A Bayesian framework for modeling individual differences in numerical cognition

Thomas J. Faulkenberry

Tarleton State University

University of Alabama – August 31, 2021

Interference effects

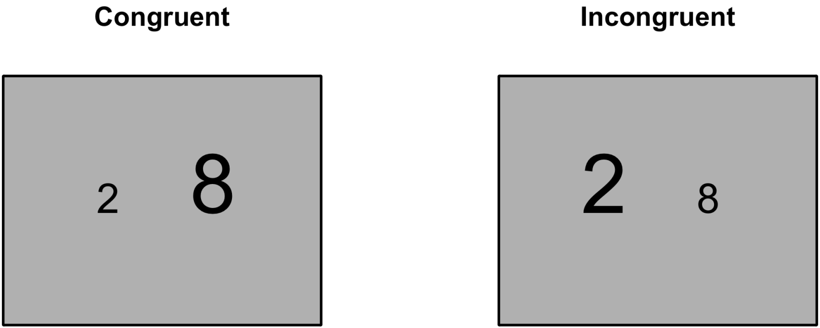
Many classic phenomena in numerical cognition present as interference effects, where responses on trials with incongruent stimuli generally take longer, on average, than responses on trials with congruent stimuli.

Two examples:

- Size congruity effect (Henik & Tzelgov, 1982)
- Unit-decade compatibility effect (Nuerk et al., 2001)

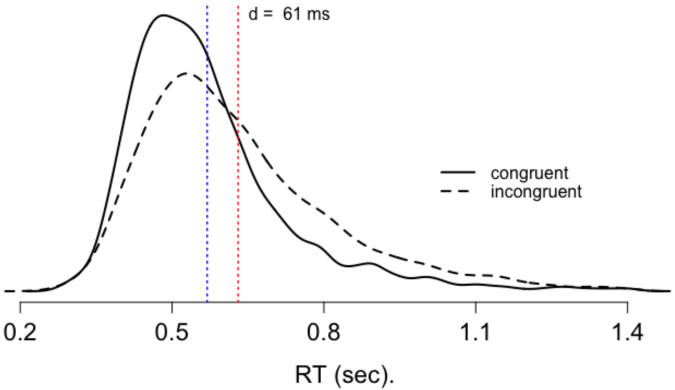
Size congruity effect

Task: choose the **physically larger** digit



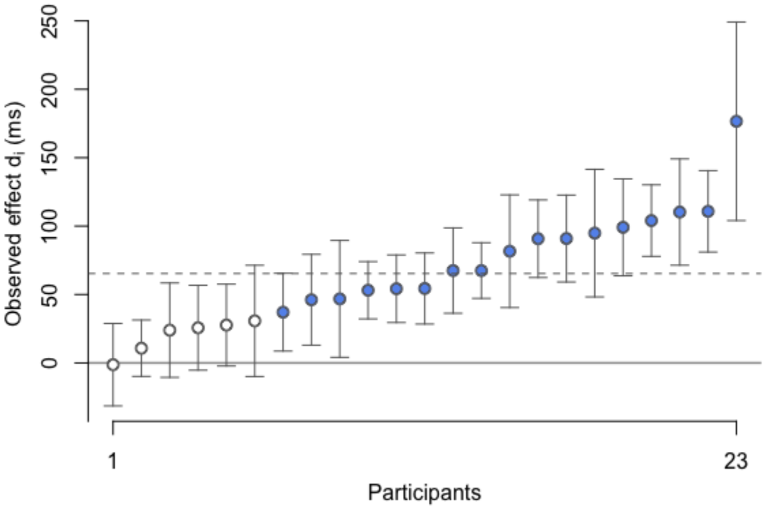
Size congruity effect

Typical result – mean RT larger for *incongruent trials*



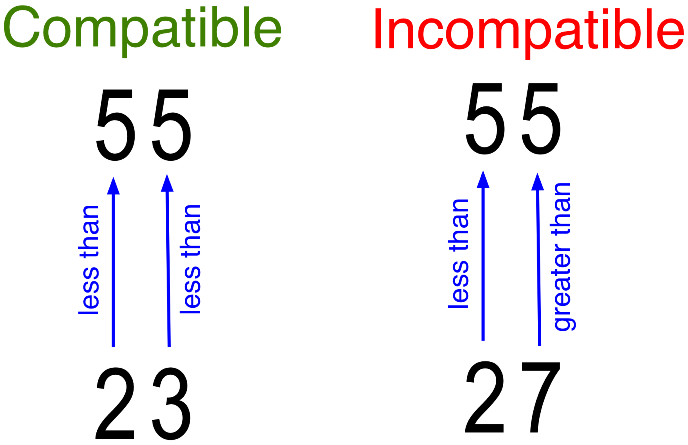
Size congruity effect

Individual effects?



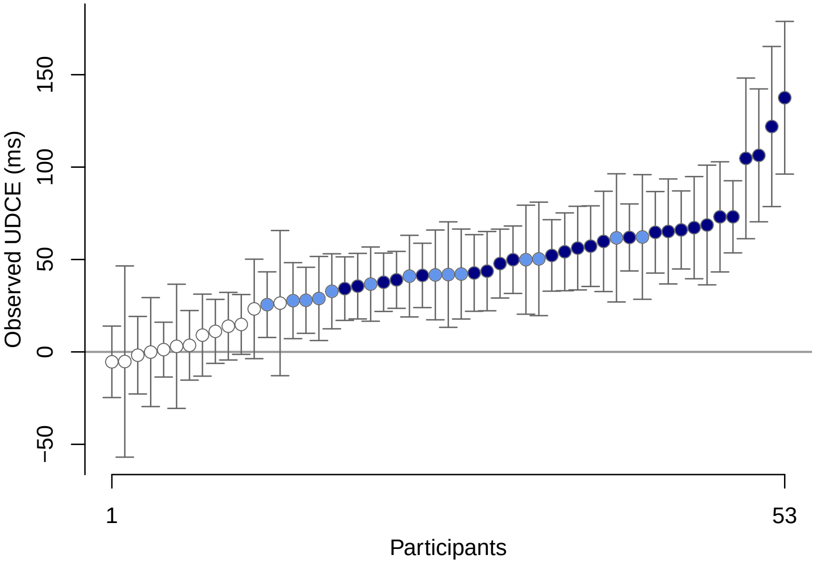
Unit-decade compatibility effect

Task – which two-digit number is larger?



