University of Amsterdam

Does Everyone?

Conceptualizing Individual Variability in Inhibition Tasks

Julia Haaf

April 29th, 2019

Individual Differences

Individual Differences

- People vary. Duh!

Individual Differences

- People vary. Duh!
- Sometimes more and sometimes less relevant for us

Individual Differences

- People vary. Duh!
- Sometimes more and sometimes less relevant for us
- Memory: We care about individual variability of memory strength more than item fit.

Individual Differences

- People vary. Duh!
- Sometimes more and sometimes less relevant for us
- Memory: We care about individual variability of memory strength more 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

Qualitative Individual Differences

1. QID are defined by research question and experimental design
2. QID point to differences in cognitive processing (not physical impairment nor experimental manipulation)

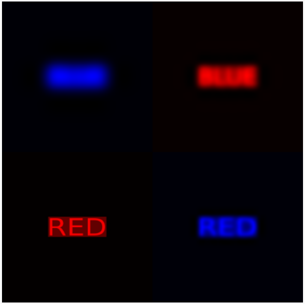
Qualitative Individual Differences

1. QID are defined by research question and experimental design
2. QID point to differences in cognitive processing (not physical impairment nor experimental manipulation)
3. QID are stable: Should be identifiable after sample noise is taken into account

Example: Stroop Effect

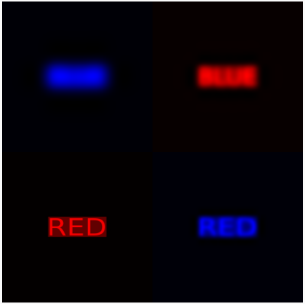
Theoretical Statement

- My favorite effect in psychology.



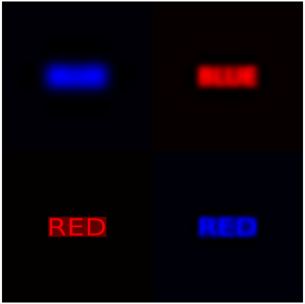
Theoretical Statement

- My favorite effect in psychology.
- Colors from congruent items are faster identified than colors from incongruent items.

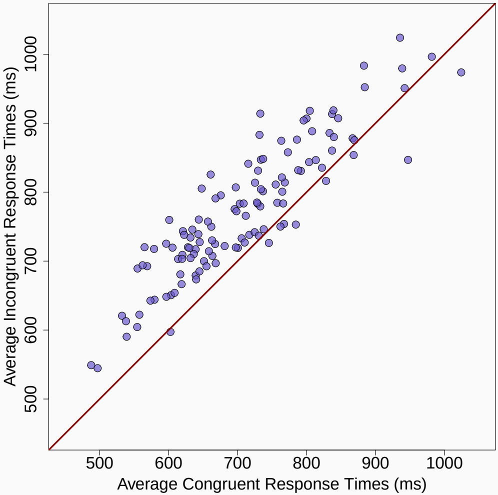


Theoretical Statement

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



Stroop Data

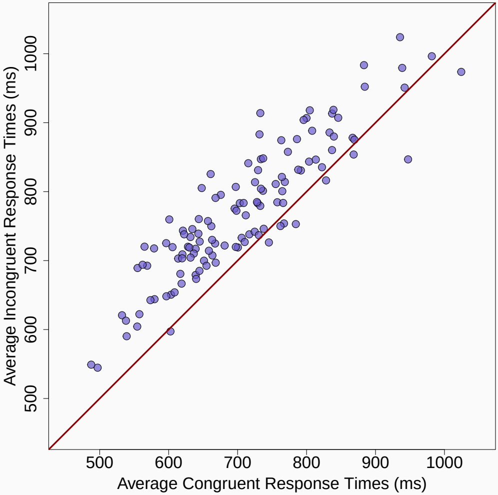


Von Bastian, Souza, & Gade (2015)

Theoretical Statement

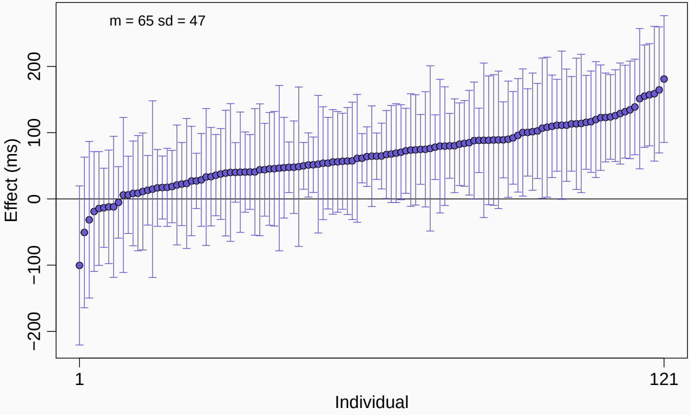
Qualitative individual differences should reflect ability to inhibit automatic reading.

