Bayesian Model Comparison in R

Part II

Julia Haaf

Slides and Material

You can find the slides (and additional materials) here: https://github.com/jstbcs/ws-bayesian-stats-r.

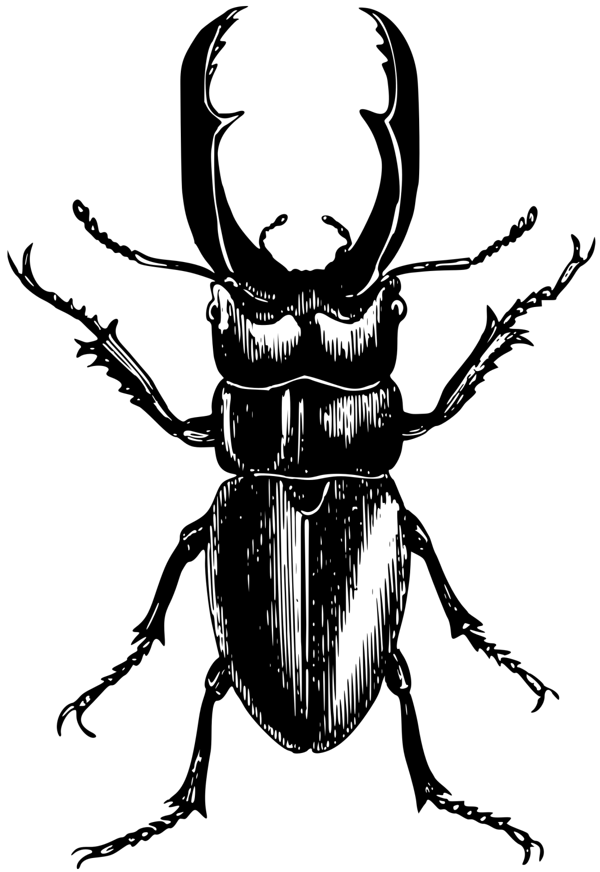
Outline

1. Ordinal constraints
2. Random effects
3. Individual differences and “Does everyone?”

Ordinal constraints

A psychologist's favorite design

With Ugly Bugs



A psychologist's favorite design

Ryan, Wilde, & Crist (2013)