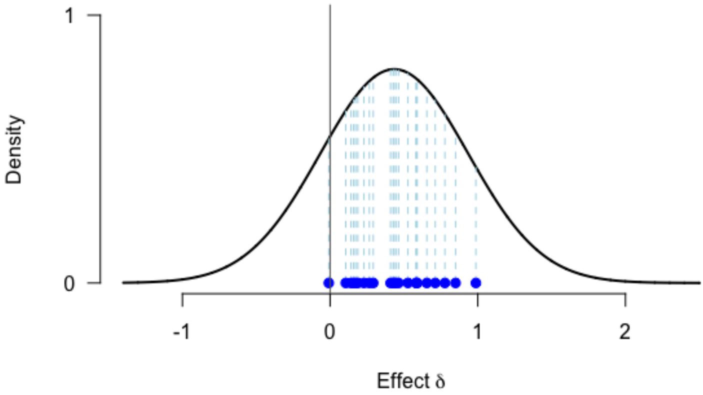
Unit-decade compatibility effect

Individual effects?



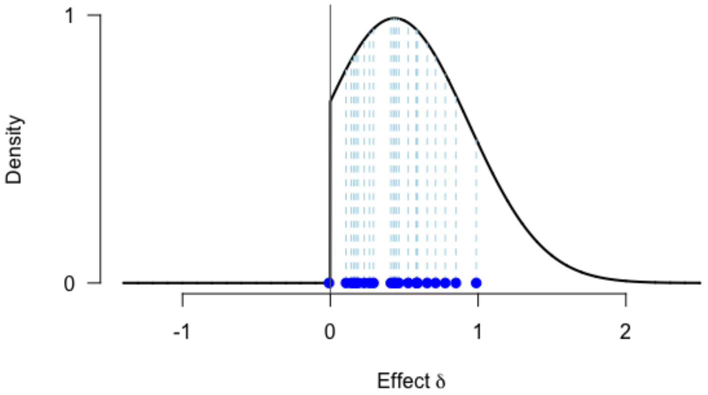
Individual differences

Suppose these observed effects d_i are drawn from population of *true* effects δ. What is the structure of this population?



Individual differences

Suppose these observed effects d_i are drawn from population of *true* effects δ. What is the structure of this population?



A new question

Does everybody *exhibit the effect?*

- if *yes*, then the effect is obligatory, resistant to strategic control, . . .
- if *no*, then the effect is complex, malleable, . . .

Importantly, both answers have downstream consequences for processing architecture of numerical cognition

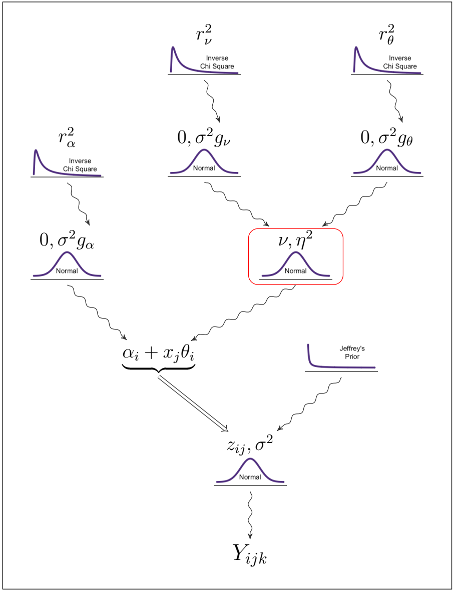
A new question

Does everybody exhibit . . .

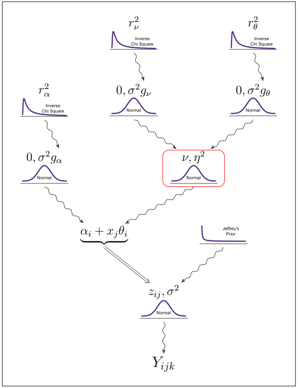
How to answer:

- build models of individual difference structures for the effect (e.g., Haaf & Rouder, 2017)
- adjudicate the models via Bayesian model comparison

Hierarchical structure



Hierarchical structure



Basic idea (Haaf & Rouder, 2017):

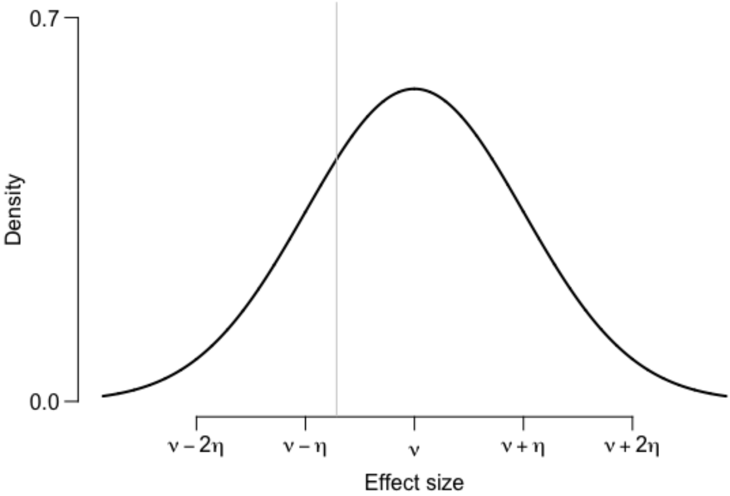
- model RTs as a random-effects linear model with effect parameter θ_i $(i = 1 \ldots, N)$
- assume (as baseline) that θ_i is drawn from a normal distribution with mean ν and variance η^2
- define competing models by **constraining** effect parameter θ_i

Four competing models

1. Unrestricted model, $\mathcal{M}_u$
2. Positive-effects model, $\mathcal{M}_+$
3. Common-effect model, $\mathcal{M}_1$
4. Null-effect model, $\mathcal{M}_0$