Stroop Data

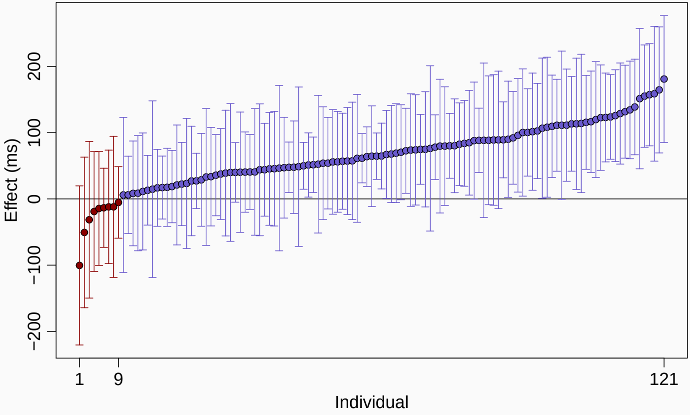


Average Incongruent Response Times (ms)
500
600
700
800
900
1000
500
600
700
800
900
1000
Average Congruent Response Times (ms)

Stroop Effects



Qualitative Differences



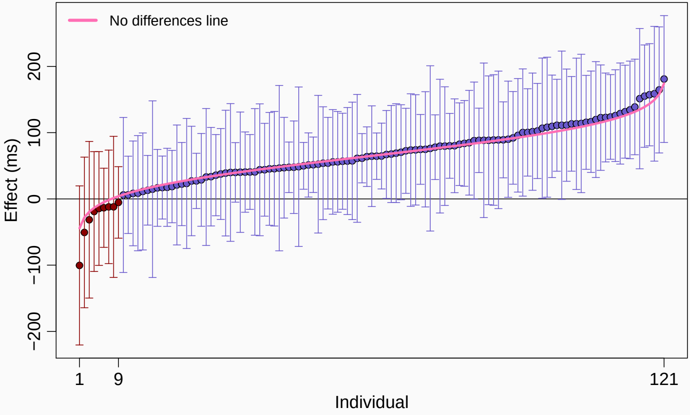
Research Questions

- **Are there qualitative individual differences of the Stroop effect?**

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

Models

- Participants: $i = 1, \ldots, I$

Models

- Participants: $i = 1, \ldots, I$
- Condition: $j = 1, 2$ (congruent, incongruent)

Models

- Participants: $i = 1, \ldots, I$
- Condition: $j = 1, 2$ (congruent, incongruent)
- Trials: $k = 1, \ldots, K_{ij}$

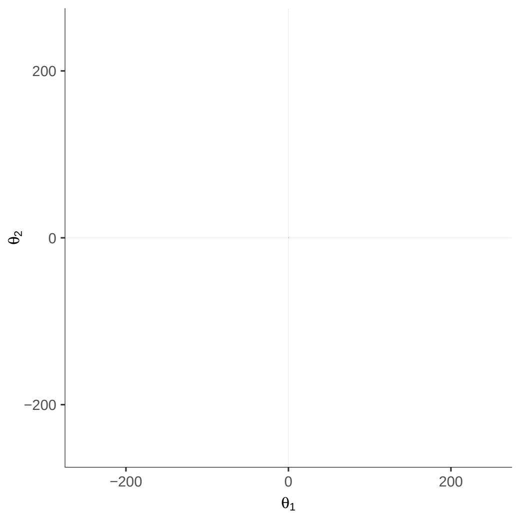
Models

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

Models

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

Models on true effects θ_i

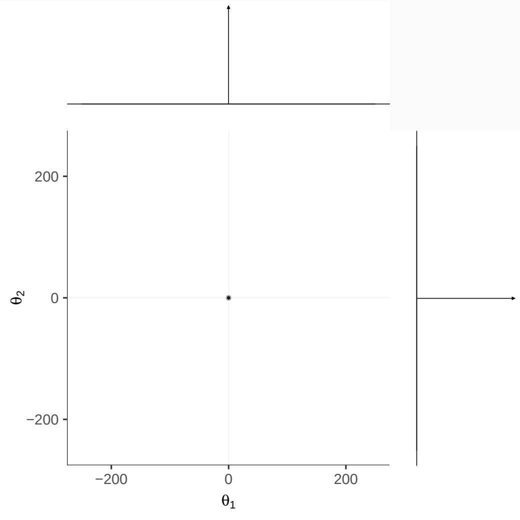


Haaf (2018); Haaf & Rouder (2017)

