new varying intercepts
 - Use simulated intercepts to simulate predictions

R code
12.34

```
# replace varying intercept samples with simulations
post <- extract.samples(m12.4)
a_actor_sims <- rnorm(7000,0,post$sigma_actor)
a_actor_sims <- matrix(a_actor_sims,1000,7)

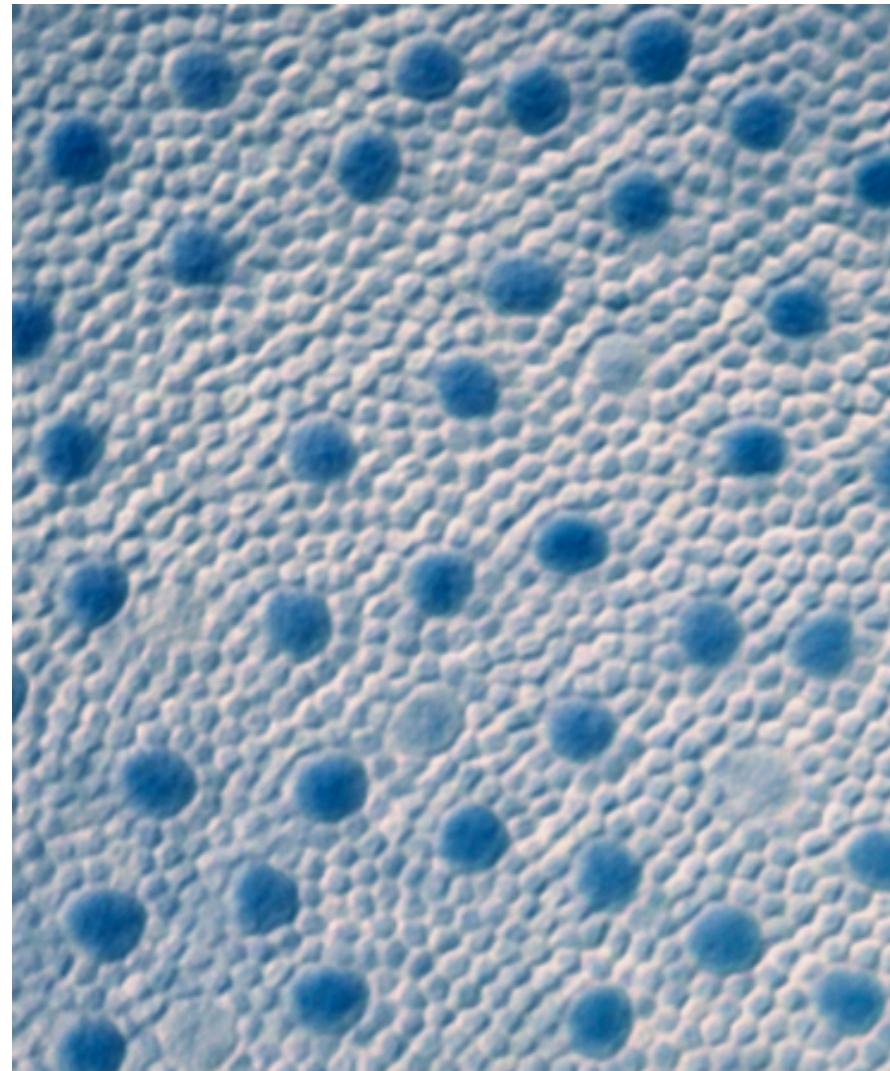
# fire up link
# note use of replace list
link.m12.4 <- link( m12.4 , n=1000 , data=d.pred ,
replace=list(a_actor=a_actor_sims) )
```





Multilevel overdispersion

- Overdispersion: Count data with residual variation greater than expectation
- Implies unmodeled heterogeneity across cases
- Can estimate that heterogeneity with varying intercepts on each case
 - Estimate varying intercept for each *observation* in the data



Human photoreceptors, up close

Multilevel islands

- Recall Oceanic tools model

	culture	population	contact	total_tools	mean_TU
1	Malekula	1100	low	13	3.2
2	Tikopia	1500	low	22	4.7
3	Santa Cruz	3600	low	24	4.0
4	Yap	4791	high	43	5.0
5	Lau Fiji	7400	high	33	5.0
6	Trobriand	8000	high	19	4.0
7	Chuuk	9200	high	40	3.8
8	Manus	13000	low	28	6.6
9	Tonga	17500	high	55	5.4
10	Hawaii	275000	low	71	6.6

