RESPONSE TIME MODELING FOR THE SIZE CONGRUITY EFFECT:

EARLY VS. LATE INTERACTION

A Thesis

by

KRISTEN A. BOWMAN

Submitted to the College of Graduate Studies of
Tarleton State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Thomas Faulkenberry, PhD
Committee Members,	Kimberly Rynearson, PhD
	Robert Newby, PhD
Head of Department,	Jamie Borchardt, PsyD
Dean, College of Graduate Studies,	Barry D. Lambert, PhD

May 2020

Major Subject: Applied Psychology

ACKNOWLEDGEMENTS

First, I would like to acknowledge certain members of the field of cognitive psychology. A special thanks to Dr. Danielle Navarro for providing the inspiration to be a thoughtful researcher rather than simply a rule follower. She truly inspired me. Next, I would like to thank Dr. Stephen Link for explaining to me the history that has led up to our current knowledge of response time models. Additionally, Dr. Stephan Lewandowsky gave me so much courage and enthusiasm that helped carry me through the end of this study. Likewise, not only has Dr. Jeff Rouder published many wonderful and insightful articles over mathematical modeling, he also shared insightful wisdom with me. Finally, Dr. E.J. Wagenmakers provided assistance by sharing figures for inclusion in my thesis; for this I am truly grateful.

As for members of the Tarleton community, I would like to thank Dr. Jennifer Blevins for her support from the very beginning of my research experience. She became a member of a support team for me that included Dr. Lesley Leach, Ms. Lacie Harris, Ms. Linda Sanders, and Dr. Eileen Faulkenberry. Without these individuals, I would not be as successful of a young woman and researcher. Next, a special thanks goes to the many mentors that I have connected with in the Department of Psychological Sciences. To name a few, Dr. Jamie Borchardt, Dr. Amber Bozer, Dr. Man’Dee Mason, Dr. Jennifer Dias, and Dr. Trina Geye played a large role in providing supportive and constructive feedback throughout this process. In addition, a very special thanks goes to Mrs. Judy Geye for her never ending happiness and enthusiasm for my well-being. In addition, a special thanks to Dr. Bob Newby and Dr. Kim Rynearson for agreeing to serve on my thesis committee. Their knowledge and feedback greatly helped to shape the outcome of this project. A very special thanks goes to my committee chair and mentor, Dr. Tom Faulkenberry. He is my main role model. I hope to one day be as awesome as he is at teaching and conducting research. Without a doubt, this thesis experiment is one to be proud of because of all the countless hours during which he and I discussed every single detail. While giving me the freedom to select my own thesis project, he also helped me to find a project that is a very solid foundation for the rest of my future as a researcher. Through all the years that we have worked together, this year has truly been the most difficult but the most rewarding.

Finally, I would like to acknowledge the love and support that I have received from my family. Specifically, I would like to acknowledge my mother for her willingness to listen and aid in my distress over my thesis. Additionally, I would like to acknowledge the total love and encouragement from my son, Caden. If anyone has reminded me of my abilities, it would be him. From giving me space to work on my study all the way to giving me supportive pats on the back, I am grateful for my son.

ABSTRACT

Bowman, Kristen A. Response time modeling for the size congruity effect: Early vs. late interaction. MASTER OF SCIENCE (Applied Psychology) May 2020, 53 pp., 1 table, 7 figures, 59 titles.

The size-congruity effect occurs when numerical magnitude interferes with judgments of physical size. Various accounts propose that this interference is either encoding-related or decision-related. To discriminate between these accounts, I used a class of mathematical models (ex-Wald, shifted Wald and EZ-Diffusion) to index the underlying cognitive processes via estimates of drift rate, response threshold, and non-decision time. I administered a single-digit physical comparison task and manipulated congruity and measured response times. First, I found that congruent trials were processed faster than incongruent trials, which is indicative of the size-congruity effect. Next, via the mathematical models, I found that the drift rate for incongruent trials was smaller than for congruent trials, indicating that incongruent trials had a faster rate of information uptake. The response threshold for incongruent trials was larger than for congruent trials, indicating that for incongruent trials more information needed to be accumulated before responding. Critically, there was no difference for non-decision time between incongruent and congruent trials. This combination of results provides support for a late interaction account of the size-congruity effect, shedding further light onto decision-related models of number processing.

TABLE OF CONTENTS

LIST OF GRAPHICS

FIGURE **Page**

CHAPTER I

INTRODUCTION TO THE SIZE CONGRUITY EFFECT

The curious nature of decision making has stumped cognitive psychologists over the past several decades. One method of studying this complex process is to limit the stimulus sets to basic units of information (e.g., words, letters, numbers, etc.). A well-known procedure for studying cognitive processing is the Stroop task (1935), in which participants were asked to name the color in which words were presented. There are two types of trials: congruent trials, where font color and printed word lead to the same decision (e.g., "purple" in purple font) and incongruent trials, where font color and printed word lead to different decisions (e.g., "purple" in green font). Even though this is a seemingly simple task, the manipulation of color font induces a lag in response times for incongruent trials compared to congruent trials, which is known as Stroop interference (MacLeod, 1991; Lindsay & Jacoby, 1994). This interference indicates that incongruent trials require a different type of cognitive processing than congruent trials.

