Demonstrating Bayesian model comparison with a class-sourced experiment in mental arithmetic

Thomas J. Faulkenberry
Department of Psychological Sciences, Tarleton State University

A "class-sourced" experiment in mental arithmetic

The factorial design is an important topic in any research methods courses, as it is one of the cornerstones of experimental design in the behavioral and brain sciences. In this activity, students in the course each administer a simple paper/pencil mental arithmetic task to 4 participants (friends, family, etc.). Using a standard 2x2 factorial design, we look for evidence of an *interaction* between problem size and problem format. Formally, we cast this as a *model comparison problem*, where we build two competing models – one with an interaction, and one without. The presence (or absence) of this interaction is diagnostic of arithmetic processing architecture, a timely and open problem in mathematical cognition.

For model comparison, we use the *Bayes factor*, which allows us to quantify the relative evidence in favor of one model over another. Specifically, if we have two competing models $\mathcal{M}_1$ and $\mathcal{M}_2$, the Bayes factor BF_{12} indexes the extent to which our observed data is better predicted by $\mathcal{M}_1$ compared to $\mathcal{M}_2$.

$$\underbrace{\frac{P(\mathcal{M}_1 \mid \text{data})}{P(\mathcal{M}_2 \mid \text{data})}}_{\text{posterior odds}} = \underbrace{\frac{P(\text{data} \mid \mathcal{M}_1)}{P(\text{data} \mid \mathcal{M}_2)}}_{\text{Bayes factor}} \cdot \underbrace{\frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}}_{\text{prior odds}}$$

To compute Bayes factors, we use JASP, freely downloadable from https://jasp-stats.org.

Background and design

Background: Problem size and format are well-known to affect arithmetic performance (Ashcraft, 1992). Large problems are completed more slowly than small problems; similarly, problems in word format are completed more slowly than problems in digit format.

Less clear is whether problem size and format interact. An additive model of mental arithmetic (Dehaene, 1990) posits no interaction, implying that encoding and calculation are independent stages. An interactive model (Campbell, 1994) states that manipulation of encoding should *directly* affect calculation processes, and thus predicts an interaction between problem size and format. The presence (or absence) of an interaction between problem size and format is completely diagnostic of an interactive (or additive) model, respectively.

Manipulations: *problem size* (small, large) & *problem format* (digits, words)

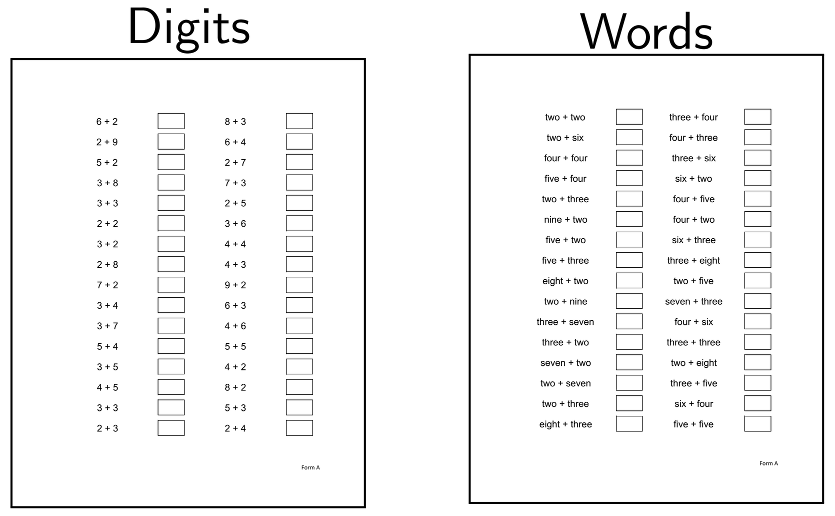
Design: mixed 2x2 factorial (problem size manipulated within subjects, format manipulated between subjects)

Primary measure: number of single-digit addition problems completed correctly in 20 seconds

Materials and procedure

Each student administers the experiment to 4 different participants. For example, a class of size 40 will result in 160 participants. Half of the participants complete "digit" problems, whereas the other half complete "word" problems. Each participant completes two sheets of arithmetic problems (in their assigned format) in counterbalanced order – one with small problems, another with large problems. Each sheet of arithmetic problems contains 32 randomly ordered problems of a given size (small, large) and format (words, digits).