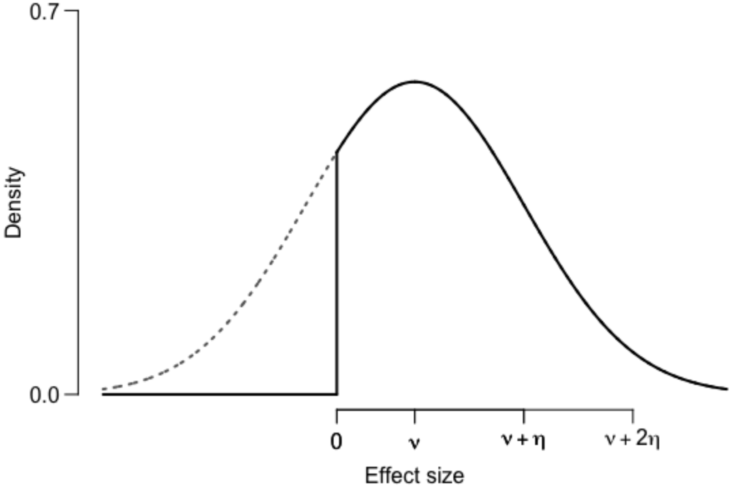
Unrestricted model

$\mathcal{M}_u : \theta_i \sim \text{Normal}(\nu, \eta^2)$



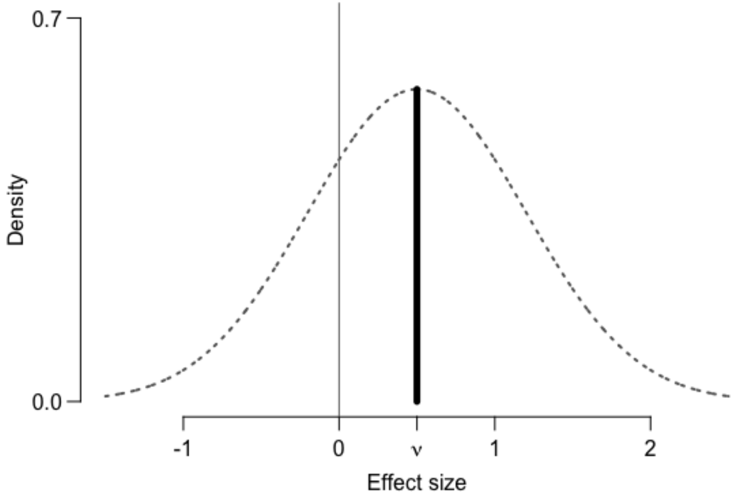
Positive-effects model

- $\mathcal{M}_+ : \theta_i \sim \text{Normal}_+(\nu, \eta^2)$



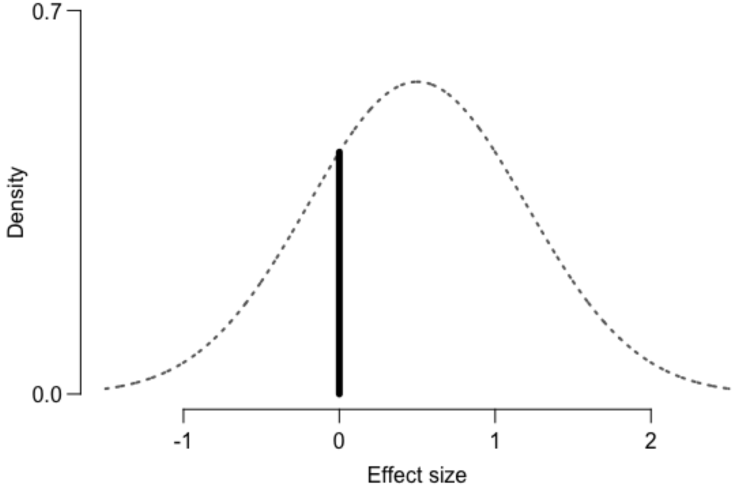
Common-effect model

- $\mathcal{M}_1 : \theta_i = \nu$



Null-effect model

- $\mathcal{M}_0 : \theta_i = 0$



Bayesian model comparison

The problem of inference

For any type of statistical inference, we fix a generative model

The problem of inference

For any type of statistical inference, we fix a generative model

(think sampling distributions)

The problem of inference

Given observed data, we then try to invert this model.

The problem of inference

Given observed data, we then try to invert this model.

The frequentist accepts or rejects $\mathcal{M}$ based on the likelihood of observing some data under a null hypothesis (i.e., the p-value)

The problem of inference

Given observed data, we then try to invert this model.

The frequentist accepts or rejects $\mathcal{M}$ based on the likelihood of observing some data under a null hypothesis (i.e., the *p*-value)

- bases decision criterion on controlling long-run error rates (i.e., α)

The problem of inference

Given observed data, we then try to invert this model.

The problem of inference

Given observed data, we then try to invert this model.

The Bayesian just directly asks: "What is the probability of this model $\mathcal{M}$, given that we've observed these data?"

The problem of inference

Given observed data, we then try to invert this model.

The Bayesian just directly asks: "What is the probability of this model $\mathcal{M}$, given that we've observed these data?"

- "posterior belief in model $\mathcal{M}$"

The problem of inference

Given observed data, we then try to invert this model.

The Bayesian just directly asks: "What is the probability of this model $\mathcal{M}$, given that we've observed these data?"

- "posterior belief in model $\mathcal{M}$"
- notation: $p(\mathcal{M} \mid \text{data})$

The problem of inference

Given observed data, we then try to invert this model.

The Bayesian just directly asks: "What is the probability of this model $\mathcal{M}$, given that we've observed these data?"

- "posterior belief in model $\mathcal{M}$"
- notation: $p(\mathcal{M} \mid \text{data})$
- no accept/reject decision

Bayes' Rule

$$\underbrace{p(\mathcal{M} \mid \text{data})}_{\substack{\text{Posterior beliefs} \\ \text{about model}}}$$

Bayes' Rule

$$\underbrace{p(\mathcal{M} \mid \text{data})}_{\substack{\text{Posterior beliefs} \\ \text{about model}}} = \underbrace{p(\mathcal{M})}_{\substack{\text{Prior beliefs} \\ \text{about model}}}$$

Bayes' Rule

$$\underbrace{p(\mathcal{M} \mid \text{data})}_{\substack{\text{Posterior beliefs} \\ \text{about model}}} = \underbrace{p(\mathcal{M})}_{\substack{\text{Prior beliefs} \\ \text{about model}}} \times \underbrace{\frac{p(\text{data} \mid \mathcal{M})}{p(\text{data})}}_{\text{predictive updating factor}}$$

Bayes' Rule

Natural action in science is to *compare* two models $\mathcal{M}_1$ and $\mathcal{M}_2$.