The peculiar finding that there is an interference in processing for incongruent trials, but not for congruent trials, started a curiosity among researchers to investigate more than colors. For example, the Numerical Stroop task is an excellent vehicle for investigating the underlying mechanisms of number processing. This task follows the design of the Stroop task, but utilizes Arabic number digits. Typically, participants are presented with two single-digit number symbols where one symbol is *physically* larger than the other. Participants are instructed to select the physically larger symbol and ignore numerical magnitude (Santens & Verguts, 2011; Schwarz & Heinze, 1998; Reike & Schwarz, 2017).

The task is comprised of two types of trials: congruent trials, where the physically larger symbol also has a larger numerical value (e.g. 3 vs 7) and incongruent trials, where the physically larger symbol has a smaller numerical value (e.g. 3 vs 7). For congruent trials, where physical size and magnitude lead to the same decision, participants are able to respond rather quickly. However, for incongruent trials, where physical size and magnitude lead to different decisions, participants require more time before a response is finalized. This interference is known as the size-congruity effect (Besner & Coltheart, 1979). Furthermore, this robust effect demonstrates that while magnitude is not needed to complete the task at hand, participants still access magnitude (Algom, Dekel, & Pansky, 1996; Besner & Coltheart, 1979; Fitousi & Algom, 2006; Henik & Tzelgov, 1982; Sobel, Puri, & Faulkenberry, 2016; Sobel, Puri, Faulkenberry, & Dague, 2017; Reike & Schwarz, 2017; Risko, Maloney, & Fugelsang, 2013).

Rather than questioning "if" this robust effect occurs, the question has become "how" it occurs and what mechanisms are utilized during the decision-making process. There are two prevalent accounts of processing for the size-congruity effect: relative speed of processing and automaticity. The relative speed of processing hypothesis assumes that when processing stimulus characteristics that vary along multiple dimensions (e.g., physical size, numerical value), parallel processing is occurring at different speeds for the multiple dimensions. Due to the limited capacity of working memory, only one of the two dimensions can take priority in processing (Rouder, Morey, Cowan, Zwilling, Morey, & Pratte, 2008; Santens & Verguts, 2011). Therefore, task instructions dictate which dimension will take priority in processing (MacLeod, 1991). Through the lens of the speed-of-processing hypothesis, the size-congruity effect occurs because identifying the physically larger number is slowed due to the attention that is required for processing magnitude. However, this interpretation only makes sense for the incongruent trials, due to the lack on interference in congruent trials.

On the other hand, the second model for the size-congruity effect is automaticity, which is typically associated with low level cognitive processing. In other words, this type of processing only requires minimal amounts of energy or effort in processing (e.g., numbers, letters, time, spatial location, etc.). In particular, processing magnitude is considered to be automatic. Furthermore, Walsh (2003) argued that time, space, and quantity are similar domains that stem from one magnitude comprehension system and are linked by certain mechanisms for processing, even though the literature has investigated them separately. Each of these domains are essential in processing and when analyzed together are known as A Theory of Magnitude (ATOM). The advantage to this theory is that it brings together three domains that are indicative of common processing mechanisms, which are used to make fast decisions about the environment. Through the lens of automaticity, the size-congruity effect occurs because automatic processing of congruent trials should occur rather quickly. However, incongruent trials are presumably slower because they require more effortful processing than congruent trials.

In contrast to automatic processing, effortful processing is associated with high level cognitive processing. In other words, this type of processing requires relatively high amounts of energy through rehearsal and elaborative processing (e.g., details, imagery, mnemonic techniques, etc.; Hasher & Zacks, 1979). Through consistent practice and mental mapping, people are able to internalize input and produce an output with more ease (Sohen, Servan-Schriber, & McClelland, 1992). Through the lens of effortful processing, the size congruity effect occurs because for incongruent trials more time is required to produce a response than for congruent trials.

While the surprising nature of the size-congruity effect can be demonstrated in the lab, there are real world implications for decision making. Within numerical decision-making tasks there are two different judgments that occur at the same time, but ask two different questions. Judgments of physical size ask “how much?” and judgments of magnitude ask “how many?” For example, when people are asked to guess how many pieces of candy are in a jar, both types of judgments can be used. For physical size, people will assess the volume of the jar to produce an answer. Whereas for numerical magnitude, people will make a judgment based on numerosity to produce an answer (Walsh, 2003). Overall, one of the main goals for cognitive psychology is to discover the underlying mechanisms of magnitude comprehension (Risko, Maloney, & Fuglsang, 2013). In the context of the size-congruity effect, automatic processing causes participants to go “against the grain” of what information is automatically accessed (MacLeod, 1991). Thus, the size-congruity effect is able to be detected. The size congruity effect contains rich information about decision making and information processing.

An Early vs. Late Interaction Account

While the size-congruity effect is robust, researchers have found conflicting evidence as to where the locus of the interference in processing occurs. Santens and Verguts (2011) were the first to frame the debate as follows; the size congruity effect stems from either the input level of representational overlap or the output level of response competition. Thus, there are two competing models for the temporal location of the interference; an early interaction account and a late interaction account (Santens & Verguts, 2011). An early interaction account suggests that the relative delay from incongruent trials occurs at the encoding stage; see Figure 1.A. During this stage, participants encode both physical size and numerical magnitude into an analog representation and the processing of information continues in a serial fashion until the appropriate response is activated. Participants may only process one unit at a time before proceeding to the next decision (Townsend & Ashby, 1983). Thus, the interference occurs at the initial stages of processing.