		Disgust	
		Low	High
Fear	Low	Low/Low	Low/High
	High	High/Low	High/High

- **"How willing are you to kill/get rid of this bug?"**

Systems of orders with bugs

M_0 : Null

M_1 : Consistent +

M_2 : + Equality

M_3 : Fear Only

M_4 : Disgust Only

M_5 : Unconstrained

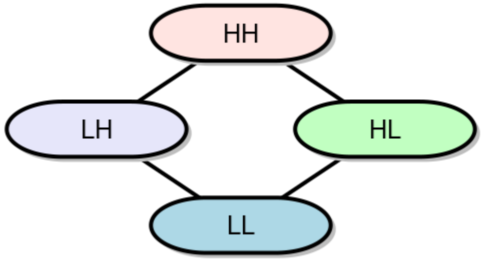
Systems of orders with bugs

M_0 : Null

M_1 : Consistent +

M_2 : + Equality

M_3 : Fear Only



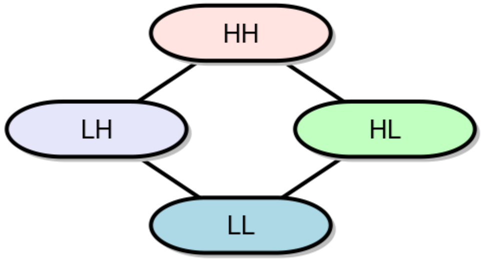
M_4 : Disgust Only

M_5 : Unconstrained

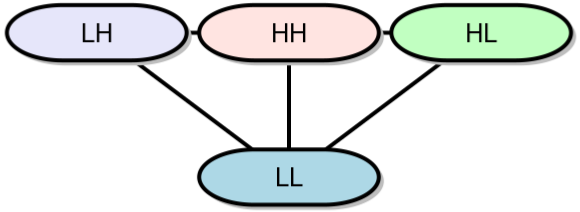
Systems of orders with bugs

M_0 : Null

M_1 : Consistent +



M_2 : + Equality

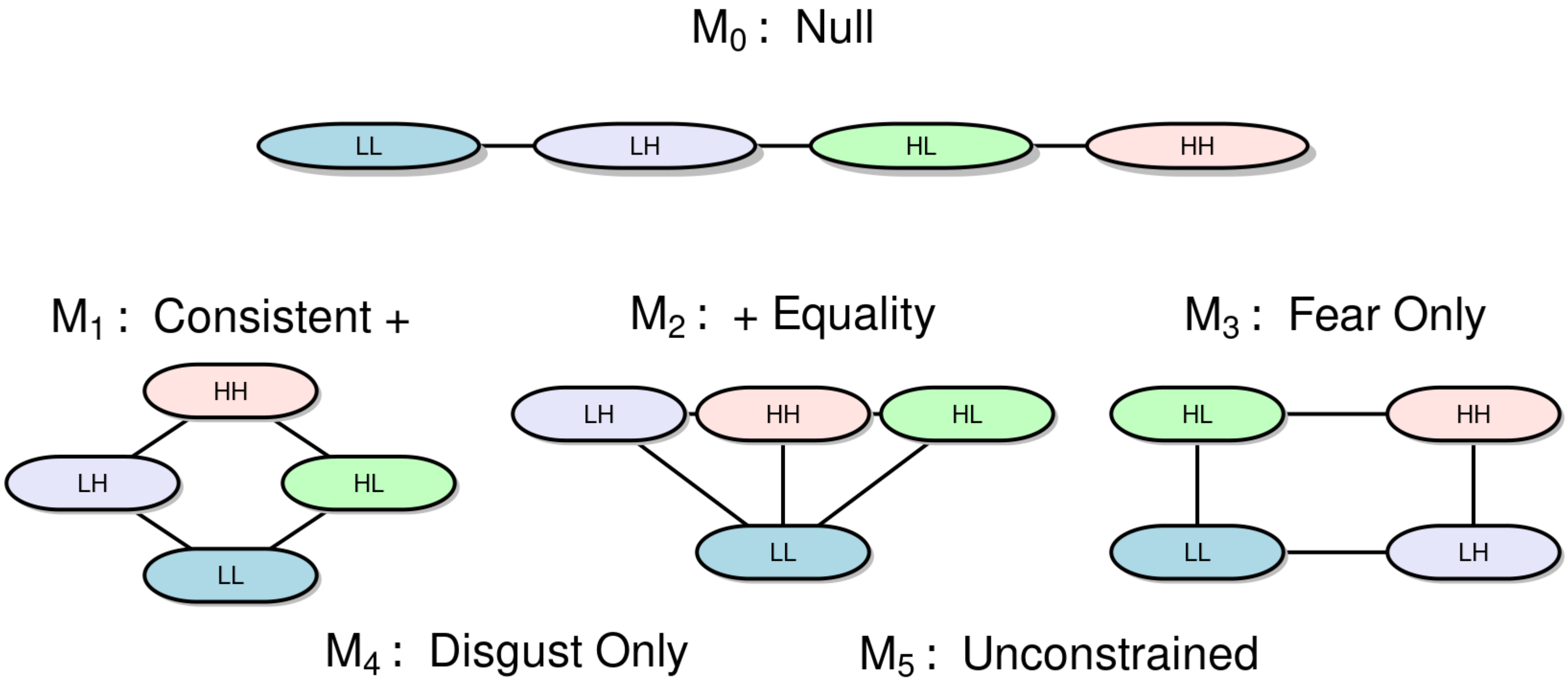


M_3 : Fear Only

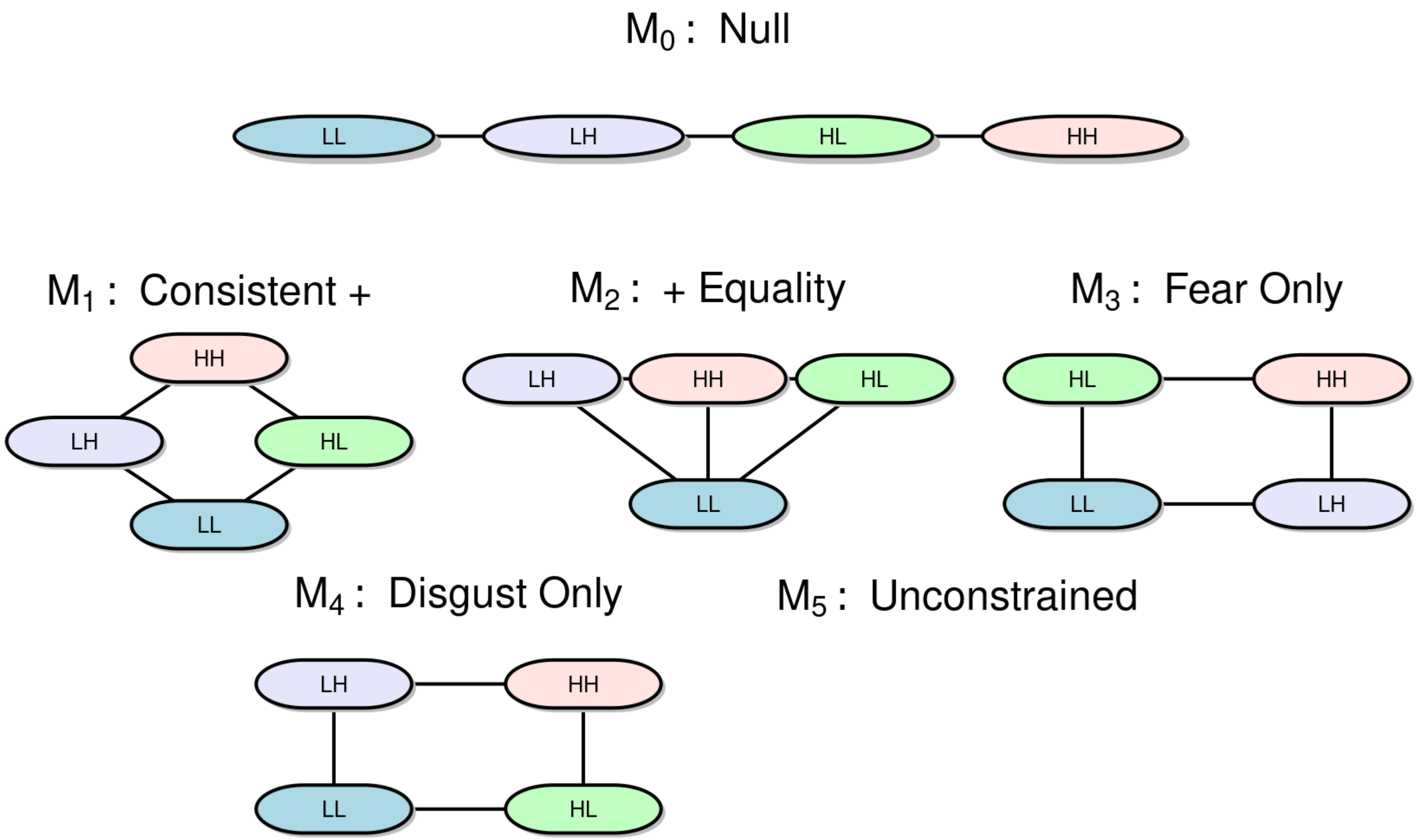
M_4 : Disgust Only

M_5 : Unconstrained

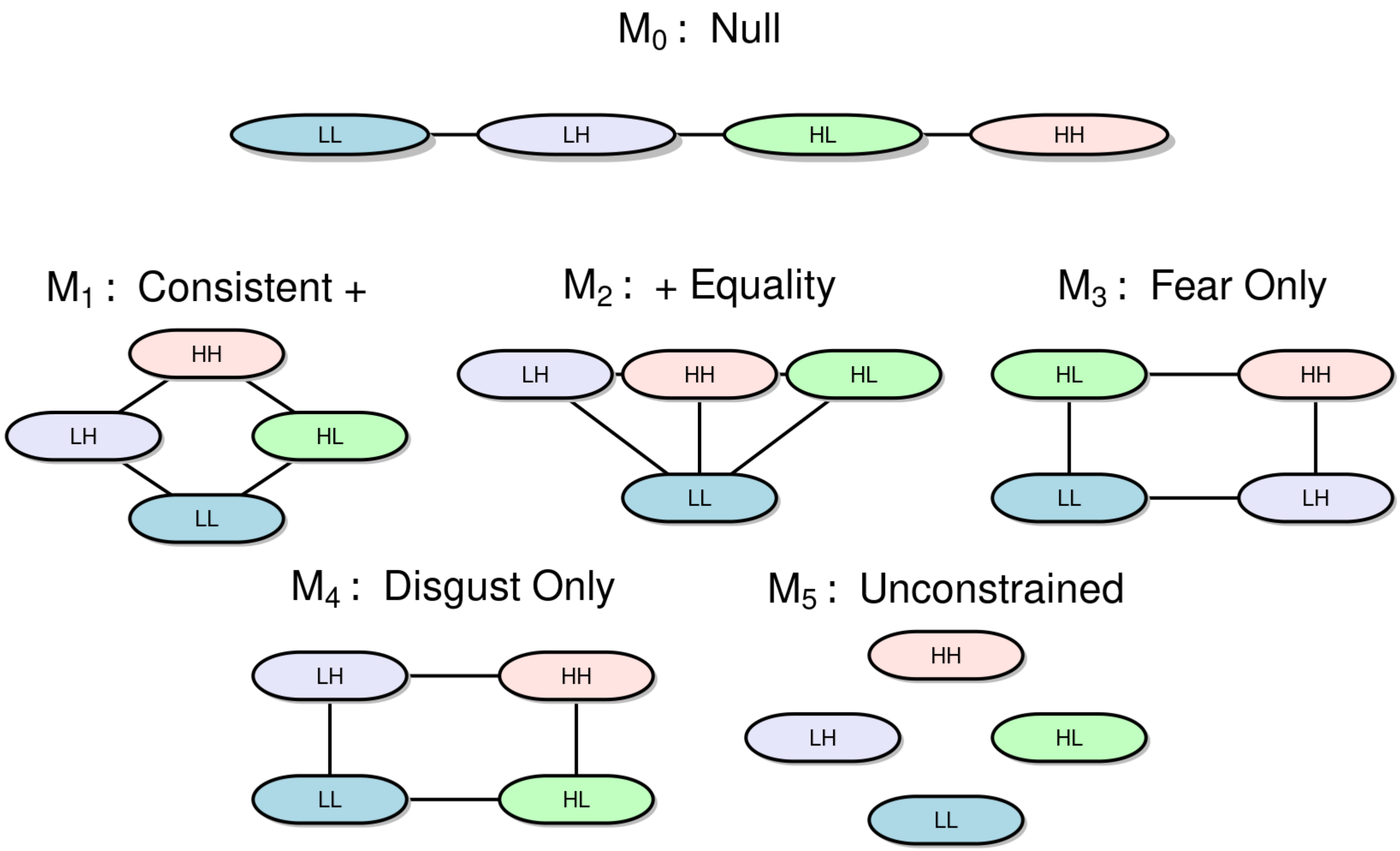
Systems of orders with bugs



Systems of orders with bugs

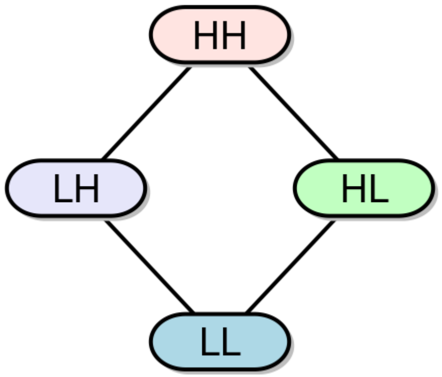


Systems of orders with bugs



Systems of orders plot

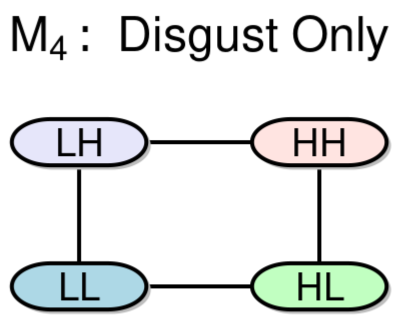
```r
library(diagram)
par(mar = c(0, 0, 0, 0))
names <- c("LL", "LH", "HL", "HH"); o <- 1:4
M <- matrix(nrow = length(o), ncol = length(o), data=0)
M[1, 2] <- ''; M[1, 3] <- ''; M[2, 4] <- ''; M[3, 4] <- ''

plotmat(M, pos = c(1, 2, 1), name = names[c(4, 2, 3, 1)]
        , curve = 0, box.type="round", box.size=.06, box.prop=.8
        , box.col = myCol[c(4, 2, 3, 1)], arr.length=0, box.cex = 1.2
        , relsize = 1, shadow.size = 0.007)
```



Ordinal constraint before seing the data

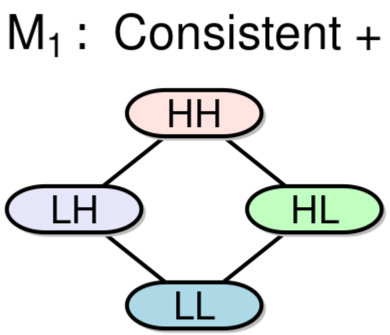
- For a Bayesian analysis with ordinal constraints we need to know the prior probability of the constraint.
- The encompassing model is a model that has all the equality constraints but lets the order free to vary.
- Restricted model space is the proportion of the prior probability space of the encompassing model that is in line with the ordinal constraint.

Let’s start simple



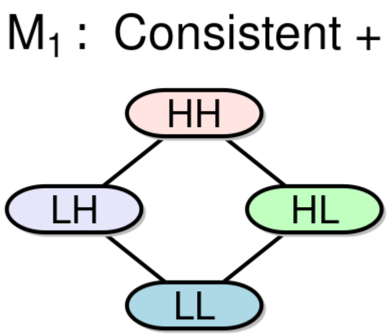
- What is the encompassing model?
- What is prior probability space of the restricted model?

A bit more complicated



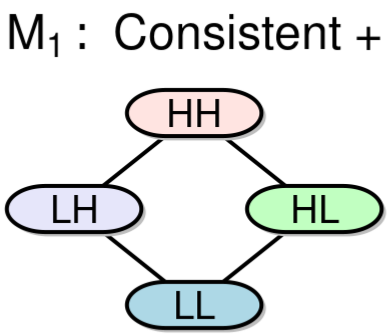
- What is the encompassing model?
- What is prior probability space of the restricted model?

A bit more complicated



Option 1

A bit more complicated

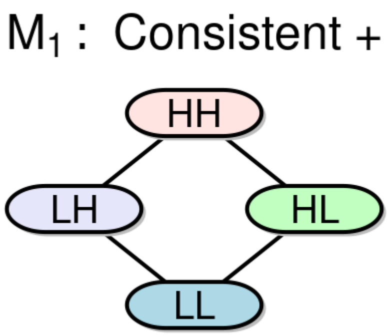


Option 1

```
M <- 100000
LL <- rcauchy(M)
LH <- rcauchy(M)
HL <- rcauchy(M)
HH <- rcauchy(M)

mean(LL < LH & LL < HL &
       LH < HH & HL < HH)
```

```
## [1] 0.0834
```

A bit more complicated



Option 2

$$
\begin{aligned}
HH &= \mu + \theta_1 + \theta_2 + \theta_3 \\
LH &= \mu - \theta_1 + \theta_2 - \theta_3 \\
HL &= \mu + \theta_1 - \theta_2 - \theta_3 \\
LL &= \mu - \theta_1 - \theta_2 + \theta_3
\end{aligned}
$$

A bit more complicated

Option 2

$$\begin{aligned}
HH &= \mu + \theta_1 + \theta_2 + \theta_3 \\
LH &= \mu - \theta_1 + \theta_2 - \theta_3 \\
HL &= \mu + \theta_1 - \theta_2 - \theta_3 \\
LL &= \mu - \theta_1 - \theta_2 + \theta_3
\end{aligned}$$

$$\begin{aligned}
HH - LH > 0 &\Leftrightarrow \theta_1 + \theta_3 > 0 \\
HH - HL > 0 &\Leftrightarrow \theta_2 + \theta_3 > 0 \\
LH - LL > 0 &\Leftrightarrow \theta_2 - \theta_3 > 0 \\
HL - LL > 0 &\Leftrightarrow \theta_1 - \theta_3 > 0
\end{aligned}$$

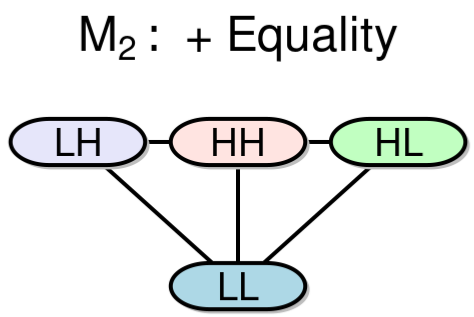
A bit more complicated

$$
\begin{aligned}
HH - LH > 0 &\Leftrightarrow \theta_1 + \theta_3 > 0 \\
HH - HL > 0 &\Leftrightarrow \theta_2 + \theta_3 > 0 \\
LH - LL > 0 &\Leftrightarrow \theta_2 - \theta_3 > 0 \\
HL - LL > 0 &\Leftrightarrow \theta_1 - \theta_3 > 0
\end{aligned}
$$

```
M <- 100000
theta1 <- rcauchy(M)
theta2 <- rcauchy(M)
theta3 <- rcauchy(M)

mean(theta1 + theta3 > 0 &
       theta2 + theta3 > 0 &
       theta2 - theta3 > 0 &
       theta1 - theta3 > 0)
```

```
## [1] 0.08201
```

Aaand a bit more complicated



- What is the encompassing model?
- Reparameterize the factors.
- New factor: *high* for HH, HL, and LH, and *low* for LL.
- What is prior probability space of the restricted model?
- That's right, it is 0.5.

Your turn

Check out worksheet 1 on github!

Random effects

Random effects

- All effects we talked about so far are fixed effects (sum-to-zero constraint).

Random effects

- All effects we talked about so far are fixed effects (sum-to-zero constraint).

```
c(est.disgust[1, "disgust-0"], est.disgust[1, "disgust-1"])
```

```
##  disgust-0  disgust-1
## -0.3538381  0.3538381
```

- Random effects have a different constraint:
- $$\theta \sim \text{Normal}(0, \sigma^2).$$

Random effects

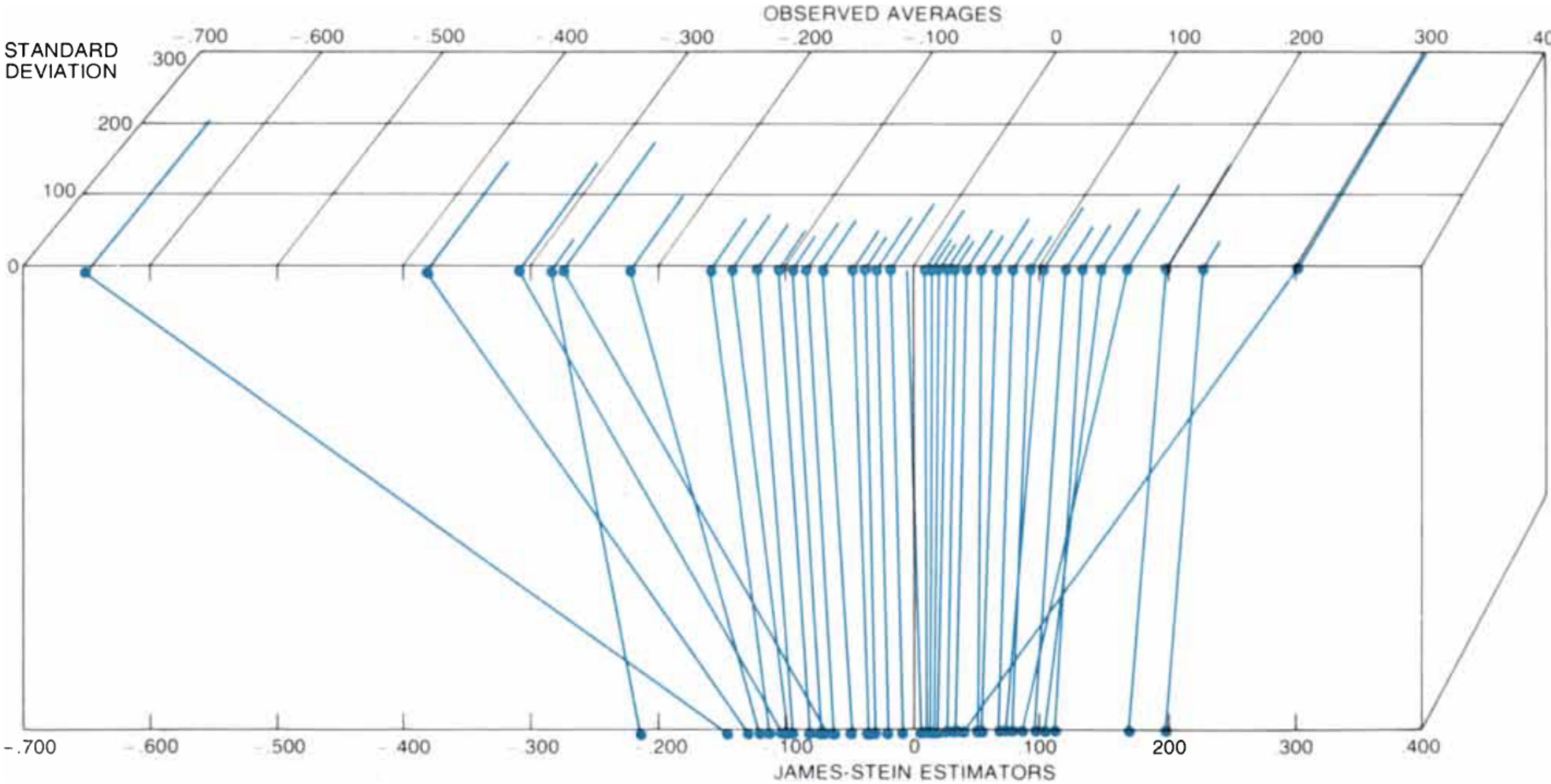
Philosophy

Gelman (2005):