Digits Words

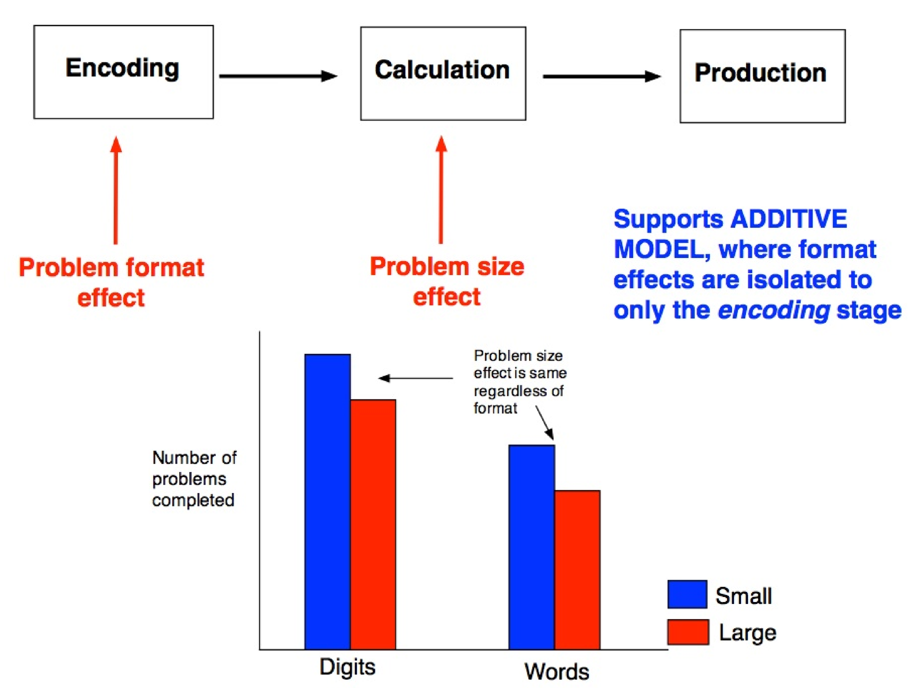


All materials, including instructions, the problem sheets, a sample writing assignment, and sample data can be downloaded from https://git.io/fpNbW.

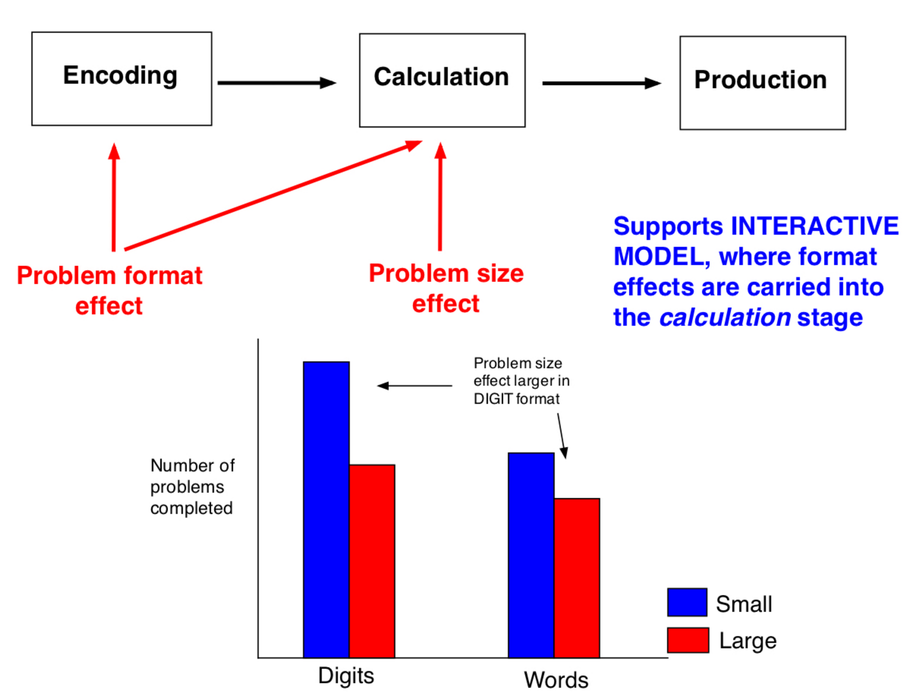
GitHub

Predictions

Additive model (Dehaene, 1990)



Interactive model (Campbell, 1994)



Models

Notation:

- Y = number of problems completed
- F = factor of *format*
- S= factor of *problem size*
- FS = interaction of format and problem size