On the other hand, a late interaction account suggests that the interference occurs at the decision stage; see Figure 1.B. In contrast to an early interaction account where there is only one merged pathway, in a late interaction account there are two separate pathways for numerical input and physical input. In other words, once the input is internalized, the two separate channels move to numerical representation and physical representation. During this stage both channels will activate "subresponses." When these two subresponses are in agreement, such as in congruent trials, processing can occur rather quickly. However, when the two subresponses are in disagreement, such as in incongruent trials, there is a relative slowdown in processing (Santens & Verguts, 2011; Schwarz & Heinze, 1998). After the information has passed through the two representational channels, the information goes to the decision stage and finally to output. Thus, because there is an interference of physical size and magnitude at the response stage, there must be parallel processing. In other words, the two separate pathways activate two different decision codes, which will compete in the decision stage and cause a delay in response time. Additionally, these two decisions are not necessarily made at the same time due to one unit requiring more processing than the other (Townsend & Ashby, 1983). The duration of this competition feeds forward into the response activation stage, which causes the relative delay in processing of incongruent trials (Faulkenberry, Cruise, Lavro, & Skaki, 2016).

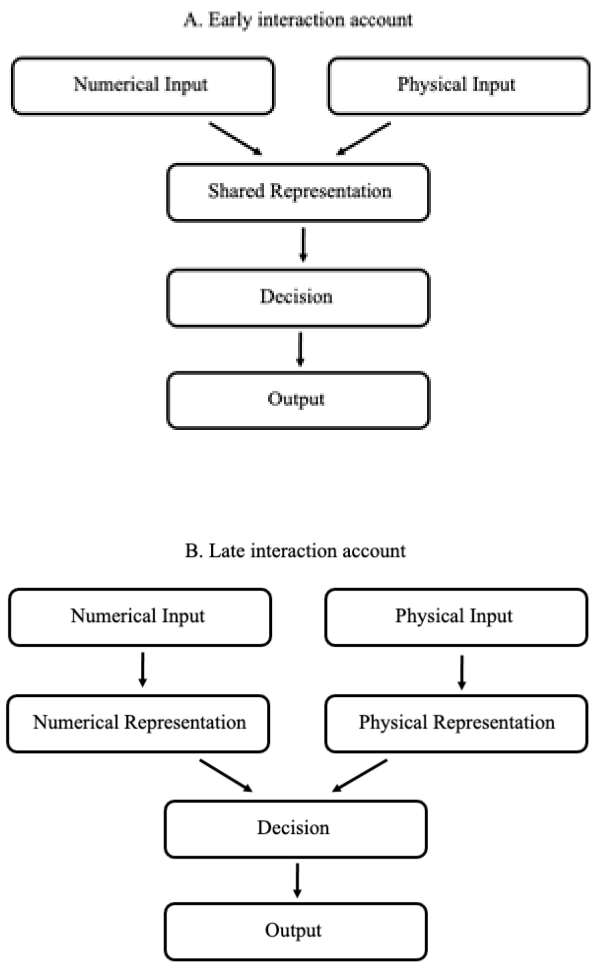

Figure 1. Illustration of an Early interaction account (A) and a Late interaction account (B). Illustration adapted from Santens and Verguts (2011).

Evidence for the Competing Accounts

Both an early interaction account and a late interaction account have received support in the literature. Schwarz and Heinze (1998) performed the size congruity task and measured response time data and event-related potentials (ERPs). They argued that the dimensions of physical size and numerical value are extracted into parallel channels and then only the prominent portions of the dimensions must be integrated into one representation in order to make a decision. They found that congruent trials were processed faster than incongruent trials. One conclusion is that incongruent trials require more effortful processing than congruent trials because participants must consider both answer choices. Contrary to this conclusion, Schwartz and Heinze (1998) monitored ERPs and found no indication of activation for incorrect responses in incongruent trials. Thus, they proposed that the locus of interference is in the early portion of the decision-related process.

On the other hand, Cohen Kadosh and Henik (2006) performed a digit comparison task using fMRI machines. They argued that if a late interaction account is accurate, researchers should see higher motor activity for incongruent trials than congruent trials due to parallel processing. Cohen Kadosh and Henik (2006) found that when they presented participants with a digit comparison task where the numerical distance was 5, there was higher workload. Thus, they concluded that because of the high motor activity, participants must be using parallel processing. However, they concluded that depending on the nature of the task, participants can change their strategy to produce an answer.

In line with the evidence for a late interaction account, Faulkenberry, Cruise, Lavro, and Shaki (2015) performed the size-congruity task with computer mouse tracking, which is considered to be reflective of the physical behavior associated with the decision-making process. The signatures of the mouse trajectories showed that incongruent trials were significantly attracted toward the incorrect response. Thus, indicating that participants consider the incorrect response for a large portion of the response time. This leads to the conclusion that there must be parallel processing of the dimensions of physical size and numerical value (see also Santens & Verguts, 2011).