1. Fixed effects are constant across individuals, and random effects vary. For example, in a growth study, a model with random intercepts α_i and fixed slope β corresponds to parallel lines for different individuals i, or the model $y_{it} = \alpha_i + \beta t$. Kreft and de Leeuw [(1998), page 12] thus distinguish between fixed and random coefficients.
2. Effects are fixed if they are interesting in themselves or random if there is interest in the underlying population. Searle, Casella and McCulloch [(1992), Section 1.4] explore this distinction in depth.
3. "When a sample exhausts the population, the corresponding variable is *fixed*; when the sample is a small (i.e., negligible) part of the population the corresponding variable is *random*" [Green and Tukey (1960)].
4. "If an effect is assumed to be a realized value of a random variable, it is called a random effect" [LaMotte (1983)].
5. Fixed effects are estimated using least squares (or, more generally, maximum likelihood) and random effects are estimated with shrinkage ["linear unbiased prediction" in the terminology of Robinson (1991)]. This definition is standard in the multilevel modeling literature [see, e.g., Snijders and Bosker (1999), Section 4.2] and in econometrics.

Random effects

Philosophy



Efron & Morris (1977)

Random effects

Philosophy

- In frequentist statistics: Data are observed and follow probability distributions.
- Parameters are not observed and are fixed (not distributed), unknown.
- Random effects are neither parameters nor data, they are distributed yet not observed.
- Random effects as latent variables: governed by distributions, but unlike data, they are not observed.
- In Bayesian statistics: Both parameters and data are governed by distributions (but only data are observed), and so are random effects, so there is not need for a third category.

Random effects

Practice

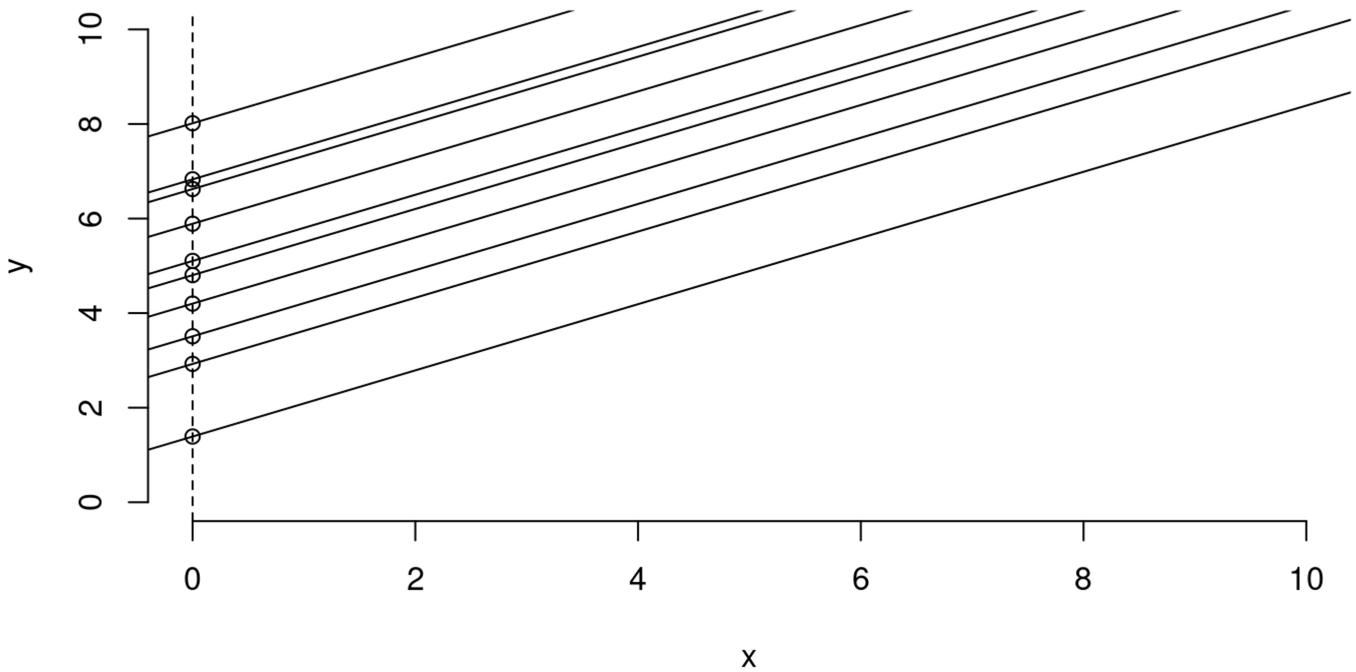
> When a sample exhausts the population, the corresponding variable is fixed; when the sample is a small part of the population the corresponding variable is random (Green & Tukey, 1960).

- Fixed effects if you deliberately planned the levels (e.g., experimental manipulations).
- Random effects if the levels represent a sample of a larger potential population (e.g., people or items).

Random effects

Practice

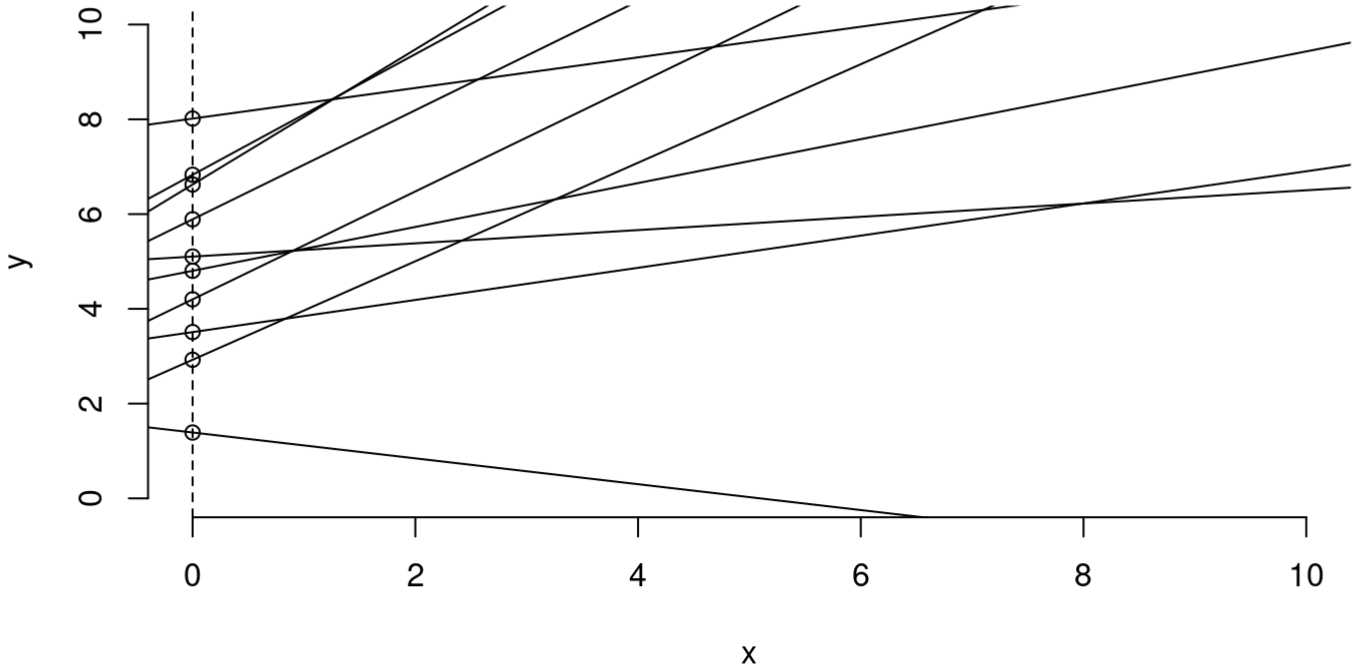
- Random intercepts



Random effects

Practice

- Random intercepts
- Random slopes



Random effects

in `BayesFactor`

- `whichRandom` is the important argument.

Random effects

in BayesFactor

- whichRandom is the important argument.

```
mod.gen <- BayesFactor::lmBF(value ~ disgust + fear + disgust:fear + Subject
                             , data = datl
                             , whichRandom = "Subject"
                             , rscaleEffects = c("disgust" = 1/2
                                                 , "fear" = 1/2
                                                 , "disgust:fear" = 1/3
                                                 , "Sub" = 1/4))
```

Random effects

in `BayesFactor`

- When `BayesFactor` was developed they were pretty confident about random intercepts, but a bit hesitant about random slopes.
- Very little documentation, but can be used.