$$T_i \sim \text{Poisson}(\mu_i)$$

$$\log(\mu_i) = \alpha + \boxed{\alpha_{\text{ISLAND}}[i]} + \beta_P \log P_i$$

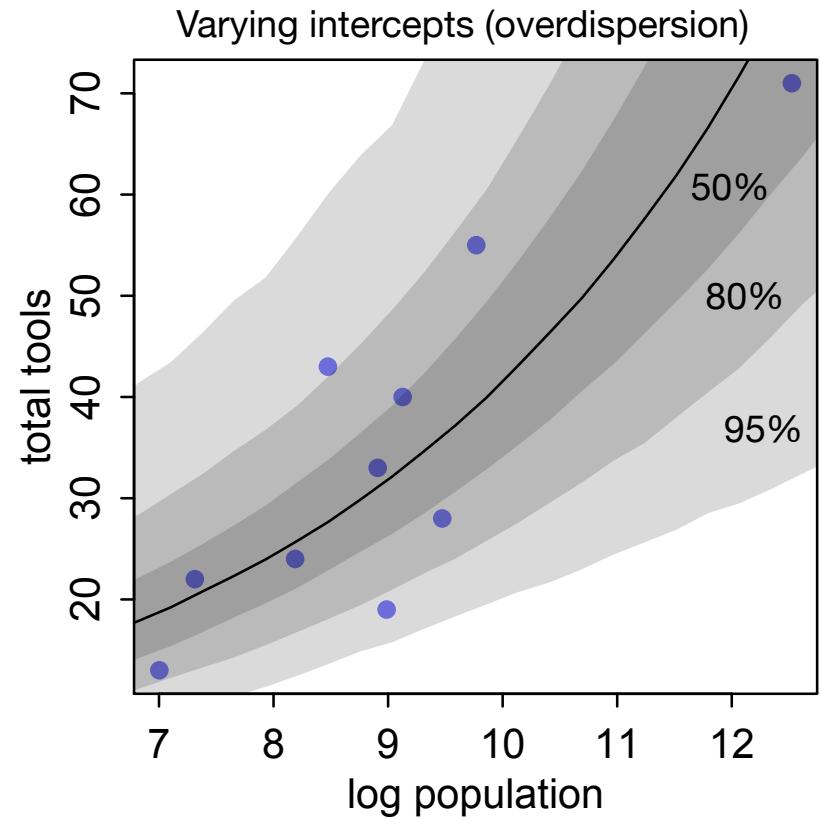
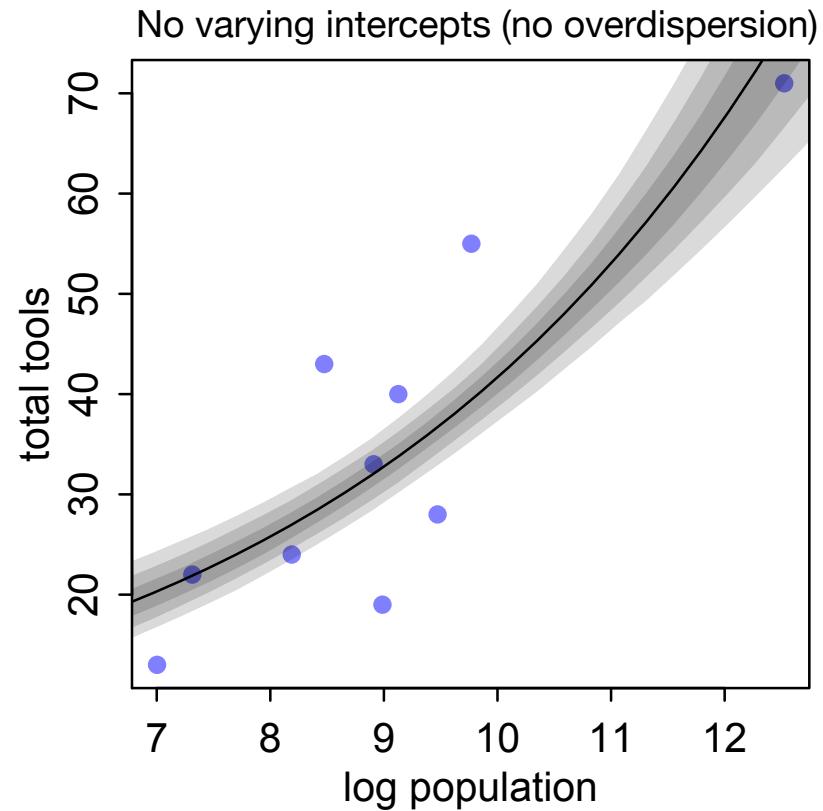
$$\alpha \sim \text{Normal}(0, 10)$$

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

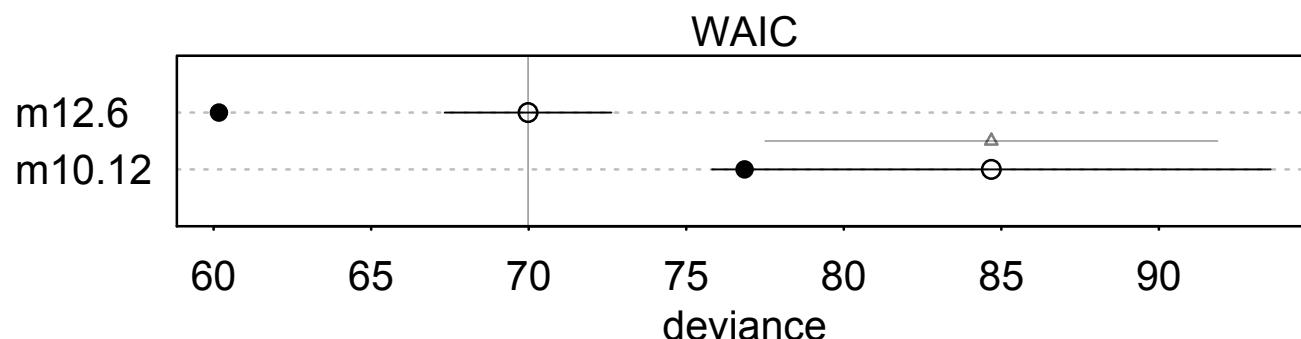
$$\alpha_{\text{ISLAND}} \sim \text{Normal}(0, \sigma_{\text{ISLAND}})$$

$$\sigma_{\text{ISLAND}} \sim \text{HalfCauchy}(0, 1)$$

Distribution of island intercepts informs amount of excess variation



	WAIC	pWAIC	dWAIC	weight	SE	dSE
m12.6	70.0	4.9	0.0	1	2.65	NA
m10.12	84.4	3.8	14.5	0	8.94	7.29



Homework

- Frogs & contraception:
12M1, 12M2, 12H1
- Next week: Chapter 13 —
varying slopes and other
wonders