Mathematical Models of Decision Making

Traditional hypothesis testing generally utilizes only the mean and standard deviation of response time curves. However, formal mathematical models allow for the entire shape of the distribution to be utilized, which helps to shed light onto the underlying decision-related processes. The purpose of mathematical modeling is to further investigate theories and laws that are not directly observable.

Maximum Likelihood Estimation

One approach to gain insight to the underlying cognitive processes of decision-making is to fit probability distributions to model the observed reaction times. A probability distribution is defined with specific parameters. One technique to estimate these parameters is maximum likelihood estimation (Myung, 2003). In addition to the practical uses of maximum likelihood estimation, there are also theoretical advantages: (1) each estimated parameter is sufficient to yield complete information about the underlying distribution; (2) each parameter is consistent with large enough samples, that is; the estimated parameters can get arbitrarily close to the true parameter value; (3) the estimation process is efficient because it produces the lowest-possible variance asymptotically; and (4) regardless of specified density function, the parameters can be transformed from one parameter formula to another.

In general, the purpose of conducting maximum likelihood estimation is to find parameter estimates that are the most likely given a set of observed data. It is important to note that likelihood is not the same as probability. Though they can have similar interpretations, the presentation is slightly different. That is:

$$f(x|\theta) = L(\theta|x),$$

where *x* represents the observed data and θ represents specific parameter values. This equation says that the likelihood function and the probability density function are related, but two separate functions. The probability density function *f* provides probabilities of observed data given specific parameter values. On the other hand, the likelihood function *L* provides likelihoods of parameter values given observed data (Myung, 2003). In other words, probability distributions are conditioned on parameters, but likelihood functions are conditioned on data. Also note that probability functions have a scale of 0 to 1, whereas likelihood functions do not have such restriction.

Gaussian and Ex-Gaussian

The first step in maximum likelihood estimation is selection of the likelihood function. One typical selection is the normal distribution, also known as the Gaussian distribution. The normal distribution provides a symmetric probability distribution. However, the symmetry of the normal distribution does not reflect the true nature of response times, which have a pronounced positive skew (Dawson, 1988; Whelan, 2008). That is, response time distributions tend to have higher density on the left that gradually decreases over time and a long positive tail on the right. One potential probability distribution that exhibits this positive skew is the ex-Gaussian distribution (Luce, 1986; Rieger & Miller, 2019; Whelan, 2008). The ex-Gaussian distribution is an exponentially modified Gaussian curve, meaning that it contains a normal component with an exponential rate of decay. The density function for the ex-Gaussian is:

$$f(x|\mu,\sigma,\tau) = \frac{1}{\tau\sqrt{2\pi}} \exp\left(\frac{\sigma^2}{2\tau^2} - \frac{x-\mu}{\tau}\right) \cdot \int_{-\infty}^{[(x-\mu)/\sigma]-(\sigma/\tau)} \exp\left(-\frac{y^2}{2}\right) dy,$$

where μ represents the mean, σ represents the standard deviation of the normal component, and τ represents rate of the exponentially distributed tail.

Model selection is vital as varying models can lead to different interpretations. Specifically, the ex-Gaussian distribution is a likelihood function, rather than a probability distribution. Hervey et al. (2006) were interested in comparing neuropsychological performances for children with and without ADHD. When fitting a Gaussian distribution, they found that children with ADHD had significantly slower response times than the children without ADHD. However, after fitting an ex-Gaussian distribution, they demonstrated that children with ADHD had significantly faster response times for the normal component of the response distribution. More interestingly, for the exponential component τ, they modeled a fatter tail, indicating that children with ADHD exhibited a greater number of response times that were beyond their mean performance than the children without ADHD. Thus, they concluded that children with ADHD are not necessarily slower than children without ADHD, rather they are more prone to attention lapses on some trials (Hervey et al., 2006). The portion of trials with attention lapses should not be the data used to draw severe conclusions. The benefit of the ex-Gaussian distribution is that such differences in groups can be analyzed separately in order to gain further insight (see also Kobor et al., 2015, for a similar finding).

Another example of the applicability of the ex-Gaussian model came from Penner-Wilger, Leth-Steensen, & LeFevre (2002), who were interested in comparing mental arithmetic differences across cultures. Specifically, they were testing the problem-size effect (i.e., larger problems, such as 7 x 8 or 9 + 7, require more time to process than smaller problems, such as 3 x 3 or 3 + 2) between Chinese and Canadian students. Through fitting an ex-Gaussian distribution to their data, they found that for the Chinese group, the effect occurs in the normal parameter μ, whereas for the Canadian group, the effect occurs in normal parameter μ and the exponential parameter τ. This indicated that for the Chinese group, the locus of the effect is in memory retrieval, whereas for the Canadian group, the locus of the effect occurs in both memory retrieval and through nonretrieval processing.