Random effects

in BayesFactor

- When BayesFactor was developed they were pretty confident about random intercepts, but a bit hesitant about random slopes.
- Very little documentation, but can be used.

```
## Bayes factor analysis
## --------------
## [1] cond                    : 28348.34     ±0%
## [2] sub                     : 1.416292e+55 ±0.01%
## [3] cond + sub              : 7.764012e+59 ±1.46%
## [4] cond + sub + cond:sub : 7.149918e+75 ±1.01%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

Random effects

Priors for random effects

Let’s start with the model setup again.

Random effects

Priors for random effects

Let's start with the model setup again.

$$Y_{ijk} \sim \text{Normal}(\mu + \alpha_i + x_j\theta, \sigma^2)$$

Random effects

Priors for random effects

Let's start with the model setup again.

$$Y_{ijk} \sim \text{Normal}(\mu + \alpha_i + x_j\theta, \sigma^2).$$

Random intercept α_i has a distribution (not the prior):

$$\alpha_i \sim \text{Normal}(0, g_\alpha \sigma^2).$$

Random effects

Priors for random effects

Let's start with the model setup again.

$$Y_{ijk} \sim \text{Normal}(\mu + \alpha_i + x_j\theta, \sigma^2).$$

Random intercept α_i has a distribution (not the prior):

$$\alpha_i \sim \text{Normal}(0, g_\alpha \sigma^2).$$

Prior on g_α, the variance scaling factor:

$$g_\alpha \sim \text{Inverse-}\chi^2(r_\alpha).$$

Random effects

Priors for random effects

Three observations:

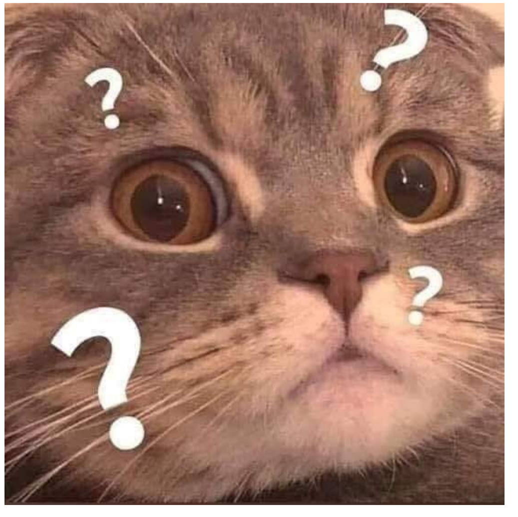
- For random effects it becomes quite difficult to distinguish between the prior and the likelihood.