The Null Model

$$\theta_i = 0$$

The Null Model

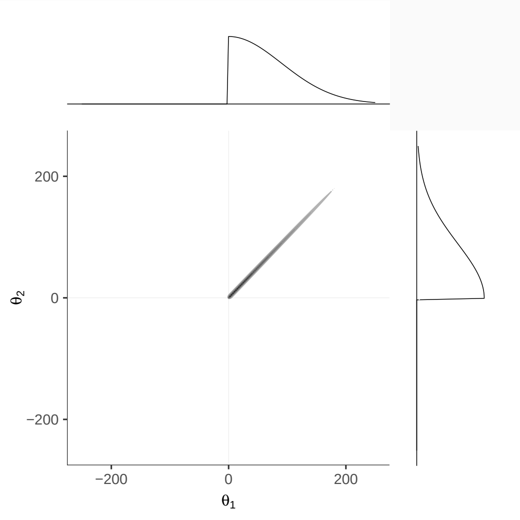


The Common-Effect Model

$$\theta_i = \nu$$

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

The Same-Effect Model

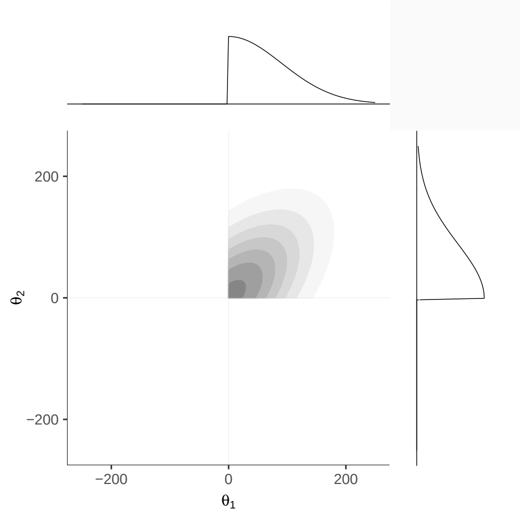


The Positive-Effects Model

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

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

The Positive-Effects Model

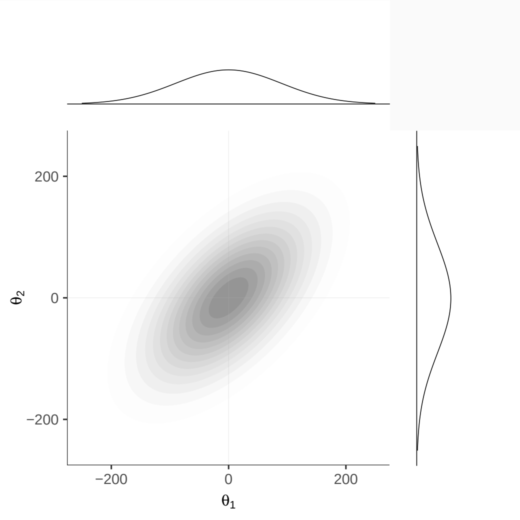


The Unconstrained Model

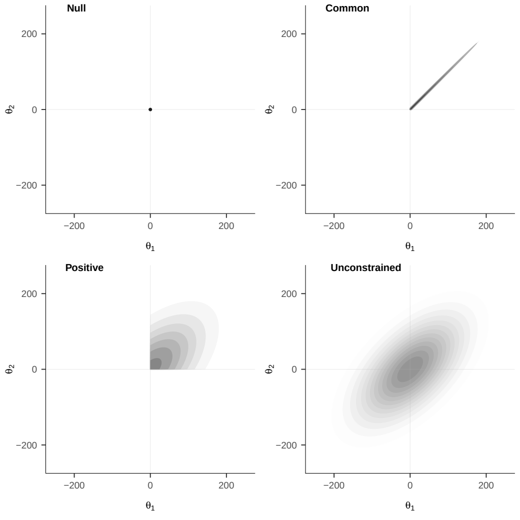
$$\theta_i \sim \text{Normal}(\nu, \tau^2)$$

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

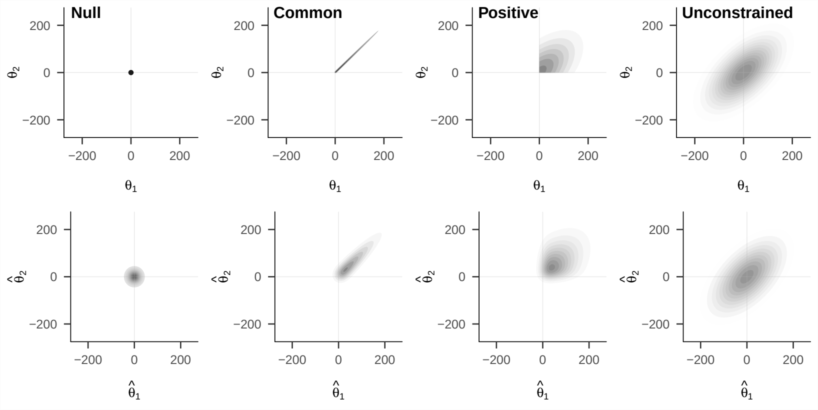
The Unconstrained Model



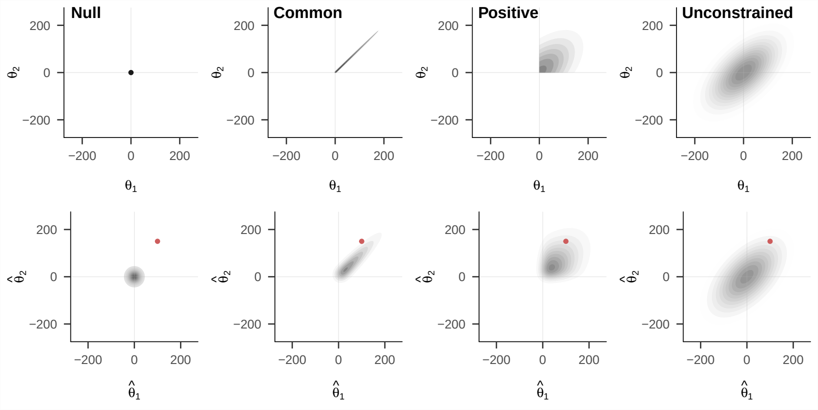
From Models. . .



From Models... to Predictions



From Models... to Predictions... to Evidence

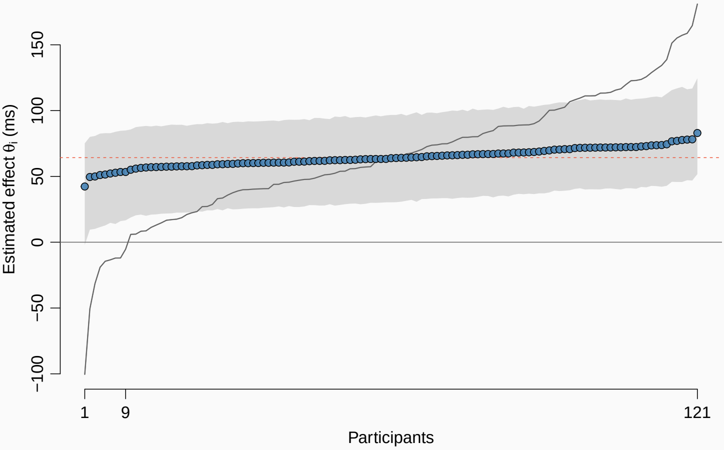


Back to Stroop

Are there qualitative individual differences of the Stroop effect?