While these examples provide further insight into group differences, researchers have been tempted to infer that these parameters directly reflect underlying cognitive processes. For example, Hohle (1965) claimed that μ reflects the motor-related portion of the decision process and that τ reflects the cognitive processes (Frigs, 2018; Hohle, 1965). Yet, there has been very little agreement on the exact process that each parameter reflects. Therefore, these parameters should be viewed as solely a descriptive tool (Matzke & Wagenmakers, 2009; Schwarz, 2001). While the reporting of descriptive parameters can be valuable, there are other tools that can enhance the inferences made about a given hypothesis.

Wald Distributions

As mentioned earlier, the major limitation of the Gaussian and ex-Gaussian distributions is that they are limited to descriptive statistics that are purely descriptive information about the distribution. However, our goal is to provide information about the unobservable, latent processes involved in decision making. One method to analyze the specific components of decision making is the use of accumulator models (Link, 1975), which are mathematical models that represent the decision-making process as a random walk towards a fixed response boundary. These models represent decision making as the process of gathering noisy partial information over time until enough information is accumulated to make a decision (Schwarz, 2001). The purpose of using an accumulator model is to increase accuracy by averaging out random fluctuations in trial-by-trial decisions (Heathcote & Hayes, 2012). Typically, accumulator models have the same three basic parameters that index cognitive processes: quality of presented information, amount of information uptake required to trigger a response, and nondecision time (Anders, Alario, & Maanen, 2016).

One such accumulator model is the Wald model, which is also known as the Inverse Gaussian. The Wald is a distribution of stopping times for a continuous diffusion process (i.e., continuous random walk) with a fixed boundary. The density function for the Wald is given by:

$$f(x|\nu, \alpha) = \frac{\alpha}{\sqrt{2\pi x^3}} \exp\left[-\frac{(\alpha - \nu x)^2}{2x} \right],$$

where x is the distribution of first passage times, γ is the drift rate, and α is response threshold. While this can be a powerful modeling tool, it is not yet suitable for response time distributions because the curve begins at zero on the x-axis. This is problematic because response times cannot be zero. Furthermore, this puts a great deal of weight on very small response times. Response times below 100 milliseconds are proven to be inaccurate representations of processing (Luce, 1986). There are two solutions to this problem: (1) similar to the ex-Gaussian distribution, we can construct an exponential modification of the Wald distribution; or (2) add a constant shift to the distribution.

The first of these solutions is the ex-Wald (Schwarz, 2001). The ex-Wald model is the mathematical combination of the Wald distribution with an exponential tail. In other words, the ex-Wald offers a rightward shift that is the convolution of two curves, which also allows for the beginning of the curve to not be at zero (Heathcote, 2004; Schwarz, 2001). First, the cumulative distribution function (CDF), *F*, is given by:

$$F(x|\nu,\sigma,\alpha) = \Phi\Big(\frac{\nu x - \alpha}{\sigma\sqrt{x}}\Big) + \exp\Big(\frac{2\alpha\nu}{\sigma^2}\Big)\cdot\Phi\Big(-\frac{\nu x + \alpha}{\sigma\sqrt{x}}\Big),$$

where ν represents drift rate which is the data-driven component of the task, σ represents a scaling parameter: without loss to generality, this is set to 1, α represents response criterion, and Φ is the cumulative standard normal distribution function. From there, the density function of the ex-Wald is given by:

$$h(x|\nu,\sigma,\alpha,\gamma) = \gamma \exp\Big[-\gamma t + \frac{\alpha(\nu - k)}{\sigma^2}\Big]\cdot F(t|k,\sigma,\alpha),$$

where γ represents drift rate and k is defined as

$$k = \sqrt{\mu^2 - 2\gamma\sigma^2} \geq 0.$$

The ex-Wald is more theoretically sound than the ex-Gaussian as a model for decision making (Schwarz, 2001), but there are only a couple of studies that utilize this mathematical density function due to its complexity; thus, it needs further research (Rieger & Miller, 2019).

On the other hand, the shifted Wald has a constant shift of the entire distribution, which is far easier to perform than the ex-Wald (Anders, Alario & Van Maanen, 2016). The density function for the shifted Wald is given by:

$$f(x|\alpha,\gamma,\theta) = \frac{\alpha}{\sqrt{2\pi(x-\theta)^3}} \cdot \exp\Big(-\frac{(\alpha-\gamma(x-\theta))^2}{2(x-\theta)}\Big),$$

where γ represents drift rate, α represents response threshold, and θ

represents nondecision time. The major benefit of the shifted Wald distribution is that each of these parameters can be interpreted as specific processing stages of the decision-making process. The level of drift rate exhibits the quality of the information presented (ambiguous, unambiguous), response threshold reflects the caution to the response (conservative, liberal), and the amount of time required for the nondecision demonstrates the speed of encoding and motor response (fast, slow) (Anders, Alario & Van Maanen, 2016) (see Figure 2).