- Bayes' rule gives us a mathematical way to do this:

$$\frac{p(\mathcal{M}_1 \mid \text{data})}{p(\mathcal{M}_2 \mid \text{data})} =$$

Bayes' Rule

Natural action in science is to *compare* two models $\mathcal{M}_1$ and $\mathcal{M}_2$.

- Bayes' rule gives us a mathematical way to do this:

$$\frac{p(\mathcal{M}_1 \mid \text{data})}{p(\mathcal{M}_2 \mid \text{data})} = \frac{p(\mathcal{M}_1) \cdot \frac{p(\text{data}|\mathcal{M}_1)}{p(\text{data})}}{p(\mathcal{M}_2) \cdot \frac{p(\text{data}|\mathcal{M}_2)}{p(\text{data})}}$$

Bayes' Rule

Natural action in science is to *compare* two models $\mathcal{M}_1$ and $\mathcal{M}_2$.

- Bayes' rule gives us a mathematical way to do this:

$$\frac{p(\mathcal{M}_1 \mid \text{data})}{p(\mathcal{M}_2 \mid \text{data})} = \frac{p(\mathcal{M}_1) \cdot \frac{p(\text{data}|\mathcal{M}_1)}{p(\text{data})}}{p(\mathcal{M}_2) \cdot \frac{p(\text{data}|\mathcal{M}_2)}{p(\text{data})}}$$

$$= \frac{p(\mathcal{M}_1) \cdot p(\text{data} \mid \mathcal{M}_1)}{p(\mathcal{M}_2) \cdot p(\text{data} \mid \mathcal{M}_2)}$$

Bayes' Rule

Natural action in science is to *compare* two models $\mathcal{M}_1$ and $\mathcal{M}_2$.

- Bayes' rule gives us a mathematical way to do this:

$$\underbrace{\frac{p(\mathcal{M}_1 \mid \text{data})}{p(\mathcal{M}_2 \mid \text{data})}}_{\substack{\text{posterior beliefs} \\ \text{about models}}} = \underbrace{\frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)}}_{\substack{\text{prior beliefs} \\ \text{about models}}} \times \underbrace{\frac{p(\text{data} \mid \mathcal{M}_1)}{p(\text{data} \mid \mathcal{M}_2)}}_{\text{predictive updating factor}}$$

Bayes factor

The predictive updating factor

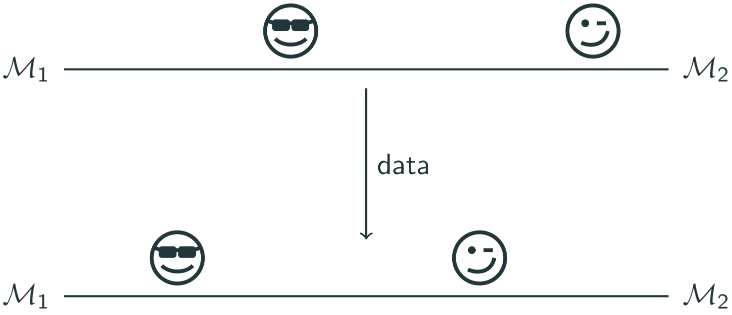
$$B_{12} = \frac{p(\text{data} \mid \mathcal{M}_1)}{p(\text{data} \mid \mathcal{M}_2)}$$

tells us how much better $\mathcal{M}_1$ predicts our observed data than $\mathcal{M}_2$.

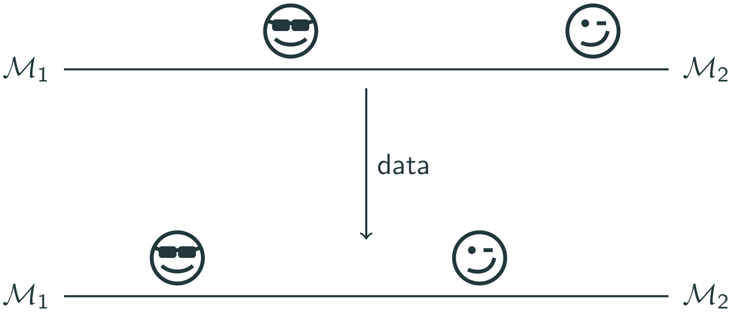
This ratio is called the **Bayes factor**

Bayes factors

Bayes factors



Bayes factors



Although 😎 and 😉 have different prior beliefs, they both shift their belief an equal amount toward $\mathcal{M}_1$.

Interpreting Bayes factors

Example 1: suppose $B_{12} = 10$.

Interpretation: the observed data are 10 times more likely under $\mathcal{M}_1$ than $\mathcal{M}_2$.

Interpreting Bayes factors

Example 1: suppose $B_{12} = 10$.

Interpretation: the observed data are 10 times more likely under $\mathcal{M}_1$ than $\mathcal{M}_2$.

Example 2: suppose $B_{12} = \frac{1}{10}$. Then $B_{21} = 10$.

Interpretation: the observed data are 10 times more likely under $\mathcal{M}_2$ than $\mathcal{M}_1$.

Interpreting Bayes factors

Example 1: suppose $B_{12} = 10$.

Interpretation: the observed data are 10 times more likely under $\mathcal{M}_1$ than $\mathcal{M}_2$.

Example 2: suppose $B_{12} = \frac{1}{10}$. Then $B_{21} = 10$.

Interpretation: the observed data are 10 times more likely under $\mathcal{M}_2$ than $\mathcal{M}_1$.

Example 3: suppose $B_{12} = 1$.

Interpretation: the observed data are equally likely under $\mathcal{M}_1$ and $\mathcal{M}_2$.