- Setting the scale for random effects is a bit more tricky than in the t-test.
- If you use `posterior()` and look at the estimates the last columns are always “g”. Now you know why. :)

Revisiting prior scales

- For non-nested data scales can be interpreted as expected effect size.
- For nested data it is a bit more complicated.
- Take Stroop.
- 2 conditions, I people, K trials.
- Typical effect size: $\frac{M_d}{SD_d}$, where $SD_d = \sqrt{\frac{\sum(d_i - M_d)}{I}}$.
- In `BayesFactor`: $\frac{\theta}{\sigma}$
- For the model including random effects σ can be interpreted as trial-by-trial variance, not between-participant variance.

Revisiting prior scales

1. How much do I expect the data to vary from trial to trial (within a person)?
2. How big is the expected effect relative to that variability (fixed effects)?
3. How much do I think people will vary relative to that variability (random effects)?



Your turn

Check out worksheet 2 on github!

Individual Differences

Individual Differences

- People vary. Duh!
- Sometimes more and sometimes less relevant for us.
- Memory: We (often) care more about individual variability of memory strength than item fit.
- Attitudes: We care about how people are persuaded to change their attitudes, (often) not how preferences differ to begin with.

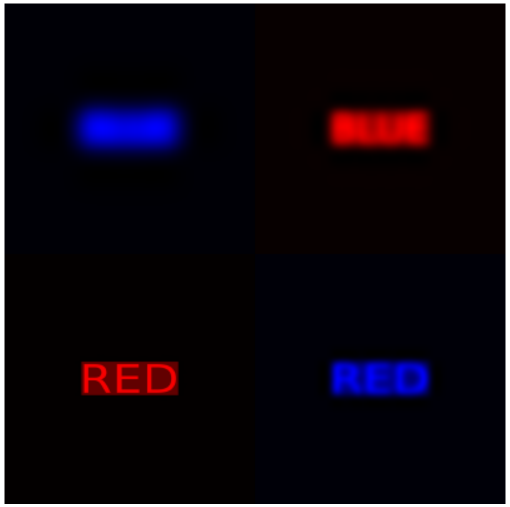
Qualitative Individual Differences

1. QID are defined by research question and experimental design.
2. QID point to differences in cognitive processing (not physical impairment or experimental manipulation).
3. QID are stable: Should be identifiable after sample noise is taken into account.