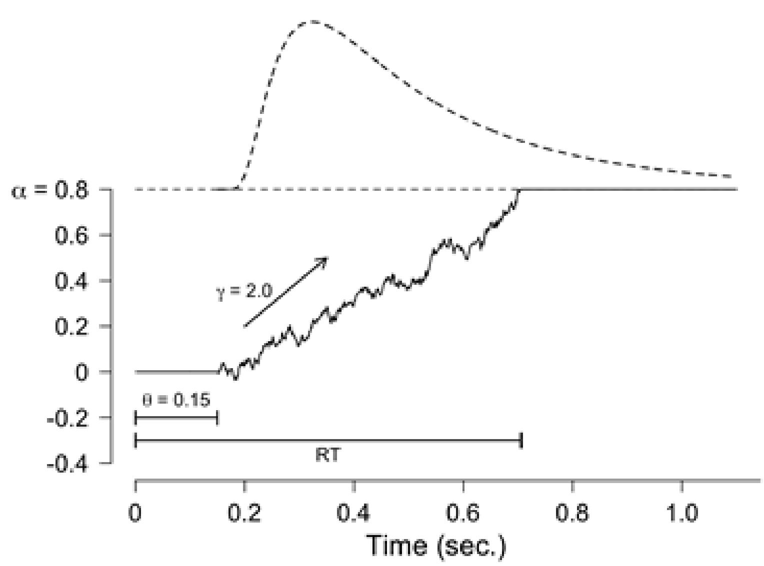

Figure 2. The shifted Wald as a cognitive model, describing RT as the time for an accumulator to drift toward and hit a single boundary α at rate γ = 2.0.

The nondecision time $\theta = 0.15$ represents the component of RT which is not due to this accumulation process. The solid black line represents the accumulator for a single trial, whereas the dashed upper curve represents the shifted Wald distribution formed by collecting RTs for many such trials.

The shifted Wald model allows for direct interpretations of response time distributions. Faulkenberry (2017) utilized this model to investigate the independence of encoding and calculation processing mental arithmetic. The manipulations in this experiment were problem size (small, large) and problem format (digit, words). By using the shifted Wald, Faulkenberry (2017) found that drift rate was affected by both problem size and format. However, response threshold and nondecision time were generally only affected by problem format, indicating that problem format impacts the calculation process. This helped to provide evidence that in mental arithmetic problems, the encoding variables (i.e., problem format) interacts with the calculation process (Campbell & Fugelsang, 2001; Frampton & Faulkenberry, 2019).

In a similar study, Faulkenberry, Vick, and Bowman (2018) fit response times from a size congruity task to the shifted Wald model. They were interested in investigating the locus of the interference that the size congruity task induces. They found a difference between congruent trials and incongruent trials for the decision related parameters: drift rate γ and response threshold α. Congruent trials had a higher drift rate γ than incongruent trials, which indicated that congruent trials are more unambiguous than incongruent trials. Congruent trials had a lower response threshold α than incongruent trials, which indicated that congruent trials induced a more liberal than incongruent trials. Critically, there was no difference in nondecision time θ, which is the parameter related to encoding. Thus, Faulkenberry et al. (2018) interpreted this finding as evidence that the interference of the size-congruity task occurs later in the decision stage of processing, rather than early in the encoding stage of processing.

EZ-Diffusion Model

The previous distributions are single boundary accumulator models, which have one response threshold. Because there is only one threshold, these distributions can only utilize the correct responses. Thus, response errors must be excluded from analysis. This is problematic because errors have the potential to hold rich information about the decision-making process. Yet, for the size congruity effect, there are two output answer choices, which yields the possibility for fitting errors and nonerrors. Therefore, researchers have used diffusion models to account for situations that lead to multiple decisions. The standard diffusion model was created by Ratcliff (1978), in which he provided a tool to mathematically fit response time data to seven parameters. These parameters are well known to be reflective of cognitive processes (Matzke, Wagenmakers, 2007; Ratcliff, Thapar, & McKoon, 2001; Ratcliff & Rouder, 2000; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008; Voss, Rothermund, & Voss, 2004).

Since then, researchers have found that while Ratcliff's (1978) diffusion model provides vast information about the response time distributions, fitting the model is very difficult. Thus, for simplicity, researchers have restricted the parameters to only three: boundary separation (response conservativeness), drift rate (the quality of information), and nondecision time (motor processes) (see Figure 3). This simplified diffusion model is known as the EZ-Diffusion model (Wagenmakers, Van Der Maas, & Grasman, 2007).

However, it is important to note that fitting the EZ-Diffusion model is fundamentally different from than maximum likelihood estimation. The EZ-Diffusion model is fit by simple mathematical formulas to summaries response time data. This process is described below.

1. First, the drift rate is calculated through the following equation:

$$\nu = sign\left(P_c - \tfrac{1}{2}\right)s\left\{\frac{logit(P_c)\left[P_c^2 logit(P_c) - P_c logit(P_c) + P_c - \frac{1}{2}\right]}{VRT}\right\}^{\frac{1}{4}},$$

   where P_c is the proportion of correct decisions, $s = 0.1$, and *VRT* is the variance of the response times for correct decisions. The logit function is defined as

$$logit(P_c) = log\left(\frac{P_c}{1 - P_c}\right).$$

2. After the drift rate has been determined, the boundary separation is determined. It is given by

$$a = \frac{s^2 logit(P_c)}{\nu}.$$

3. Following the calculation of drift rate and boundary separation, mean decision time (*MDT*) is calculated as follows:

$$MDT = \left(\frac{a}{2\nu}\right)\frac{1 - \exp(-\frac{\nu\alpha}{s^2})}{1 + \exp(-\frac{\nu\alpha}{s^2})}.$$