Bayes factors

Jeffreys (1961) proposed the following thresholds for evidence:

Bayes factor	Evidence
1-3	anecdotal
3-10	moderate
10-30	strong
30-100	very strong
100-	extreme

Models $\leftrightarrow$ hypotheses

Full Bayesian inference requires specification of generative models for data. This is often difficult.

Also, we are typically trained to evaluate hypotheses about effects.

To reconcile the two, several teams (e.g., Rouder, Morey, Wagenmakers, et al.) have developed *default* Bayesian hypothesis tests. The key idea is that we define models on effect size.

Models $\leftrightarrow$ hypotheses

Specifying models on effect size

Models $\leftrightarrow$ hypotheses

Specifying models on effect size

- let $\delta = \dfrac{\mu}{\sigma}$ (think Cohen's d, but at the population level)

Models $\leftrightarrow$ hypotheses

Specifying models on effect size

- let $\delta = \dfrac{\mu}{\sigma}$ (think Cohen's *d*, but at the population level)
- define competing models on δ:

Models $\leftrightarrow$ hypotheses

Specifying models on effect size

- let $\delta = \dfrac{\mu}{\sigma}$ (think Cohen's d, but at the population level)
- define competing models on δ:
  - $\mathcal{H}_0 : \mu = 0$ (the effect size is 0)

Models $\leftrightarrow$ hypotheses

Specifying models on effect size

- let $\delta = \dfrac{\mu}{\sigma}$ (think Cohen's *d*, but at the population level)
- define competing models on δ:
  - $\mathcal{H}_0 : \mu = 0$ (the effect size is 0)
  - $\mathcal{H}_1 : \mu \neq 0$ (the effect size is not 0)

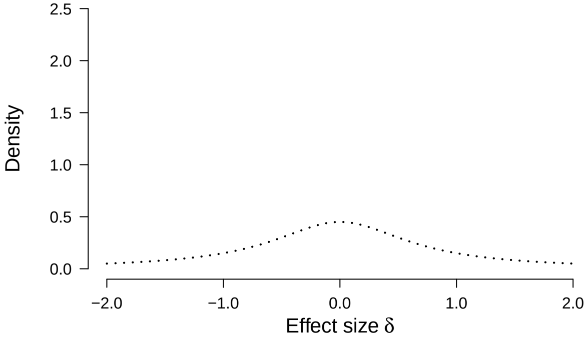
Models $\leftrightarrow$ hypotheses

Specifying models on effect size

- let $\delta = \frac{\mu}{\sigma}$ (think Cohen's d, but at the population level)
- define competing models on δ:
  - $\mathcal{H}_0 : \mu = 0$ (the effect size is 0)
  - $\mathcal{H}_1 : \mu \neq 0$ (the effect size is not 0)
- use Bayes' rule to compute

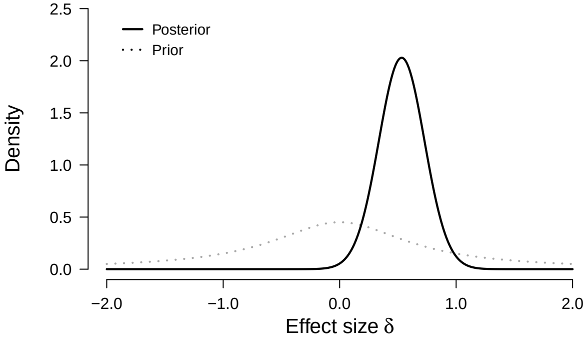
$$p(\mathcal{H}_1 \mid \text{data}) = p(\mathcal{H}_1) \times \frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data})}$$

Generic default Bayesian test



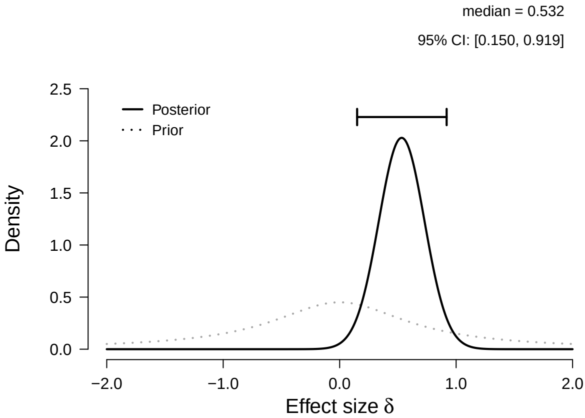
Start with prior belief about expected effect sizes δ.

Generic default Bayesian test



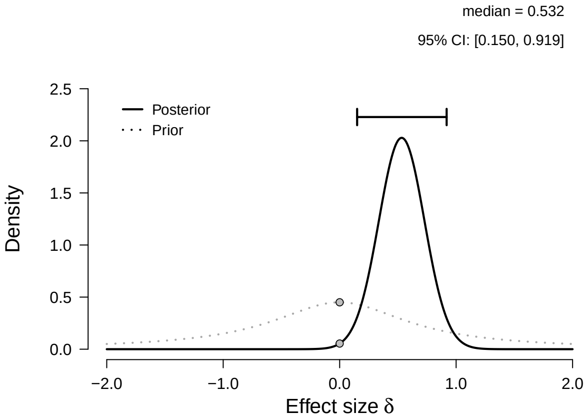
Observing data updates our prior to a posterior.

Generic default Bayesian test



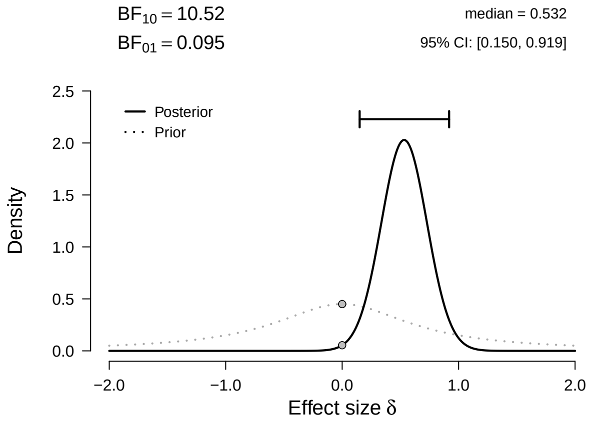
We can extract posterior estimates of δ

Generic default Bayesian test



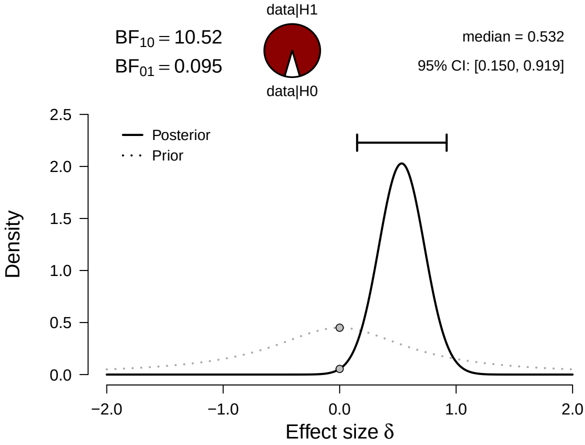
The Bayes factor is the ratio of the densities of $\delta = 0$ in the posterior and prior.

