Name: Tom Faulkenberry

One-sentence question: When people exhibit Stroop interference during decisions involving both physical and numerical magnitude, does the interference occur early or late in the processing stream?

Background: Decisions involving comparisons of Arabic number digits often exhibit an interference between the physical size of the digit and the implied numerical magnitude, a phenomenon called the size-congruity effect (Henik & Tzelgov, 1982). For example, suppose a participant is presented with two number symbols side by side, and one of the numbers is presented in a larger font than the other (e.g., 2 versus 8). Suppose further that the participant is asked to ignore numerical value and choose the *physically* larger digit. Even though numerical magnitude is irrelevant to this comparison task, participants are usually slower to respond on trials where physical and numerical magnitude are incongruent with each other (e.g., a large 2 paired with a small 8), compared to trials on which physical and numerical size are congruent (e.g., a small 2 paired with a large 8).

Related research over the past four decades has yielded two competing models of the phenomenon: an early interaction account (Schwarz & Heinze, 1998), where interference between numerical and physical magnitude occurs at an early encoding stage, and a late interaction account (Faulkenberry, Cruise, Lavro, & Shaki, 2016), where the interference occurs downstream as response competition during the decision process. Given these mixed results, we cannot at present disentangle these two competing theories. This study is designed to directly test these two accounts of the size-congruity effect.

Method: We will administer a timed numerical comparison task to approximately 50 adults. The task will consist of a set of single-digit number pairs. On each trial, participants will be asked to quickly and accurately indicate (via button press) the physically larger of each pair (ignoring numerical value). Half of the trials will consist of congruent number pairs (i.e., the physically larger digit is also numerically larger), whereas the other half will consist of incongruent pairs (i.e., the physically larger digit is numerically smaller). Response time (in milliseconds) will be recorded on each trial.

Analysis plan: To get an index of early versus late processing, the raw response time distributions for each participant and each experimental condition will be decomposed via the EZ-diffusion model (Wagenmakers, van der Maas, & Grasman, 2007) into drift rate, response threshold, and nondecision time. Drift rates will serve as an index of late processing, whereas nondecision times will serve as an index of early processing. Response thresholds will be assumed to be constant across conditions.

Next, participants' drift rates and nondecision times will be compared across experimental conditions (congruent versus incongruent) via a Bayesian paired samples *t*-test. For each of these dependent variables, we will define two competing models on the population effect size δ: a null hypothesis $\mathcal{H}_0\colon \delta = 0$ versus an alternative hypothesis $\mathcal{H}_1\colon \delta > 0$.

Our goal with this setup is twofold – first, we wish to see which of the two models best predicts the observed data. We will do this with the Bayes factor, which expresses the marginal likelihood of the observed data under one model compared to the other. If indeed we find support for $\mathcal{H}_1$, we will then estimate the value of the population effect size δ using the posterior distribution for δ under $\mathcal{H}_1$.

All Bayesian inference requires the specification of a prior distribution for the parameter of interest under $\mathcal{H}_1$. We will use the default prior specification in JASP, which is a Cauchy prior with scale 0.707. Under this prior, there is a 50% *a priori* probability of observing an effect size of magnitude less the 0.707, which is reasonable in the context of effects on response times in cognitive tasks. To mitigate the potential negative perception of choosing this specific prior, we will perform a robustness check in JASP, which displays the impact of prior width on the obtained Bayes factors.

Predicted results:

If the early interaction model is correct, then we should see support for $\mathcal{H}_1$ on nondecision times, but support for $\mathcal{H}_0$ on drift rates. On the other hand, if the late interaction model is correct, we should see the opposite: that is, support for $\mathcal{H}_0$ on nondecision times, but support for $\mathcal{H}_1$ on drift rates.

References:

Faulkenberry, T. J., Cruise, A., Lavro, D., & Shaki, S. (2016). Response trajectories capture the continuous dynamics of the size congruity effect. *Acta Psychologica, 163*, 114–123. https://doi.org/10.1016/j.actpsy.2015.11.010

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition, 10*(4), 389–395. https://doi.org/10.3758/bf03202431

Schwarz, W., & Heinze, H. J. (1998). On the interaction of numerical and size information in digit comparison: A behavioral and event-related potential study. *Neuropsychologia, 36*(11), 1167–1179. https://doi.org/10.1016/s0028-3932(98)00001-3

Wagenmakers, E.-J., van der Maas, H., & Grasman, R. (2007). An EZ-diffusion model for response time and accuracy. Psychonomic Bulletin & Review, 14(1), 3-22. https://doi.org/10.3758/BF03194023