This is a vital step in order to calculate nondecision time. Response times are equal to the sum of the *MDT* and the nondecision time:

$$MRT = MDT + T_{er}.$$

Thus, by subtracting *MDT* from the mean response time for correct decisions (*MRT*), the remaining amount will be considered the nondecision time (Wagenmakers et al., 2007), denoted T_{er}. That is,

$$T_{er} = MRT - MDT.$$

While the EZ-Diffusion model is relatively new, this model provides estimates about the unobservable variables: drift rate, response threshold, and nondecision time. Wagenmakers and colleagues (2007) note that Ratcliff's diffusion model is not to be replaced by the EZ-Diffusion model, but rather is another tool for researchers to model response time data. If the goal is to study the relationship between correct and incorrect response times, them the Ratcliff diffusion model is the optimal tool. However, if the goal is to create estimates for unobservable variables, then the EZ-Diffusion model is the optimal tool.

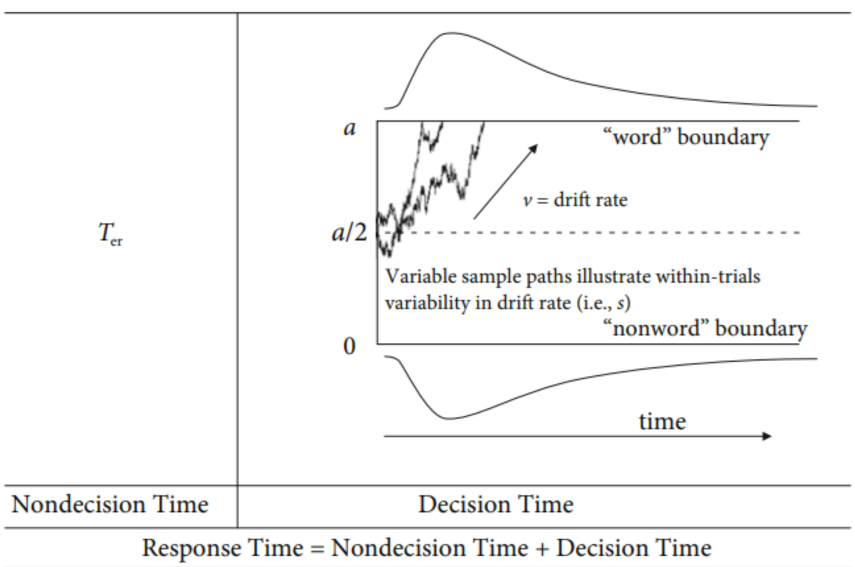

Figure 3. The EZ-diffusion model from "An EZ-diffusion model for response time and accuracy." by Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grasman, R. P. P. P., 2007, *Psychonomic Bulletin & Review*, p. 8. Reprinted with permission.

The Present Study

The purpose of the present study was to provide evidence for either an early or a late interaction account of the size-congruity effect. I fit the observed response times of the size-congruity task to four different mathematical models. Based on the calculated parameter estimates of each model, inferences were made about the underlying cognitive processes. The distributions of response times were used to gain insight onto the nature of the size-congruity effect.

I predicted the following:

1. For the ex-Gaussian model, the normal components μ, σ are believed to be reflective of the retrieval process. However, the exponential component τ is thought to be reflective of the decision-making process. If there is a notable horizontal rightward shift of the normal component between incongruent trials and congruent trials, then I can infer that the interaction occurs early in the decision-related process. However, if there is a notably a fatter tail for incongruent trials than congruent trials, then there is evidence for a late interaction model.
2. For the ex-Wald, if there are notable differences in the tail between incongruent and congruent trials, I can infer that the interference occurred in the encoding stage. This would be is evidence for an early interaction model. However, if the impact of congruency is in the Wald portion of the curve γ and α, then I can infer that the interference occurred in the final decision stage. This would be evidence for a late interaction model.

3. For the shifted Wald, the parameters α and γ of the curve are thought to be reflective of the decision-related process. If there is a notable difference in nondecision time between incongruent and congruent trials, I can infer that the interference occurred in the encoding stage. This would be evidence for an early interaction model. However, if there are notable differences between α and γ, but no difference in nondecision time θ, I can infer that the interference is occurring in the final decision stage. This would be evidence for a late interaction model.
4. For the EZ-Diffusion model, the predictions of the parameter estimate differences are similar to the shifted Wald. Additionally, even though participants commit very view errors overall (Wagenmakers, et al., 2007), I can test whether errors are systematically slower or faster than correct responses.

CHAPTER II

METHOD

Participants

I recruited 53 Tarleton State University students to participate in this experiment (20 males, 33 females, mean age 19.45 years, SD = 1.44). Students were recruited and offered partial course credit for volunteering. In order to keep participant confidentiality, participants were given an identification number and the data that was collected referred this number. Furthermore, only the participant identification number was used in data analysis.

Stimuli and Procedure

Participants were presented with pairs of single-digit Arabic numerals chosen from the stimulus set (2, 3, 4, 5, 6, 7, and 8). In order to balance numerical distance between numerals, the following 12 pairs were selected: 2-3, 3-4, 4-5 (distance 1); 2-4, 3-5, 4-6 (distance 2); 2-5, 3-6, 4-7 (distance 3); 2-6, 3-7, 4-8 (distance 4). The primary manipulation was the font size of the stimuli. The physically larger digits were presented in 36-point font, whereas the physically smaller digits were presented in 28-point font (see Figure 4).