Generic default Bayesian test

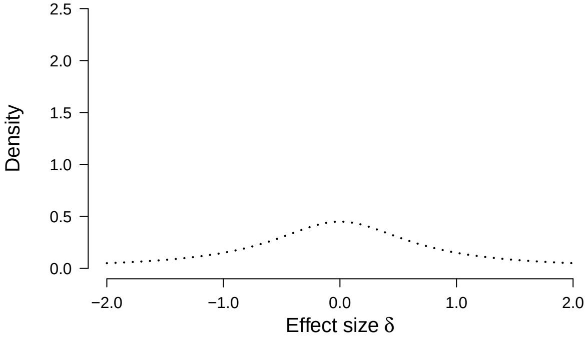


Observing data reduced our belief that $\delta = 0$ by a factor of 10.52

Generic default Bayesian test

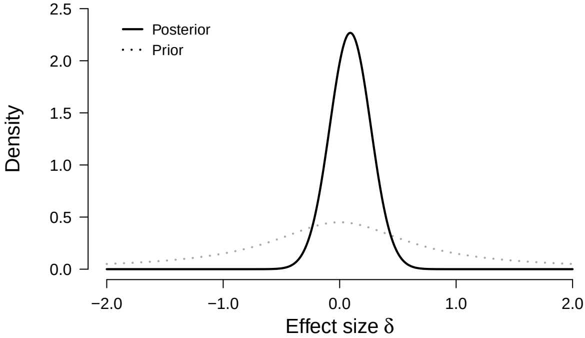


Generic default Bayesian test



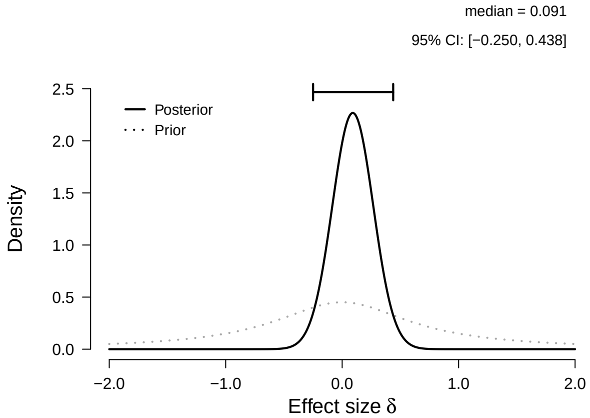
What happens if the null is supported instead?

Generic default Bayesian test



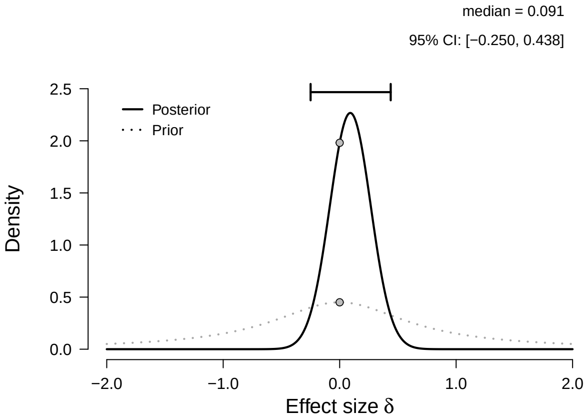
Observing data updates our prior to a posterior.

Generic default Bayesian test



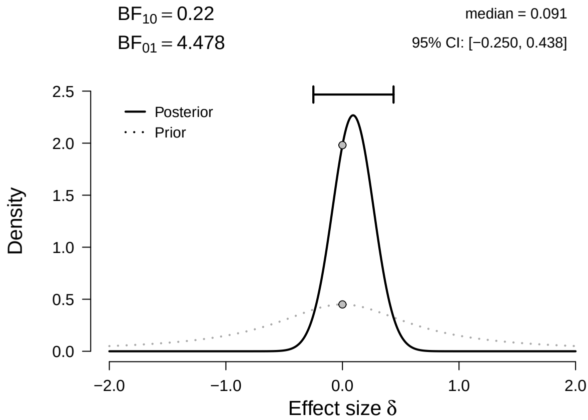
We can extract posterior estimates of δ

Generic default Bayesian test



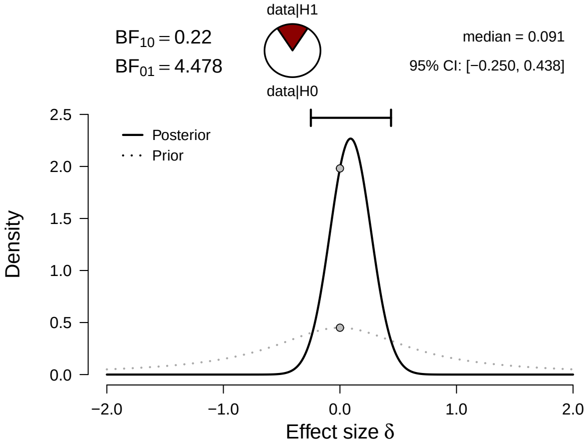
The Bayes factor is the ratio of the densities of $\delta = 0$ in the posterior and prior.