Does everyone show a positive Stroop effect after taking sample noise into account?

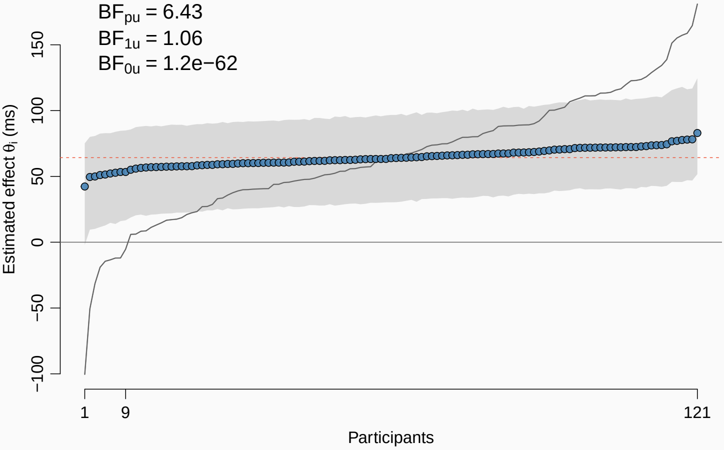
Taking sample noise into account



Taking sample noise into account

- Residual variability is low ($sd \approx 7$ ms).

Bayes factor analysis of qualitative individual differences



Conclusion

- Residual variability is low ($sd \approx 7$ ms).

Conclusion

- Residual variability is low ($sd \approx 7$ ms).
- The positive-effects model is preferred

Conclusion

- Residual variability is low ($sd \approx 7$ ms).
- The positive-effects model is preferred
- There are individual differences, but not in a meaningful way

Conclusion

- Residual variability is low ($sd \approx 7$ ms).
- The positive-effects model is preferred
- There are individual differences, but not in a meaningful way
- Caveat: There was no theoretically informed alternative model for qualitative individual differences

In R

https://github.com/jstbcs/play/tree/master/Qualitative%20Individual%20Differences

```
filename <- curl::curl("https://raw.githubusercontent.com/
                        jstbcs/play/master/Qualitative%20
                        Individual%20Differences/quid.R")
source(filename)
```

```
res <- quid(id = data$subject
            , condition = data$condition
            , rt = data$rt)
```

Mixtures of Populations

Theoretically meaningful model of qualitative individual differences

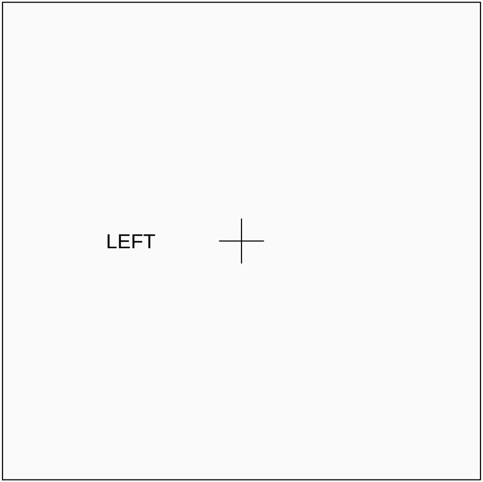
- Negative Stroop effects: Stroop pathology?

Theoretically meaningful model of qualitative individual differences

- Negative Stroop effects: Stroop pathology?
- Very unlikely that individuals truly Stroop negative

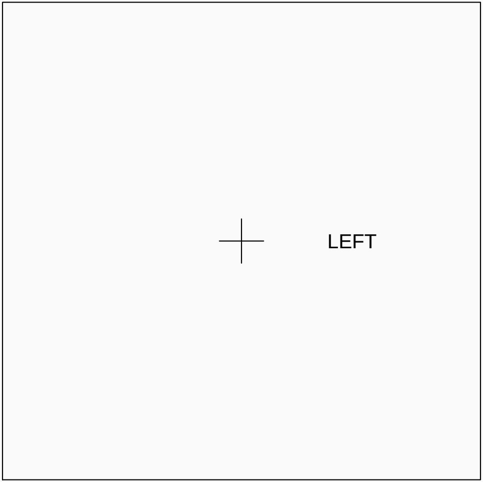
Example II: Another Stroop Effect

Location Stroop



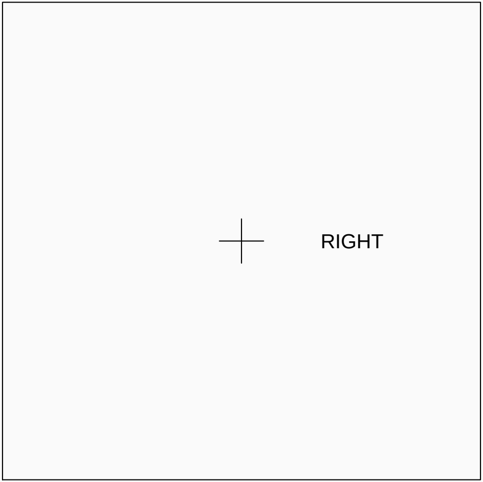
Pratte, Rouder, Morey, & Feng (2010)