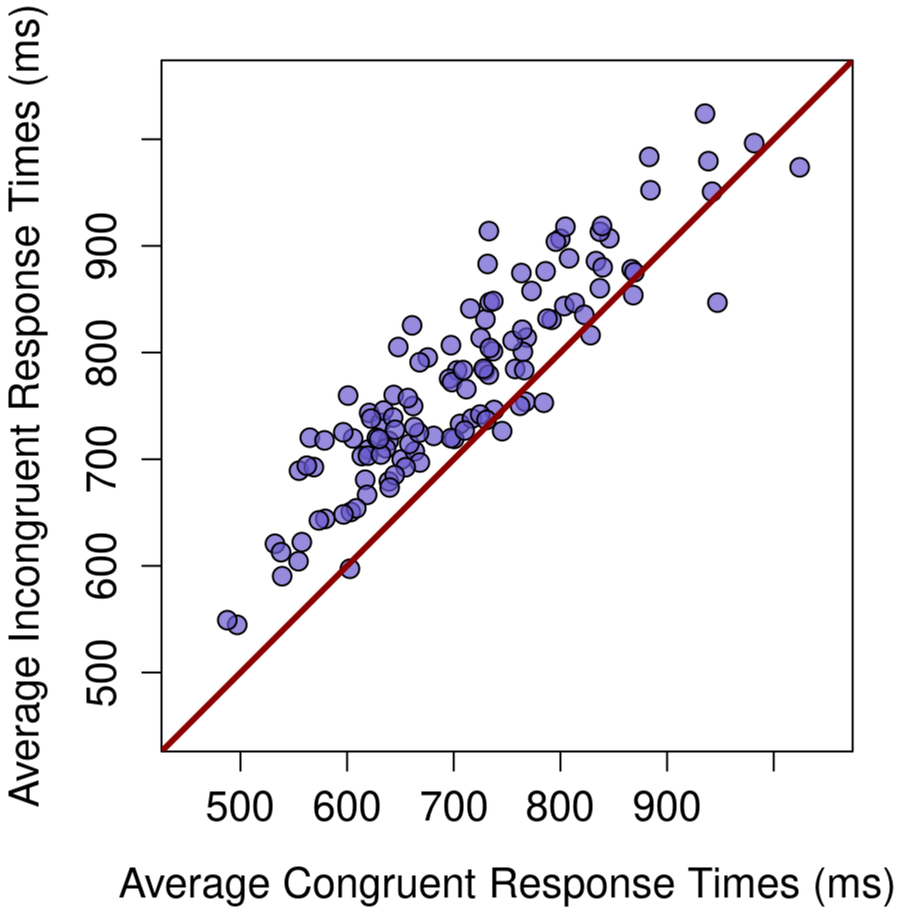
Example: Stroop Effect

Theoretical Statement

- Colors from congruent items are faster identified than colors from incongruent items.
- Qualitative individual differences should reflect ability to inhibit automatic reading.

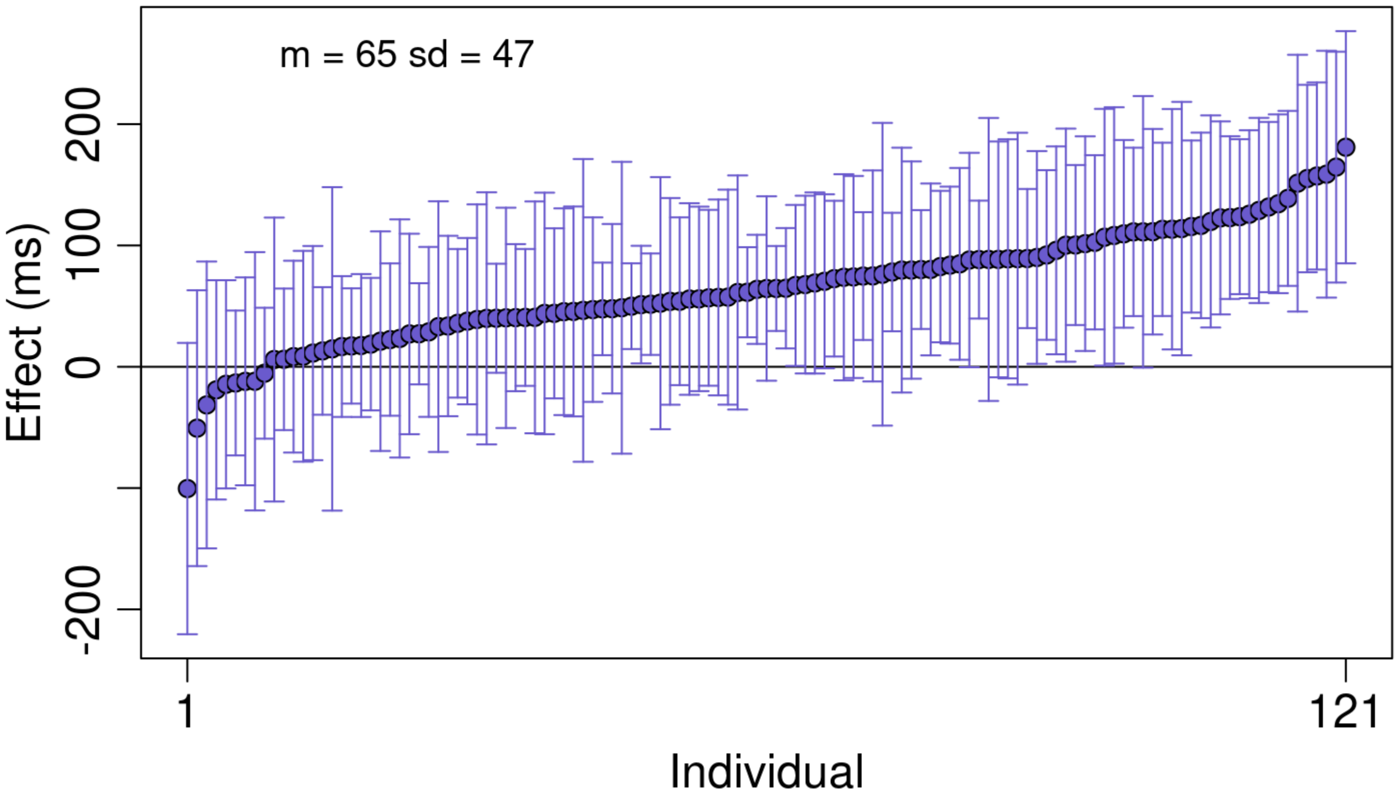


Stroop Data

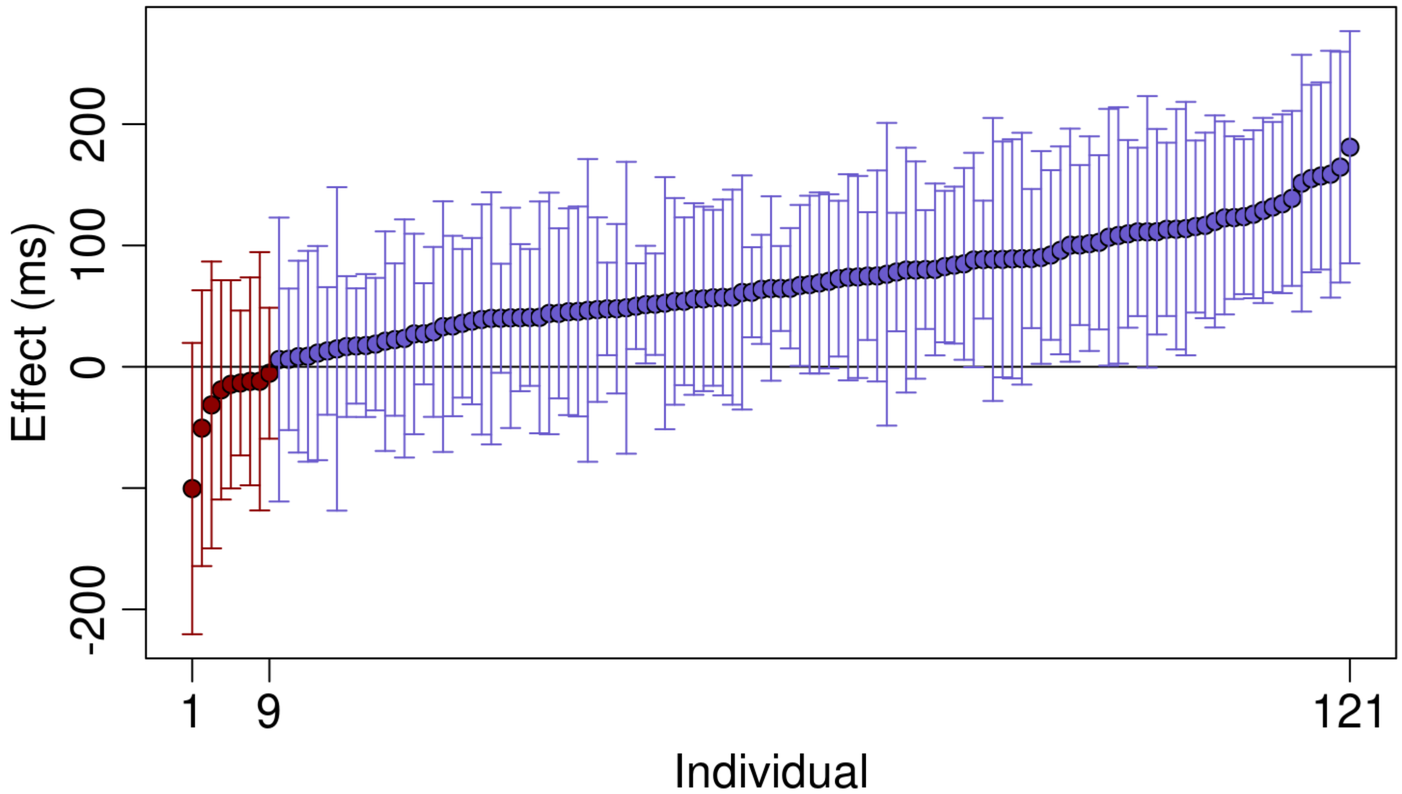


Von Bastian, Souza, & Gade (2015)

Stroop Effects



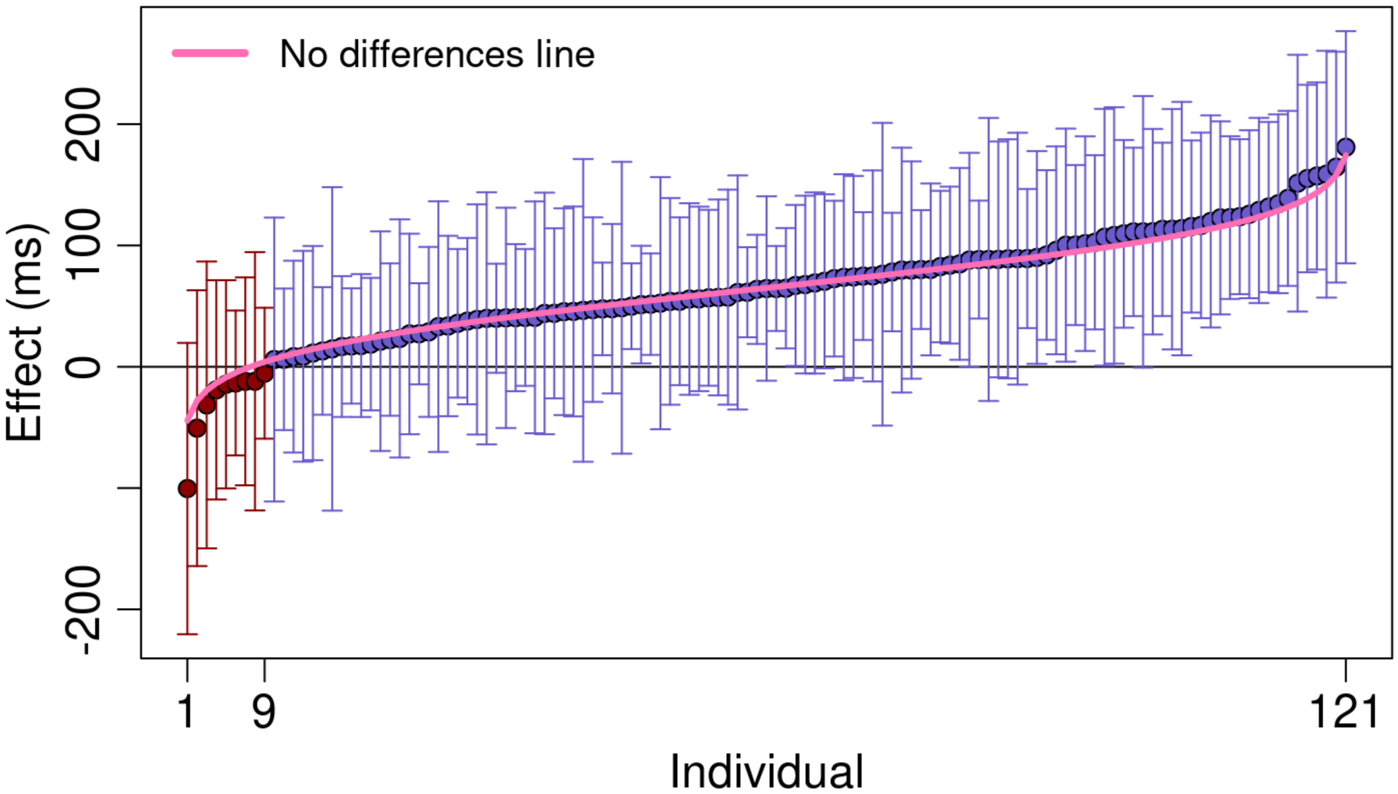
Qualitative Differences



Research Questions

- **Are there qualitative individual differences of the Stroop effect?**
- **Are there qualitative individual differences after sample noise is taken into account, or does everyone show a Stroop effect?**

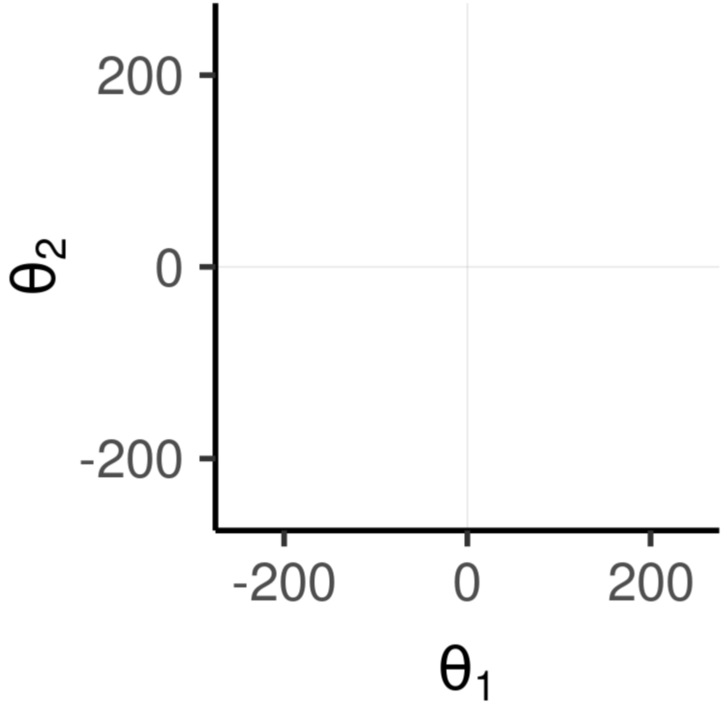
Stable Differences



Models

- Participants: $i = 1, \ldots, I$
- Condition: $j = 1, 2$ (congruent, incongruent)
- Trials: $k = 1, \ldots, K_{ij}$
- $$Y_{ijk} \sim \text{Normal}(\alpha_i + x_j\theta_i, \sigma^2)$$
- θ_i is the effect

Models on true effects θ_i

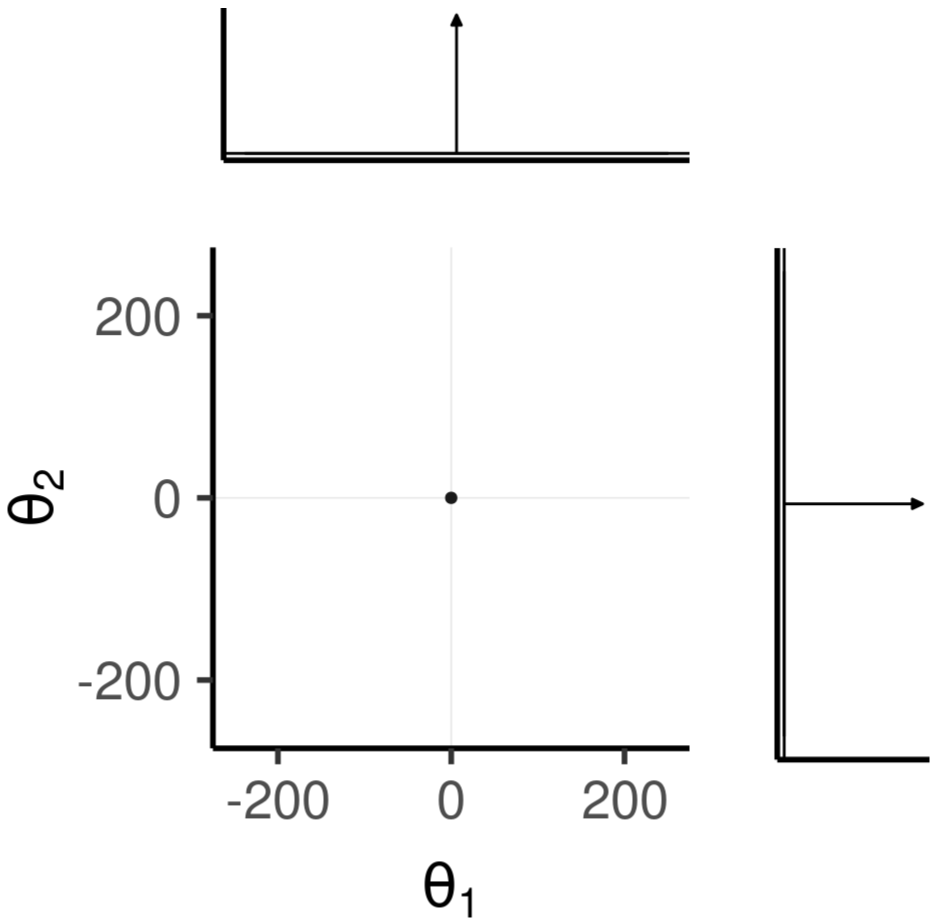


Haaf (2018); Haaf & Rouder (2017)

The Null Model

$$\theta_i = 0$$

The Null Model

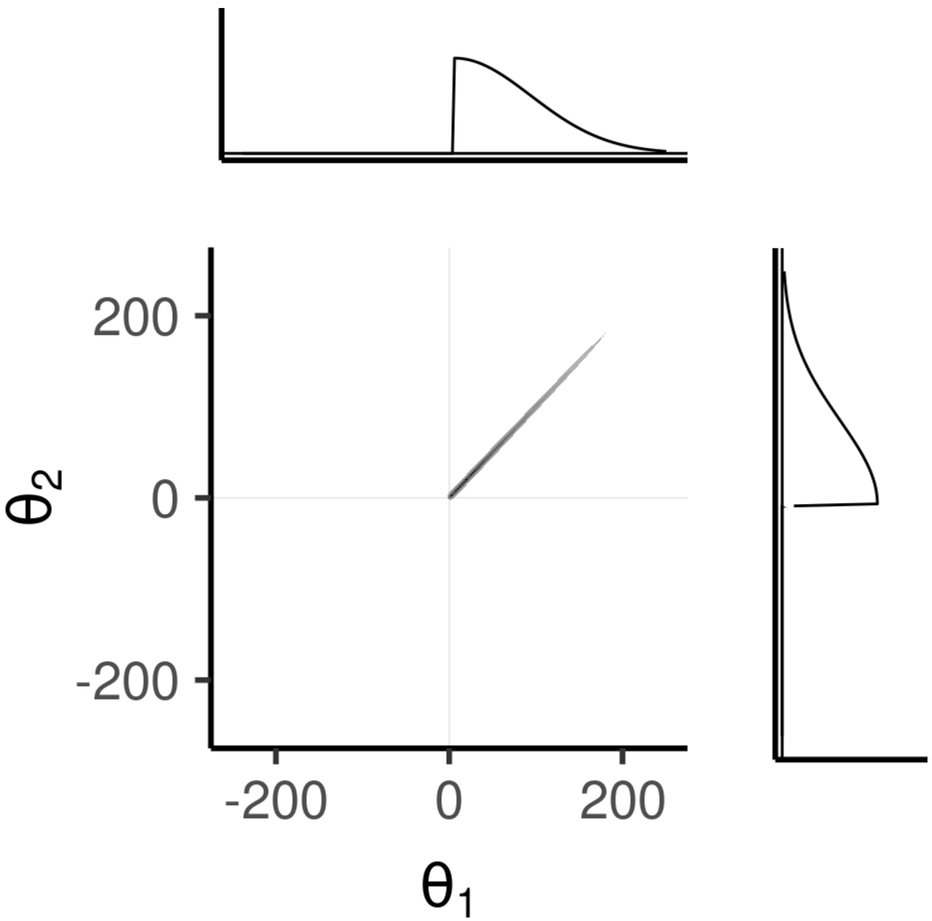


The Common-Effect Model

$$\theta_i = \nu$$

$$\nu \sim \text{Truncated-Normal}(0, \eta^2)$$

The Same-Effect Model

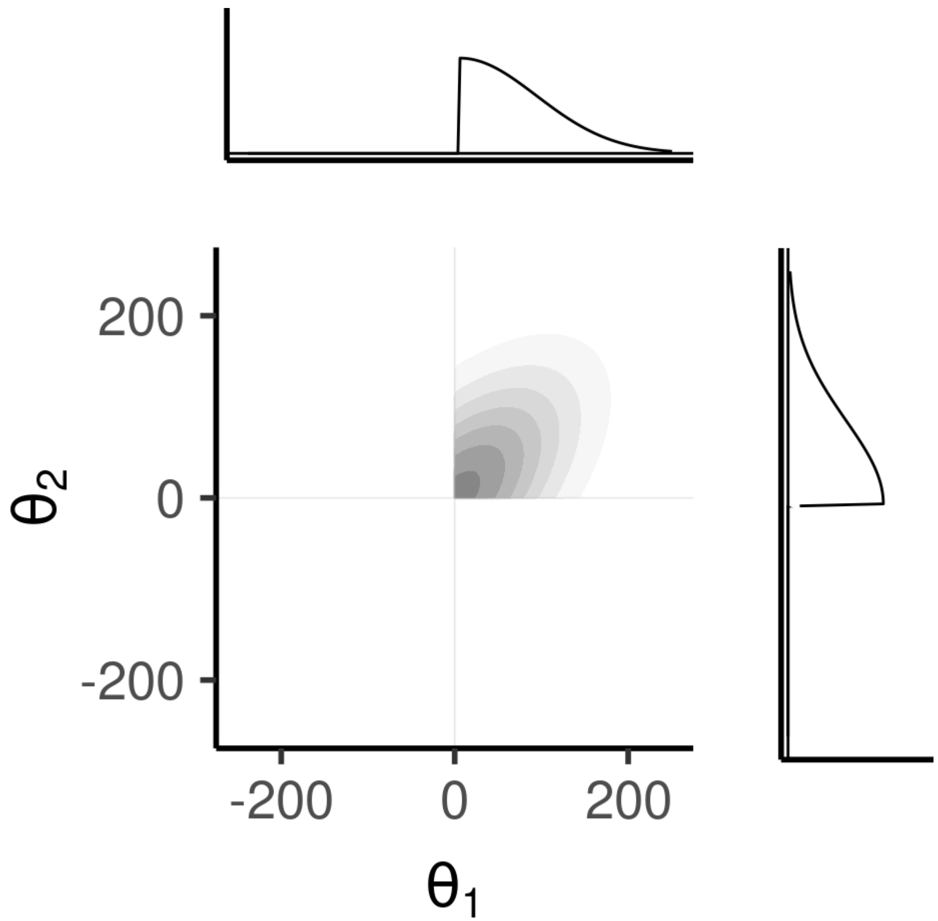


The Positive-Effects Model

$$\theta_i \sim \text{Truncated-Normal}(\nu, \tau^2)$$

$$\nu \sim \text{Truncated-Normal}(0, \eta^2)$$

The Positive-Effects Model

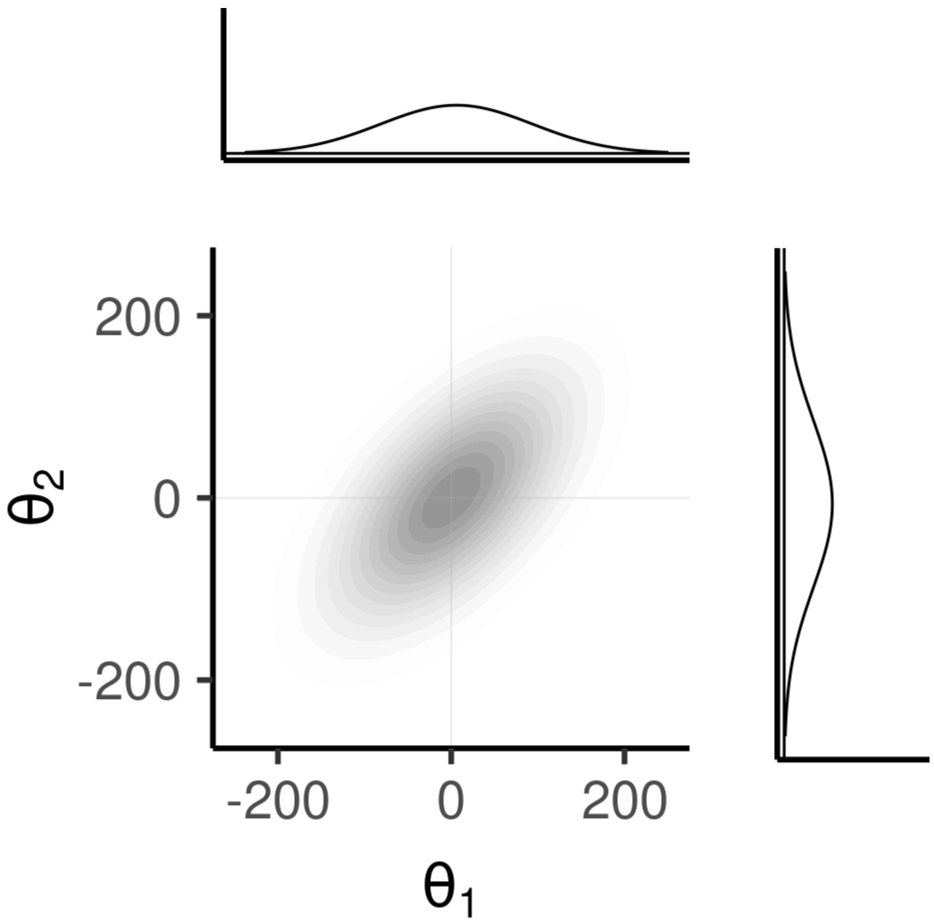


The Unconstrained Model

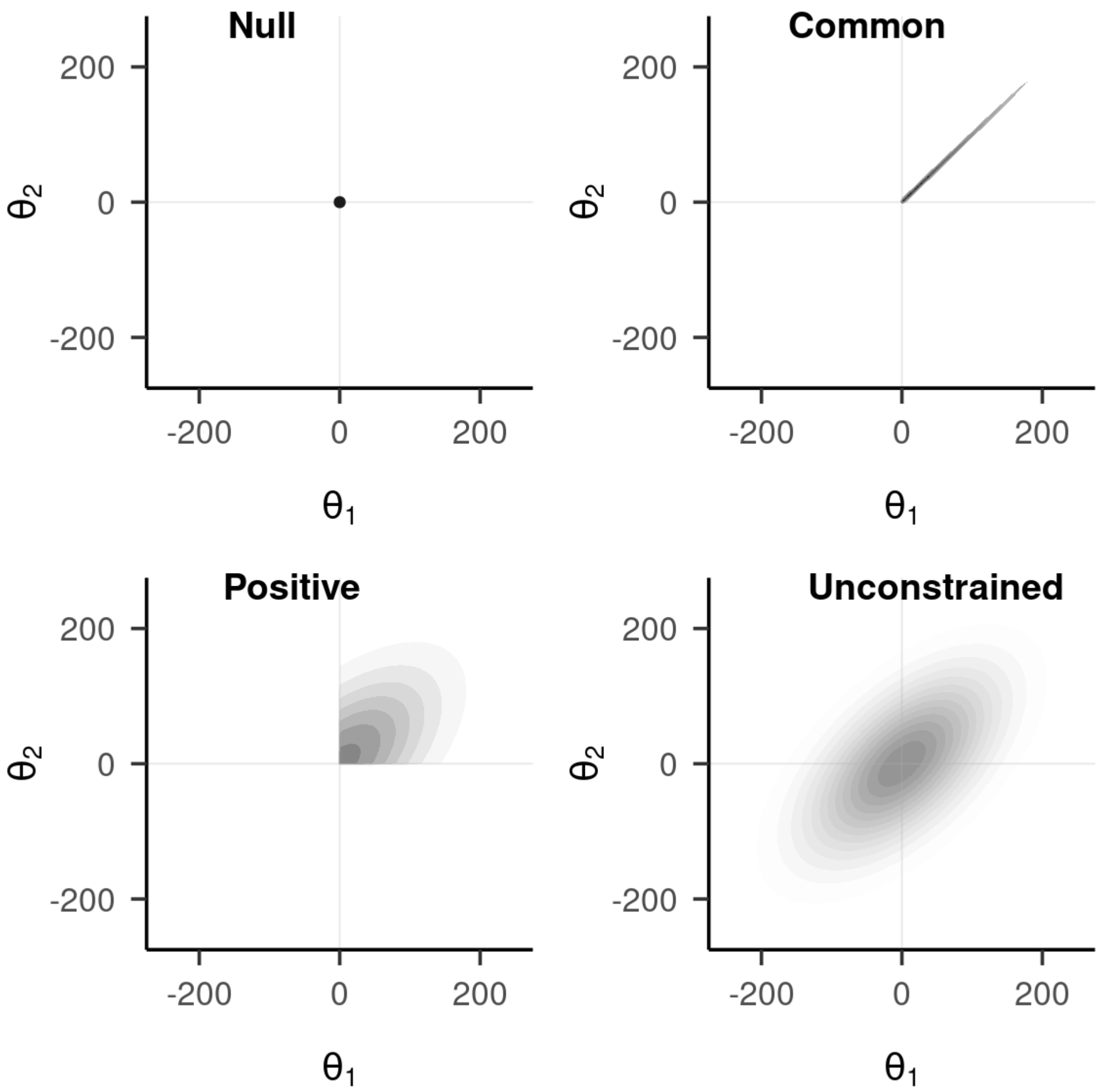