Generic default Bayesian test



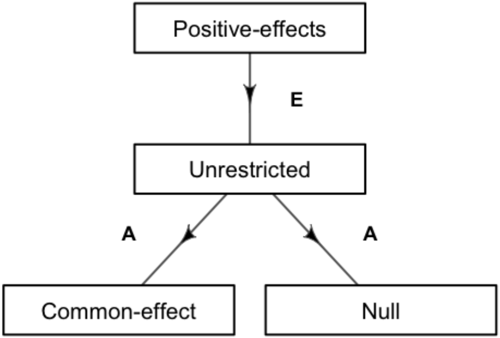
Observing data increased our belief that $\delta = 0$ by a factor of 4.478

Generic default Bayesian test



Bayes factor computations

So $B_{ab} = \dfrac{p(\text{data} \mid \mathcal{M}_a)}{p(\text{data} \mid \mathcal{M}_b)}$. How do we compute this?



Bayes factor computations

So $B_{ab} = \dfrac{p(\text{data} \mid \mathcal{M}_a)}{p(\text{data} \mid \mathcal{M}_b)}$. How do we compute this?

$$p(\text{data} \mid \mathcal{M}) = \int_{\boldsymbol{\xi} \in \Xi} p(\text{data} \mid \boldsymbol{\xi}, \mathcal{M}) p(\boldsymbol{\xi} \mid \mathcal{M}) d\boldsymbol{\xi}$$

Bayes factor computations

So $B_{ab} = \dfrac{p(\text{data} \mid \mathcal{M}_a)}{p(\text{data} \mid \mathcal{M}_b)}$. How do we compute this?

$$p(\text{data} \mid \mathcal{M}) = \int_{\boldsymbol{\xi} \in \Xi} p(\text{data} \mid \boldsymbol{\xi}, \mathcal{M}) p(\boldsymbol{\xi} \mid \mathcal{M}) d\boldsymbol{\xi}$$

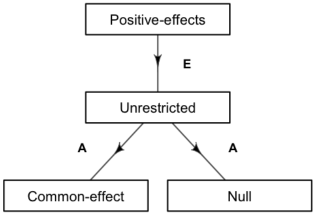
Problem: for our models $\mathcal{M}$, the parameter vectors $\boldsymbol{\xi}$ look like

$$\boldsymbol{\xi} = (\mu, \sigma^2, \nu, \alpha_1, \ldots, \alpha_N, \theta_1, \ldots, \theta_N, g_\alpha, g_\nu, g_\theta)$$

so the integral is carried out in $\mathbb{R}^{2N+6}$.

For $N = 35$, this would be a 76-dimensional integral!

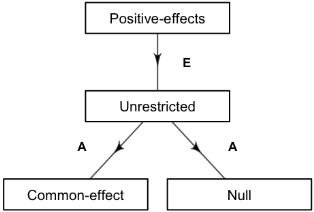
Bayes factor computations



A = *analytic approach*

- Zellner & Siow (1980); Rouder et al. (2012)
- place g-priors on individual intercepts and effect parameters
- everything *except* the g-parameters integrates symbolically
- g-parameters can be well approximated with MCMC sampling
- techniques coded into BayesFactor package in R

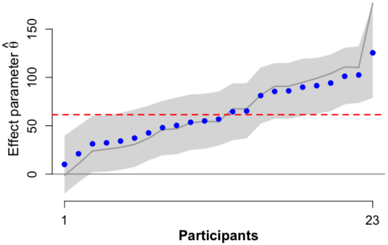
Bayes factor computations



E = *encompassing approach*

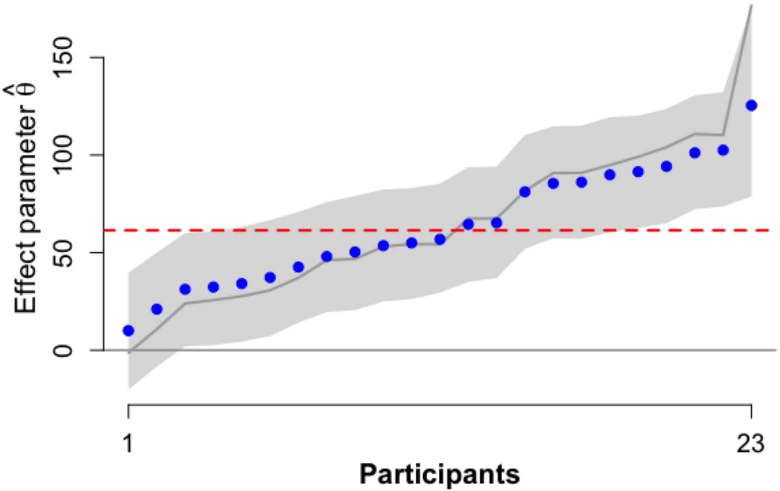
- Klugkist et al. (2005)
- generalization of Savage-Dickey density ratio
- $B_{+u} = \dfrac{P(\boldsymbol{\theta} > 0 \mid \text{data}, \mathcal{M}_u)}{P(\boldsymbol{\theta} > 0 \mid \mathcal{M}_u)}$
- probabilities computed as fraction of MCMC samples from unrestricted model that are **positive** for all individuals (both in the prior and posteriori)

Results - size congruity effect

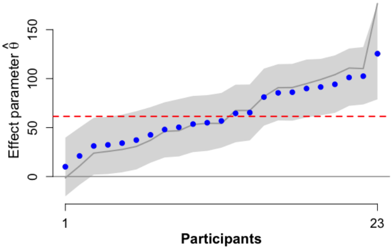


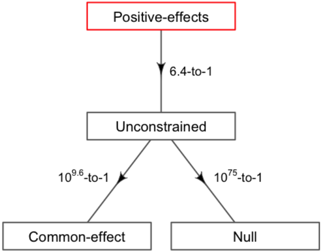
- Red line $=$ estimated effect θ from $\mathcal{M}_1$
- Blue dots $=$ individual effect estimates θ_i
- Gray line $=$ estimates from mean differences d_i
- Gray area $=$ 95% credible intervals