Location Stroop



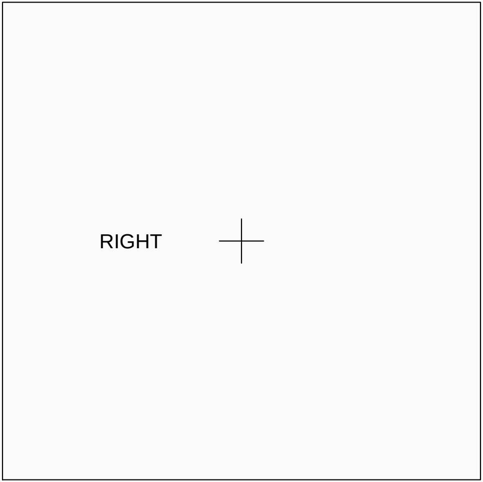
Pratte et al. (2010)

Location Stroop



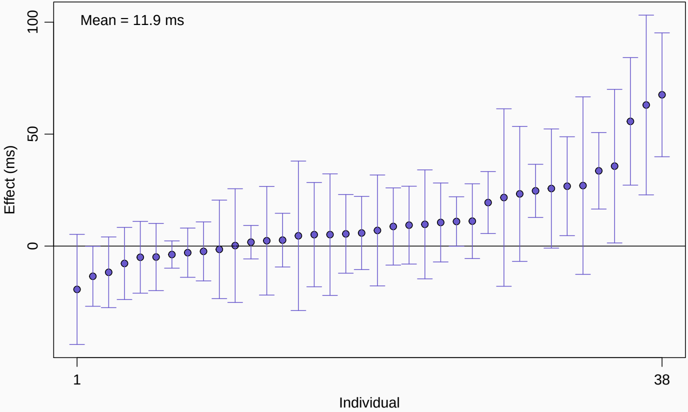
Pratte et al. (2010)

Location Stroop



Pratte et al. (2010)

Location Stroop data



Location Stroop strategies

- Participants use different strategies

Location Stroop strategies

- Participants use different strategies
- Do the task as expected

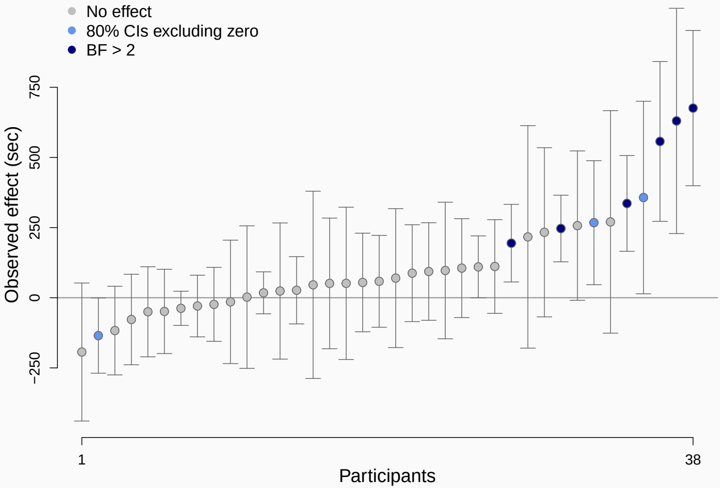
Location Stroop strategies

- Participants use different strategies
- Do the task as expected
- Or squint and get out of the room faster

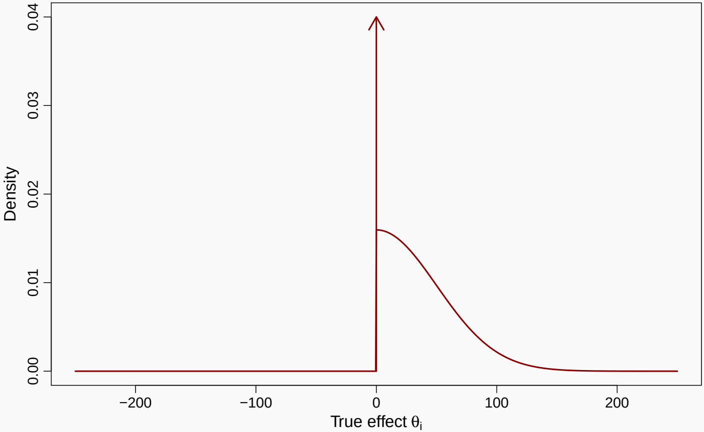
Location Stroop strategies

- Participants use different strategies
- Do the task as expected
- Or squint and get out of the room faster
- How would we know?