$$\theta_i \sim \text{Normal}(\nu, \tau^2)$$

$$\nu \sim \text{Normal}(0, \eta^2)$$

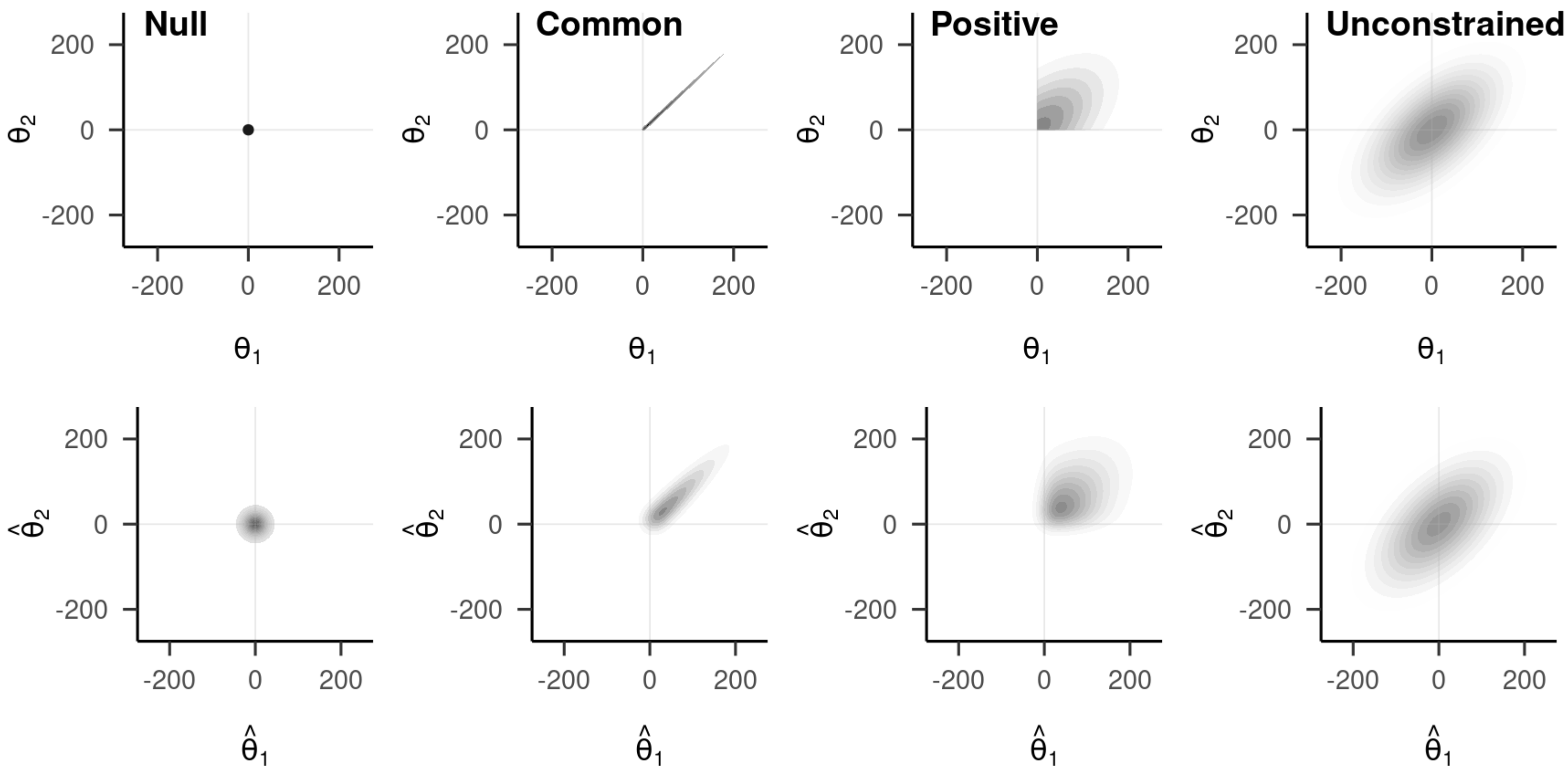
The Unconstrained Model



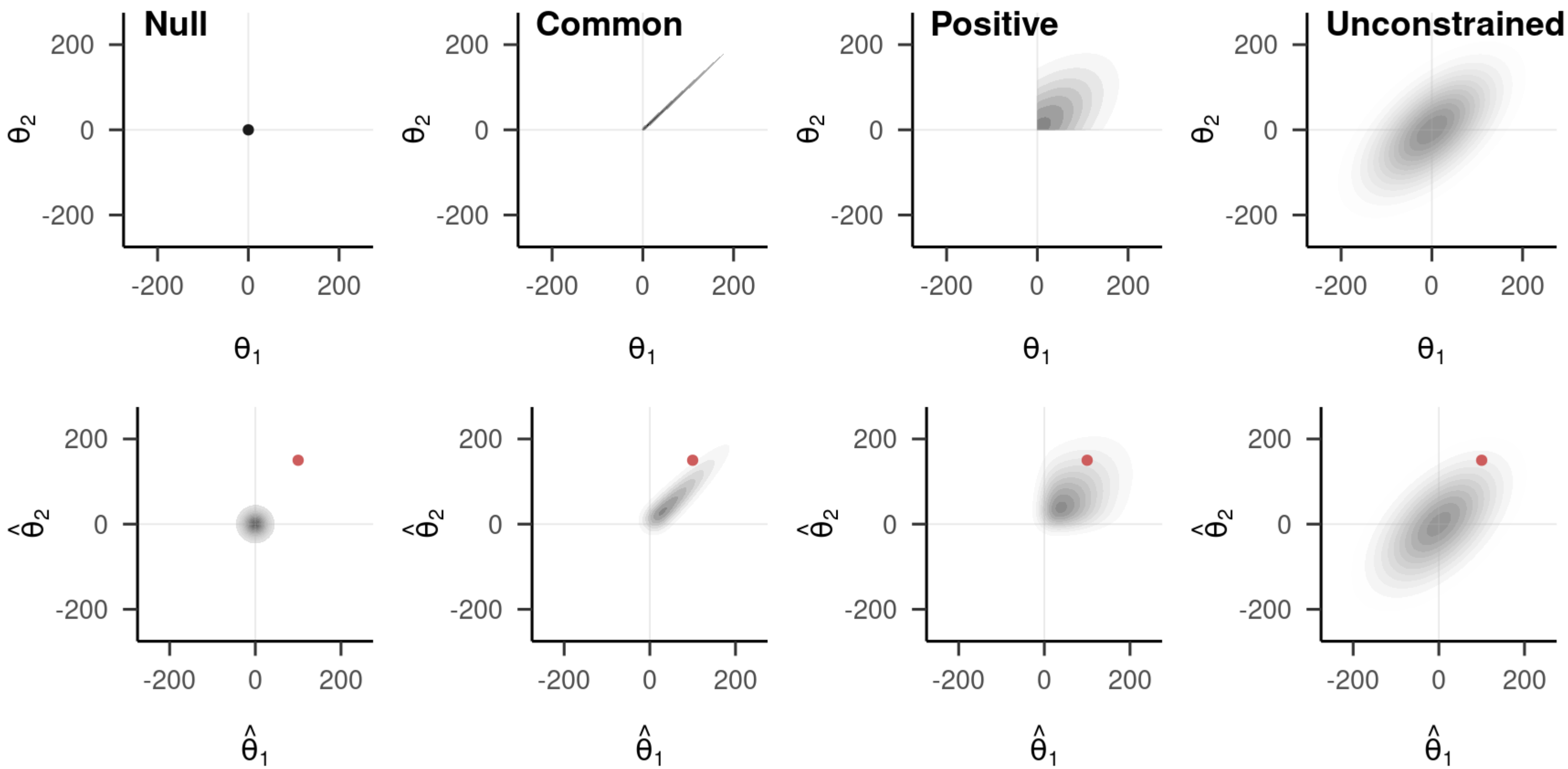
From Models...



From Models… to Predictions



From Models… to Predictions… to Evidence

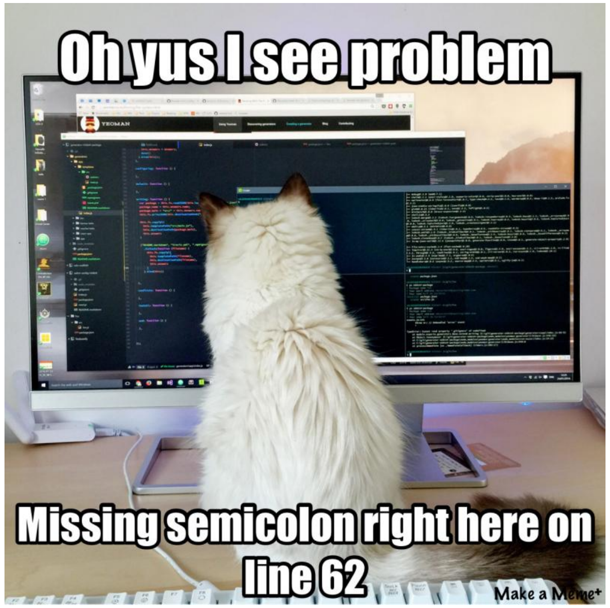


Back to Stroop

In worksheet 3. :)

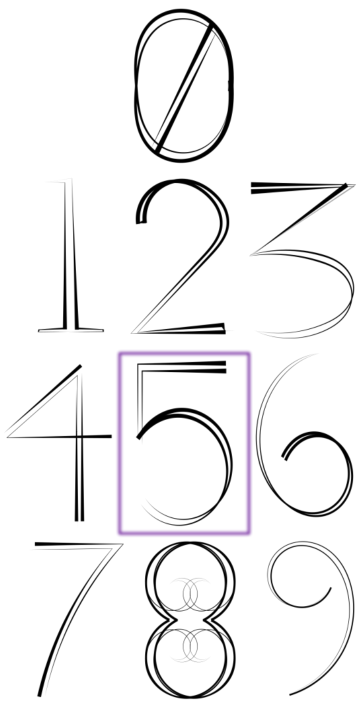
“Does everyone” analysis in R

- Thesis student Lukas Klima has developed an R package: https://github.com/lukasklima/quid.
- Backend is BayesFactor.
- New argument `whichConstraints` specifies the ordinal constraints tested.



Example: Distance from 5

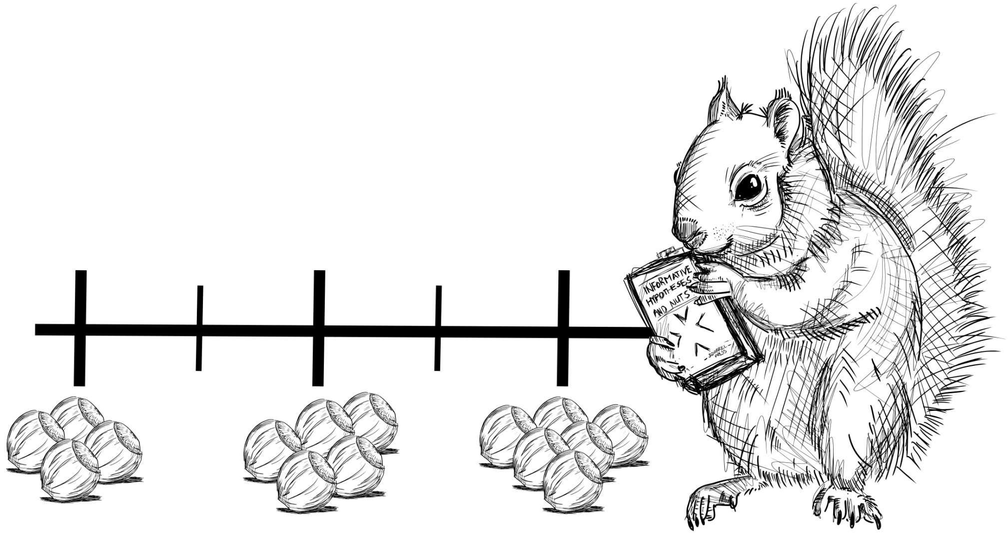
How do we represent numbers internally?



How do we represent numbers internally?

Theoretical Positions

1. Everyones uses analog representation.



How do we represent numbers internally?

Theoretical Positions

1. Everyones uses analog representation.
2. Everyone uses propositional representation.

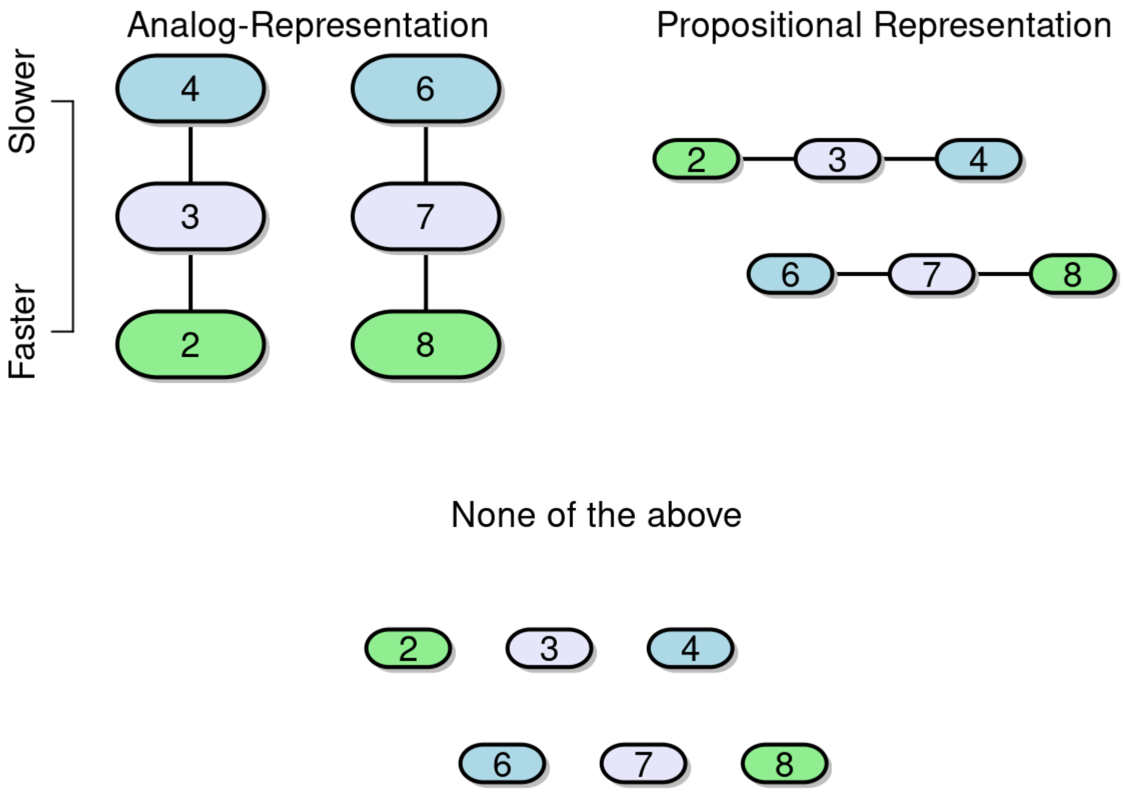
 $<$  $=$ true 

How do we represent numbers internally?

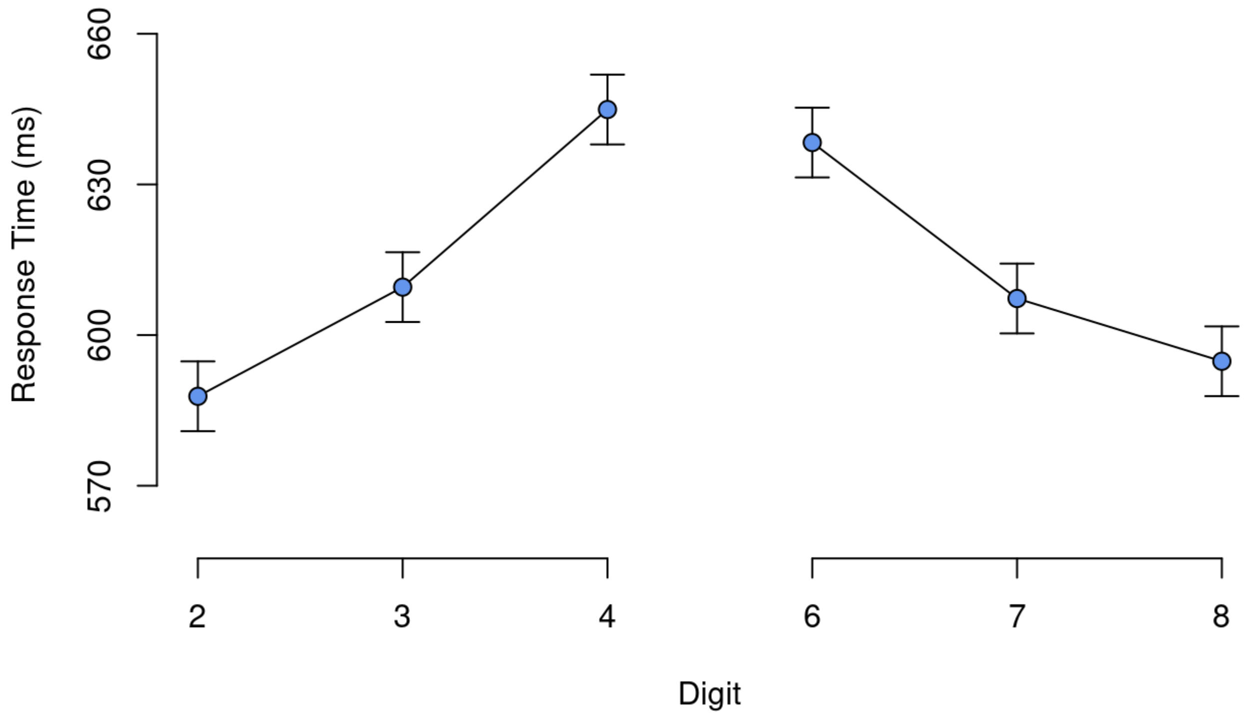
Theoretical Positions

1. Everyones uses analog representation.
2. Everyone uses propositional representation.
3. None of the above.

Symbolic Distance

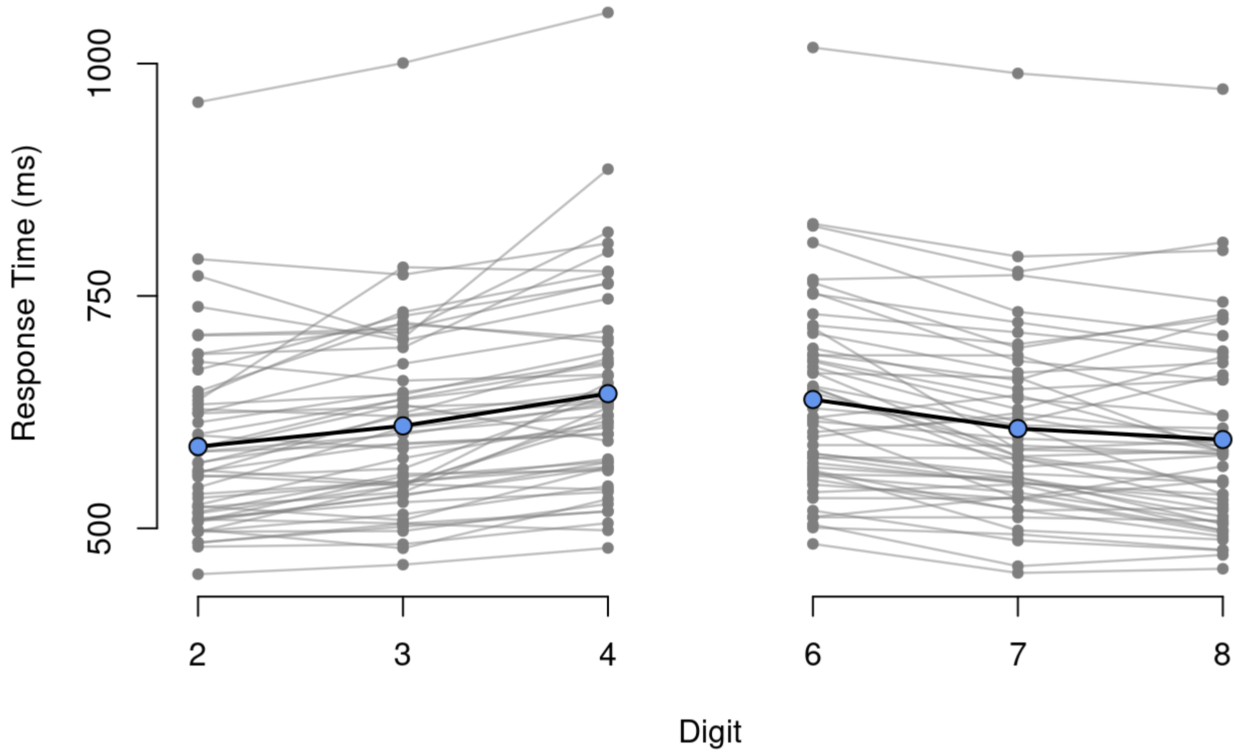


Data



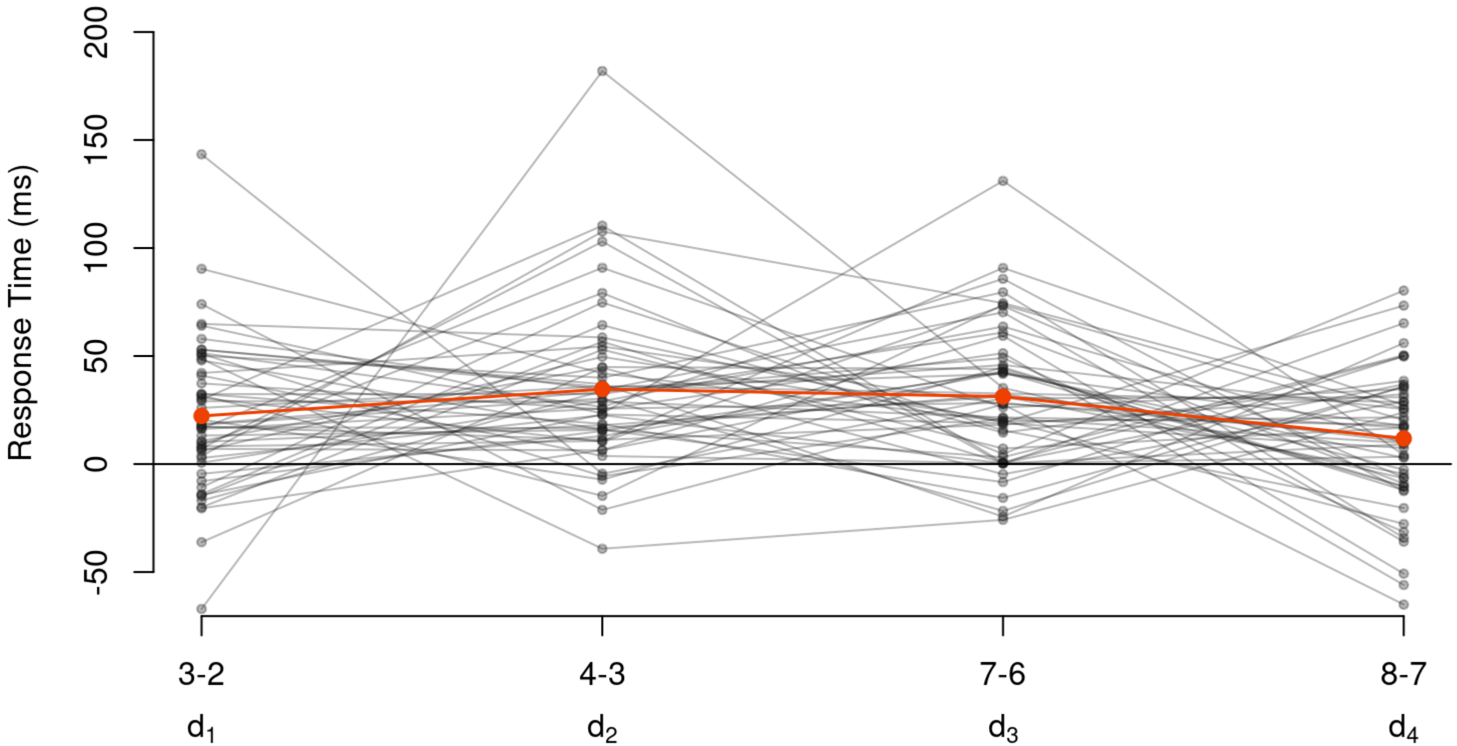
Rouder, Lu, Speckman, Sun, & Jiang (2005)

Individual differences



Rouder, Lu, Speckman, Sun, & Jiang (2005)

Individual differences



Rouder, Lu, Speckman, Sun, & Jiang (2005)

Does everyone in R

Install and load `quid`.