$$\mathcal{M}_4 : \ Y \sim F + S + FS$$

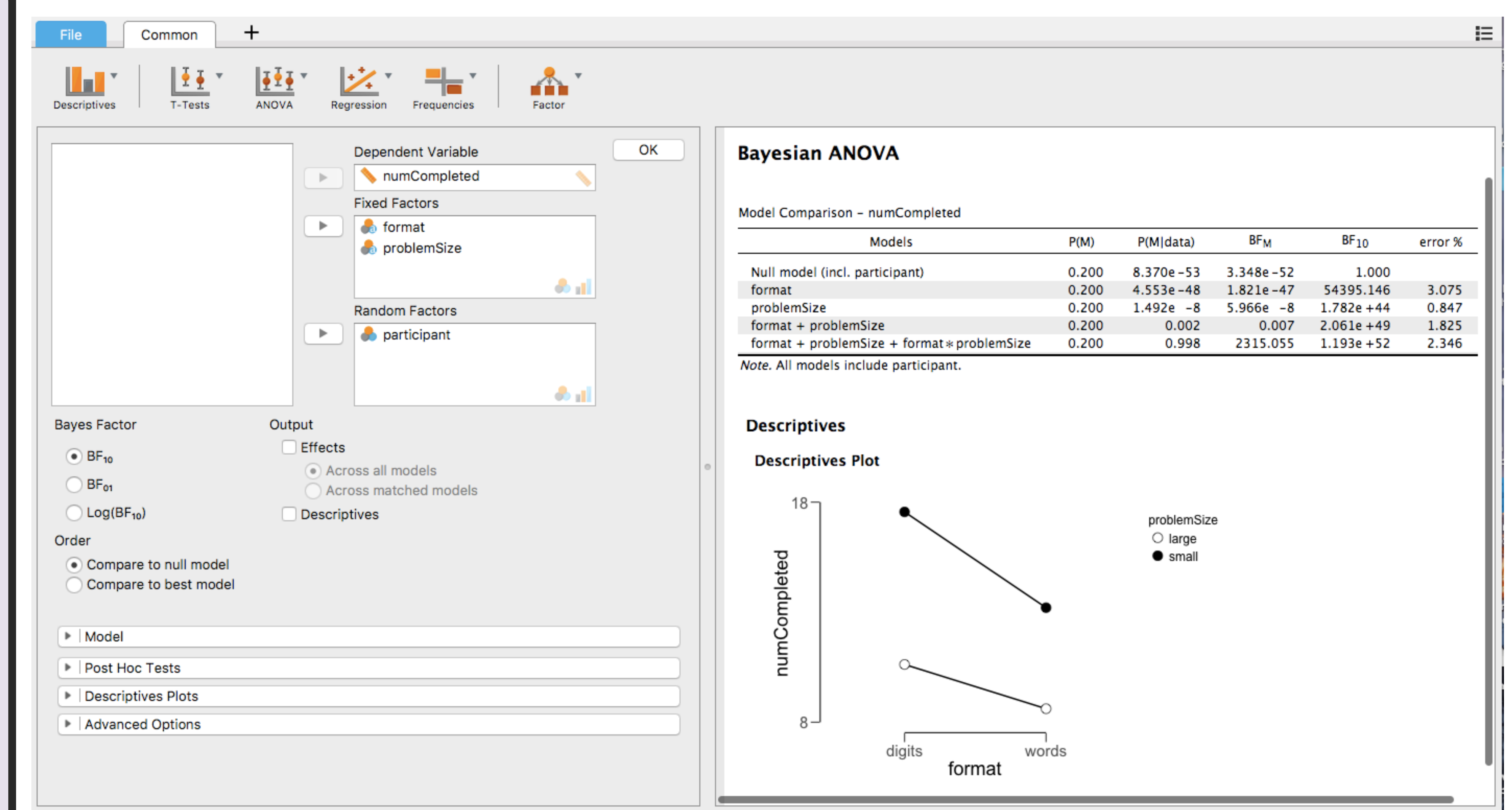
$$\updownarrow$$

$$\mathcal{M}_3 : \ Y \sim F + S$$

$$\mathcal{M}_1 : \ Y \sim F \qquad \mathcal{M}_2 : \ Y \sim S$$

$$\mathcal{M}_0 : \ Y \sim .$$

Example results from JASP



Interpretation of JASP output

Column	Interpretation
$P(M)$	All models are set to be equally likely, *a priori*. Since there are 5 models, each model has prior probability $1/5 = 0.2$
$P(M \mid \text{data})$	The prior probability for each model is updated to a *posterior* probability after seeing the data. The model with the largest posterior probability is the "main effects + interaction" model.
BF_M	BF_M represents the factor by which the *model odds* are updated after observing data. For example, the interactive model has prior odds $\frac{0.2}{0.8} = 0.25$ and posterior odds $\frac{0.998}{0.002} = 499$. Thus, the model odds have been increased by a factor of 2315.
BF_{10}	BF_{10} represents the Bayes factor for the given model over the *null model*. To compute the Bayes factor for the interactive model over the additive model, we can simply divide the two corresponding values for BF_{10}, which gives us a Bayes factor of $\frac{1.193\mathrm{e}52}{2.061\mathrm{e}49} = 578.8$.
error %	The computation of Bayes factors requires integration, which is accomplished using MCMC methods. If error percentage is large, the user can increase the number of MCMC samples in the Advanced Options menu.

Example writeup

"These models were compared via a Bayesian analysis of variance (Rouder, 2012) using the Bayesian ANOVA function in JASP. Equal prior probabilities were assigned to the five competing models. The interactive model received the largest posterior probability, $p(\mathcal{M}_4 \mid \text{data}) = 0.998$. This was much larger than the posterior probability of the additive model, which received a posterior probability of $p(\mathcal{M}_3 \mid \text{data}) = 0.002$. In terms of model odds, the interactive model received the most support from the data, which updated the prior odds for the model by a factor of 2,315. In comparison, the prior odds for the additive model were actually decreased by a factor of $1/0.007 = 143$. Finally, the Bayes factor directly comparing the two models was equal to 578.8, indicating that the observed data is approximately 579 times more likely under the interactive model than under the additive model."