Classification for location Stroop



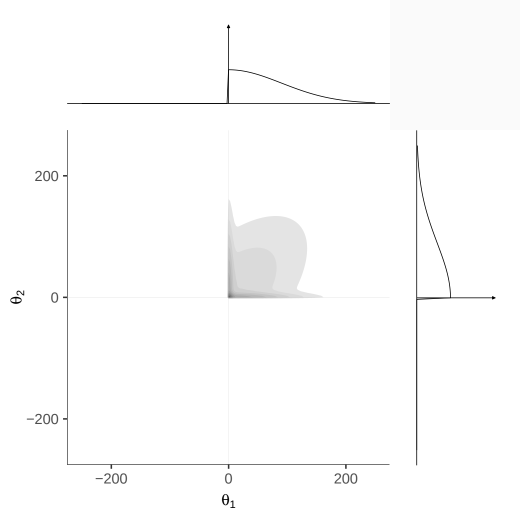
Some do some don't



Modeling Approach: Some do some don't model

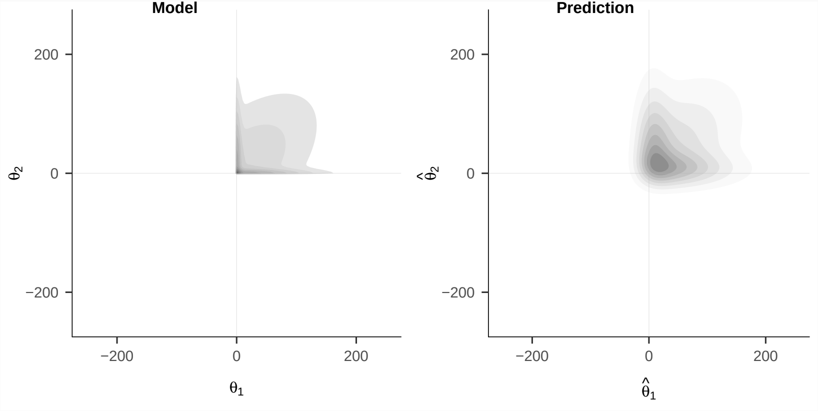
$$\mathcal{M}_{SS}: \qquad \begin{aligned} &\theta_i|(z_i = 1) \sim \text{Truncated-Normal}(\nu, \tau^2), \\ &\theta_i|(z_i = 0) = 0. \end{aligned}$$

Modeling Approach: Some do some don't model

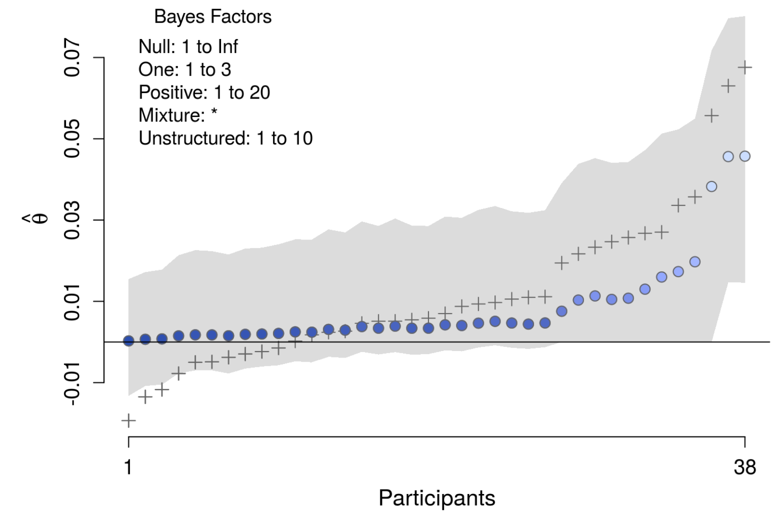


Haaf & Rouder (2019)

Modeling Approach: Some do some don't model



Location Stroop results



Conclusion

- The mixture model is preferred.

Conclusion

- The mixture model is preferred.
- There is evidence for qualitative individual differences, probably different strategies.

Conclusion

- The mixture model is preferred.
- There is evidence for qualitative individual differences, probably different strategies.
- Evidence is relatively weak, little variability left after sample noise is taken into account ($sd \approx 6$ ms).

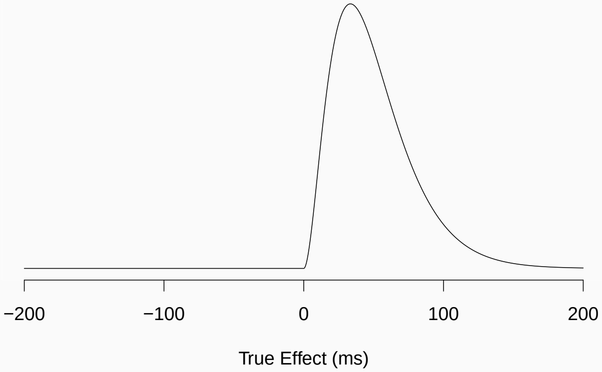
General lack of variability

(#tab:tab)

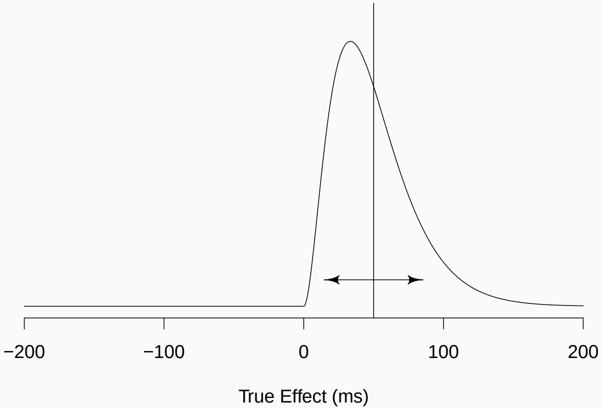
Inhibition experiments

	Total	Participants	Trials	Mean Effect	$\hat{\sigma}$	s_d	s_θ	$\hat{\tau}$
Stroop								
von.Bastian	11,245	121	46	64.68	198	47	8	19
Pratte.1	11,114	38	146	90.98	264	50	27	34
Pratte.2	12,565	38	165	11.85	160	20	6	12
Rey.Mermet.1	48,937	264	93	54.02	155	30	11	18
Rey.Mermet.2	48,966	261	94	58.88	175	69	59	64
Hedge	43,408	53	410	69.51	188	32	27	29
Simon								
von.Bastian.1	23,453	121	97	78.73	128	36	22	28
Pratte.1.1	17,343	38	228	16.81	186	24	10	15
Pratte.2.1	12,266	38	161	30.25	175	30	15	20
Flanker								
von.Bastian.2	11,215	121	46	1.38	152	32	4	12
Rey.Mermet.1.1	49,300	265	93	30.25	147	24	5	12
Rey.Mermet.2.1	39,275	207	95	36.43	107	43	37	40
Hedge.1	43,384	53	409	44.09	100	16	13	15
Other								
Rouder.1	11,346	52	109	50.29	165	28	9	16
Rouder.2	16,859	58	145	141.75	352	72	42	52
Mean	26,712	115	156	51.99	177	37	20	26
Median	17,343	58	109	50.29	165	32	13	19

Low variability comes from the ordinal constraint



Low variability comes from the ordinal constraint



What happens if we study individual differences and there are none?