```
# load libraries
library(devtools)
install_github("lukasklima/quid")
library(quid)
```

```
##
## Attaching package: 'quid'
```

```
## The following object is masked _by_ '.GlobalEnv':
##
##     stroop
```

Does everyone in R

Documentation is still under development...

```r
# check what arguments the function takes
args(quid::constraintBF)
```

```
## function (formula, data, whichRandom, ID, whichConstraint, rscaleEffects,
##     iterationsPosterior = 10000, iterationsPrior = iterationsPosterior *
##         10, burnin = 1000, ...)
## NULL
```

```r
## LD5
# inspect data
str(ld5)
```

```
## 'data.frame':    17031 obs. of  9 variables:
##  $ sub     : Factor w/ 52 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ block   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ trial   : int  20 21 22 23 24 26 27 28 29 30 ...
##  $ stim    : int  5 1 3 4 1 5 0 4 4 0 ...
##  $ resp    : int  1 0 1 1 0 1 0 1 1 0 ...
##  $ rt      : int  470 476 507 603 493 535 463 431 569 509 ...
```

Does everyone in R

```r
# analysis
resLD5 <- quid:::constraintBF(formula = rt ~ sub * distance + side,
                              data = ld5,
                              whichRandom = c("sub"),
                              ID = "sub",
                              whichConstraint = c(distance = "1 > 2", distance = "2>3"))
```

Does everyone in R

Priors

```
# analysis
resLD5 <- quid::constraintBF(formula = rt ~ sub * distance + side,
                               data = ld5,
                               whichRandom = c("sub"),
                               ID = "sub",
                               whichConstraint = c(distance = "1 > 2", distance = "2>3"),
                               rscaleEffects = c("sub" = 1,
                                                 "side" = 1/6,
                                                 "distance" = 1/6,
                                                 "sub:distance" = 1/10))
```

Does everyone in R

BF for the analog representation model:

```
resLD5
```

```
##
## Bayes factor analysis
## --------------
## [1] sub                                 : 4.36765e+1044  ±0.01%
## [2] distance                            : 1.523779e+41   ±0.01%
## [3] sub + distance                      : 1.439468e+1100 ±0.7%
## [4] sub + distance + sub:distance       : 1.340429e+1100 ±0.66%
## [5] side                                : 0.052699       ±0%
## [6] sub + side                          : 2.340764e+1043 ±1.62%
## [7] distance + side                     : 7.511926e+39   ±0.93%
## [8] sub + distance + side               : 7.387908e+1098 ±1.35%
## [9] sub + distance + sub:distance + side : 7.337836e+1098 ±2.13%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
##
```

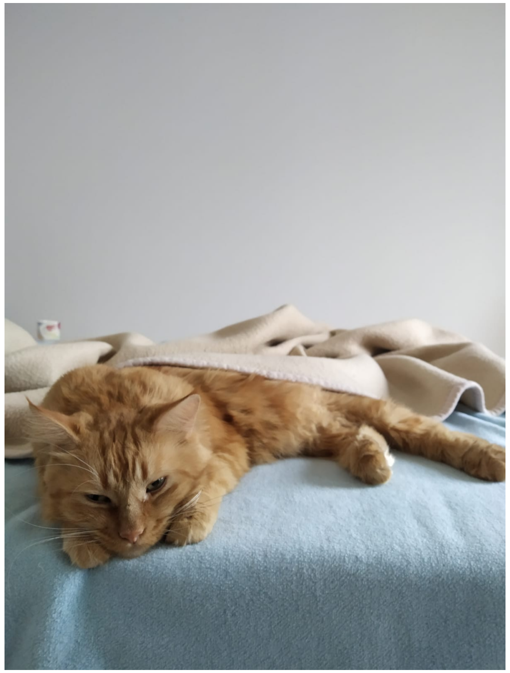
Your turn

Check out worksheet 3 on github!

Bringing it all together

- Ordinal constraints help match theoretical positions to statistical analysis.
- Both on the aggregate level and on the individual level.
- Random effects can be used to control for person/item variability but also to answer substantive questions.

Thank you!



Haaf, J. M. (2018). *A hierarchical Bayesian analysis of multiple order constraints in behavioral science* (PhD thesis). University of Missouri.

Haaf, J. M., & Rouder, J. N. (2017). Developing constraint in Bayesian mixed models. *Psychological Methods*, *22*(4), 779–798.

Rouder, J. N., Lu, J., Speckman, P. L., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin and Review*, *12*, 195–223.

Von Bastian, C. C., Souza, A. S., & Gade, M. (2015). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General*, *145*(2), 246–258.