Results - size congruity effect

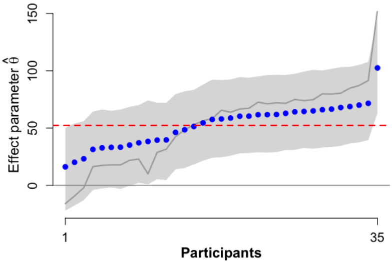


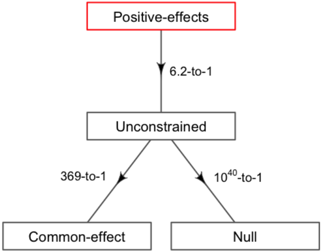
Results - size congruity effect





Results - another SCE dataset





Sensitivity to prior specifications

Experiment 1:

r_ν	r_θ	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_+$	$\mathcal{M}_u$
$\frac{1}{6}$ (50 ms)	$\frac{1}{10}$ (30 ms)	5.6e-77	4.2e-11	*	0.16
$\frac{1}{12}$ (25 ms)	$\frac{1}{20}$ (15 ms)	1.8e-76	7.7e-11	*	0.16
$\frac{1}{12}$ (25 ms)	$\frac{1}{5}$ (60 ms)	1.9e-77	8.1e-12	*	0.05
$\frac{1}{3}$ (100 ms)	$\frac{1}{20}$ (15 ms)	1.9e-76	1.9e-10	*	0.39
$\frac{1}{3}$ (100 ms)	$\frac{1}{5}$ (60 ms)	3.0e-77	3.1e-11	*	0.16

Note: Bayes factors computed against the "winning" model, denoted by *

Sensitivity to prior specifications

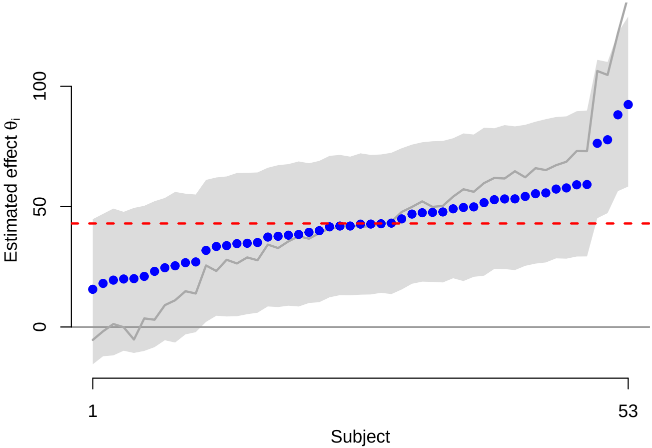
Experiment 2:

r_ν	r_θ	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_+$	$\mathcal{M}_u$
$\frac{1}{6}$ (50 ms)	$\frac{1}{10}$ (30 ms)	4.3e-41	0.0004	*	0.17
$\frac{1}{12}$ (25 ms)	$\frac{1}{20}$ (15 ms)	1.2e-40	0.0007	*	0.16
$\frac{1}{12}$ (25 ms)	$\frac{1}{5}$ (60 ms)	2.6e-41	0.0002	*	0.06
$\frac{1}{3}$ (100 ms)	$\frac{1}{20}$ (15 ms)	1.4e-40	0.0017	*	0.42
$\frac{1}{3}$ (100 ms)	$\frac{1}{5}$ (60 ms)	4.1e-41	0.0005	*	0.19

Note: Bayes factors computed against the “winning” model, denoted by *

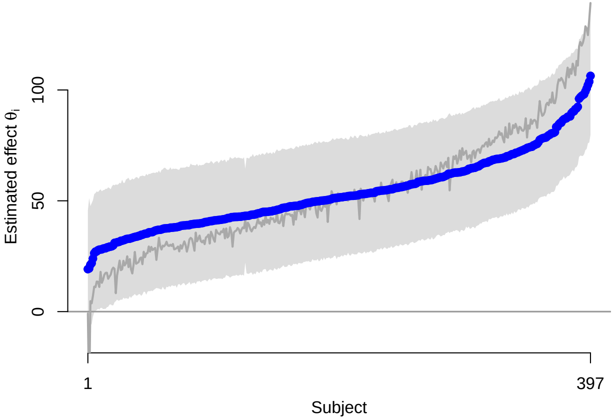
Results - unit decade compatibility effect

Data from Connolly, Bahnmeuller, Bowman, Faulkenberry, & Cipora (in preparation)



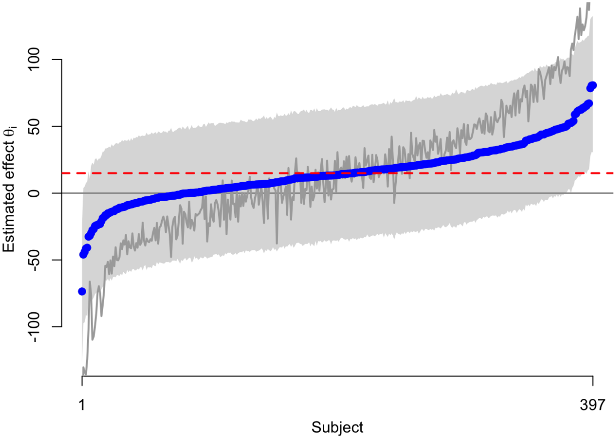
Results - numerical distance effect

Data from Vogel, Faulkenberry, & Grabner (2021)



Results - reverse distance effect

Data from Vogel, Faulkenberry, & Grabner (2021)



Summary points

Does everybody . . .

- if *yes*, then effect is obligatory, resistant to strategic control.
- if *no*, then effect is complex, malleable.

Importantly, both answers have downstream consequences for processing architecture of numerical cognition

- e.g., for SCE, what does this say about *early vs. late* interaction debate (e.g., Faulkenberry et al., 2016; Sobel et al., 2016; 2017; Faulkenberry, Vick, & Bowman, 2019)

Summary points

Some other benefits:

- Bayes factors easy to interpret
- hierarchical structure removes trial noise from individual estimates
- common effect (CE) model provides important self-check:
    - if CE model is best, is our design adequate to capture individual differences
- Might be good approach to disentangle competing theories of mental arithmetic
    - Does everyone exhibit size-by-format interaction?
    - Does everyone reflect fast counting in small addition problems?

Thank you!

- slides available at `https://tomfaulkenberry.github.io/`
- Twitter: `@tomfaulkenberry`
- Email: `faulkenberry@tarleton.edu`