The case of inhibition

- Tasks like Stroop are used in inhibition research

The case of inhibition

- Tasks like Stroop are used in inhibition research
- Examples: Simon, Flanker, Global/Local, Stop Signal, Anti-Saccade, Priming

The case of inhibition

- Tasks like Stroop are used in inhibition research
- Examples: Simon, Flanker, Global/Local, Stop Signal, Anti-Saccade, Priming
- Key: Effect = difference in performance

The case of inhibition

- Tasks like Stroop are used in inhibition research
- Examples: Simon, Flanker, Global/Local, Stop Signal, Anti-Saccade, Priming
- Key: Effect = difference in performance
- These tasks assess how well people inhibit irrelevant information like reading, location, other stimuli ...

The case of inhibition

- Tasks like Stroop are used in inhibition research
- Examples: Simon, Flanker, Global/Local, Stop Signal, Anti-Saccade, Priming
- Key: Effect = difference in performance
- These tasks assess how well people inhibit irrelevant information like reading, location, other stimuli ...
- Research Question: Is inhibition in all these tasks based on one underlying ability or separate abilities?

Problem: Low correlations

Table 2: Correlation between Stroop and Flanker Tasks

Study	Correlation
Friedman & Miyake (2004)	0.18
Pettigrew & Martin (2014)	0.03
Von Bastian et al. (2015)	0.00
Hedge et al. (2018)	-0.06
Rey-Mermet et al. (2018)	-0.09

Problem: Why are the correlations so low?

- Statistical artifact (as argued by Hedge, Powell, & Sumner, 2018)

Problem: Why are the correlations so low?

- Statistical artifact (as argued by Hedge, Powell, & Sumner, 2018)
- Psychological reality (as argued by Rey-Mermet, Gade, & Oberauer, 2018)

Problem: Why are the correlations so low?

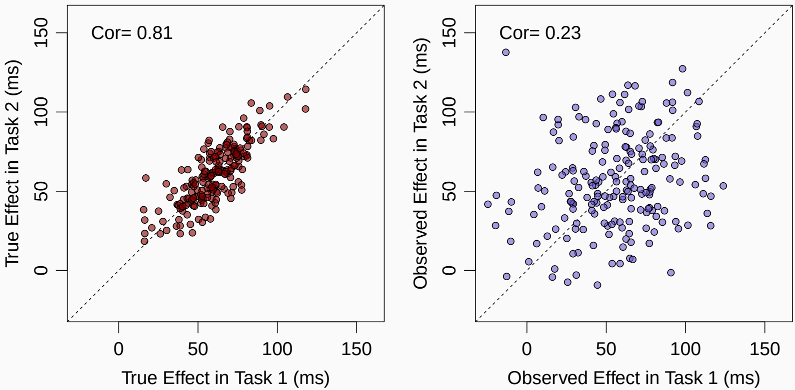
- Statistical artifact (as argued by Hedge, Powell, & Sumner, 2018)
- Psychological reality (as argued by Rey-Mermet, Gade, & Oberauer, 2018)
- There are hundreds of studies that collect hundreds of thousands of observations.

Problem: Why are the correlations so low?

- Statistical artifact (as argued by Hedge, Powell, & Sumner, 2018)
- Psychological reality (as argued by Rey-Mermet, Gade, & Oberauer, 2018)
- There are hundreds of studies that collect hundreds of thousands of observations.
- And we still don't know if the lack of correlation between Stroop and Flanker is a statistical artifact or psychological reality.

The complication of sample noise

Small individual differences + sample noise



Can we recover latent structure?

Hierarchical model for recovering correlations

- i: Individual, j: Task, k: Condition (congruent/incongruent), ℓ: Replicate

(Rouder, Kumar, & Haaf, 2019; Rouder & Haaf, 2019)

Hierarchical model for recovering correlations

- i: Individual, j: Task, k: Condition (congruent/incongruent), ℓ: Replicate
- $$Y_{ijk\ell} \sim \text{Normal}(\alpha_{ij} + x_k\theta_{ij}, \sigma^2)$$

(Rouder, Kumar, & Haaf, 2019; Rouder & Haaf, 2019)

Hierarchical model for recovering correlations

- i: Individual, j: Task, k: Condition (congruent/incongruent), ℓ: Replicate
- $$Y_{ijk\ell} \sim \text{Normal}(\alpha_{ij} + x_k \theta_{ij}, \sigma^2)$$
- $$\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{iJ})$$

(Rouder, Kumar, & Haaf, 2019; Rouder & Haaf, 2019)

Hierarchical model for recovering correlations

- i: Individual, j: Task, k: Condition (congruent/incongruent), ℓ: Replicate
- $$Y_{ijk\ell} \sim \text{Normal}(\alpha_{ij} + x_k\theta_{ij}, \sigma^2)$$
- $$\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{iJ})$$
- $$\boldsymbol{\theta}_i \sim \text{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right),$$

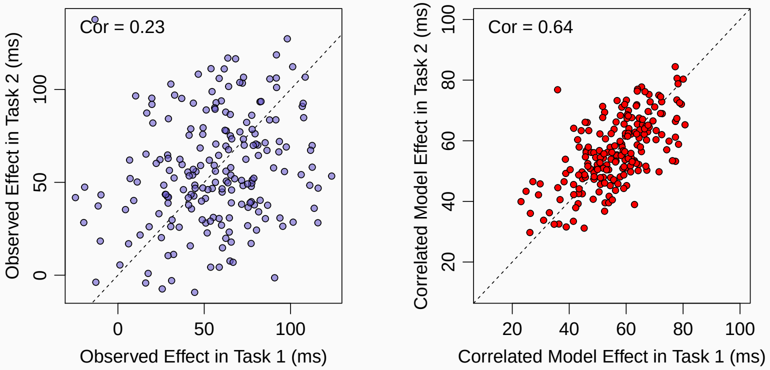
(Rouder, Kumar, & Haaf, 2019; Rouder & Haaf, 2019)

Hierarchical model for recovering correlations

- i: Individual, j: Task, k: Condition (congruent/incongruent), ℓ: Replicate
- $$Y_{ijk\ell} \sim \text{Normal}(\alpha_{ij} + x_k \theta_{ij}, \sigma^2)$$
- $$\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{iJ})$$
- $$\boldsymbol{\theta}_i \sim \text{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right),$$
- Σ is a matrix of free parameters that form a proper covariance matrix

(Rouder, Kumar, & Haaf, 2019; Rouder & Haaf, 2019)

Can it work?



Cor = 0.23
Observed Effect in Task 2 (ms)
Observed Effect in Task 1 (ms)
0
50
100
0
50
100
Cor = 0.64
Correlated Model Effect in Task 2 (ms)
Correlated Model Effect in Task 1 (ms)
20
40
60
80
100
20
40
60
80
100

Does it really work though?

Simulation

Settings based on realistic study values

- $I = 200$ People

Simulation

Settings based on realistic study values

- $I = 200$ People
- $L = 100$ Replicates

Simulation

Settings based on realistic study values

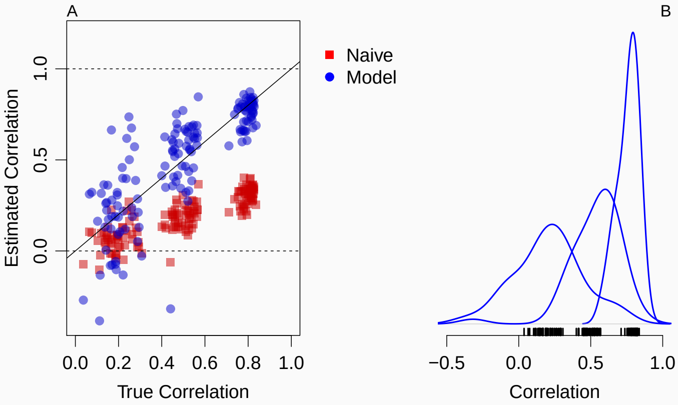
- $I = 200$ People
- $L = 100$ Replicates
- $\sigma = 200$ms trial variability

Simulation

Settings based on realistic study values

- $I = 200$ People
- $L = 100$ Replicates
- $\sigma = 200$ms trial variability
- $\sigma_\theta = 20$ms between ppl variability

Simulation 1: Two tasks



Conclusion

- When studying individual differences we are interested in stable, qualitative variability.

Conclusion

- When studying individual differences we are interested in stable, qualitative variability.
- When taking sample noise into account little variability is left.

Conclusion

- When studying individual differences we are interested in stable, qualitative variability.
- When taking sample noise into account little variability is left.
- The reason may be stochastic dominance (i.e. everyone shows the effect in the expected direction).

Conclusion

- When studying individual differences we are interested in stable, qualitative variability.
- When taking sample noise into account little variability is left.
- The reason may be stochastic dominance (i.e. everyone shows the effect in the expected direction).
- Even if no one shows a negative effect there still can be meaningful differences (some do some don't).

Conclusion

- When studying individual differences we are interested in stable, qualitative variability.
- When taking sample noise into account little variability is left.
- The reason may be stochastic dominance (i.e. everyone shows the effect in the expected direction).
- Even if no one shows a negative effect there still can be meaningful differences (some do some don't).
- When true variability is small it can be hard to study individual differences across tasks.

What can be done?

Task design

- Increase number of trials per person

Analysis

What can be done?

Task design

- Increase number of trials per person
- Get rid of conditions you don't care about

Analysis

What can be done?

Task design

- Increase number of trials per person
- Get rid of conditions you don't care about
- Strong manipulation to increase the overall effect

Analysis

What can be done?

Task design

- Increase number of trials per person
- Get rid of conditions you don't care about
- Strong manipulation to increase the overall effect

Analysis

- Use hierarchical modeling

What can be done?

Task design

- Increase number of trials per person
- Get rid of conditions you don't care about
- Strong manipulation to increase the overall effect

Analysis

- Use hierarchical modeling
- Use models that answer your theoretical question

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.

Haaf, J. M., & Rouder, J. N. (2019). Some do and some don't? Accounting for variability of individual difference structures. *Psychonomic Bulletin and Review*. Retrieved from https://psyarxiv.com/zwjtp/

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavioral Research Methods*.

Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception & Psychophysics*, *72*, 2013–2025.

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Retrieved from http://dx.doi.org/10.1037/xlm0000450

Rouder, J. N., Kumar, A., & Haaf, J. M. (2019). *Why most studies of individual differences with inhibition tasks are bound to fail*. Retrieved from https://psyarxiv.com/3cjr5/download?format=pdf

Rouder, J., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin and Review*. Retrieved from doi.org/10.3758/s13423-018-1558-y

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.