Thus, I had two different conditions, congruent trials and incongruent trials. For congruent trials, the physical size and magnitude led to the same decision (e.g., large font and large magnitude). For incongruent trials, the physical size and magnitude led to different decisions (e.g., large font and small magnitude).

Each pair was presented in different left-right orders and different font configurations (smaller/left; larger/right; or smaller/right; larger/left). In total, there were

12 pairs x 2 congruity conditions x 2 orders x 2 font configurations = 96 experimental trials per block. Participants completed 4 blocks for a total of 384 trials per participant. The completion time for each participant was approximately 10 minutes.

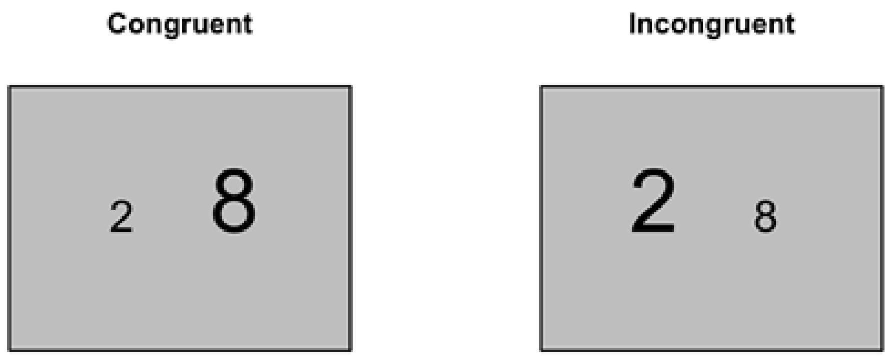

Figure 4. Example stimuli in a physical size comparison task. The left panel depicts a congruent trial, where the physically larger number (8) is also the numerically larger digit. The right panel depicts an incongruent trial, where the physically larger number (2) is the numerically smaller digit.

For this experiment I used the OpenSesame software package (Mathôt, Schreij, & Theeuwes, 2012) and ran it on a Lenovo Thinkpad X220 computer with a 12.5 inch display at a resolution of 1366 x 768 pixels. I used a standard computer keyboard for responses, in which the “A” key represented the selection of the leftmost number and the “L” key represented the selection of the rightmost number. Prior to the start of the experiment, students were instructed to select the physically larger of the two numbers that was presented on the screen. Additionally, participants were instructed to answer as quickly and as accurately as possible.

Each trial started with a fixation point that was displayed for 500 milliseconds. Immediately after, the stimulus pair was presented. Each pair remained on the screen until a response was given. If participants selected the correct response (the physically larger number), they were given no feedback and the next trial began. However, if participants selected the incorrect response, they were given a red “X” for one second and the next trial began immediately after (see Figure 5).

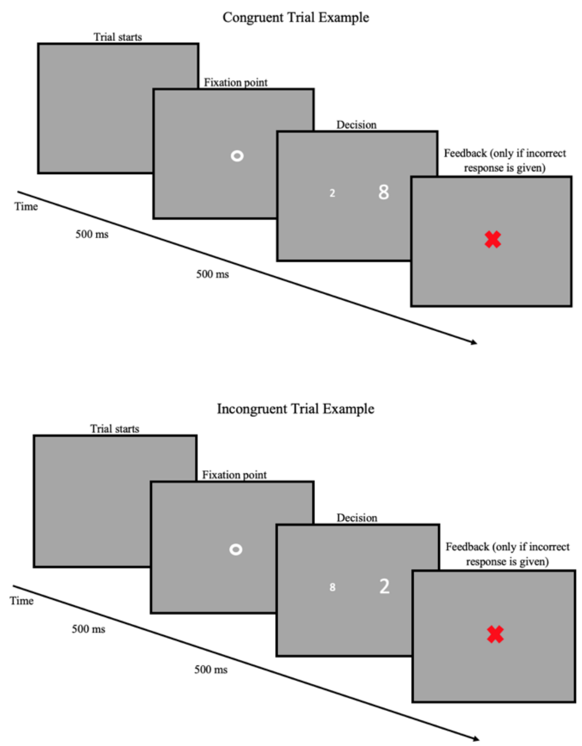

Figure 5. Sequence of experimental trials.

Upon completion of the experimental trials, participants were given a short demographic survey and thanked for their participation. All anonymous response time data was archived to GitHub, under a “born-open” data protocol (Rouder, 2015). Data sets that are archived in this manner are known as born-open data because they are automatically uploaded nightly and time stamped. Some of the main advantages of using born-open data storage are that (1) data backup is automatic; (2) it provides simplified sharing of data between labs and collaborators; and (3) it increases long-term availability of data. Therefore, the data collected from this experiment were publicly stored in a repository. However, the data were anonymous, and participants were informed that their anonymous data were open to the public via GitHub.

Analysis Plan

The response time data for this experiment were fit to a variety of mathematical models. After a cleaning procedure to remove errors and extreme outliers, the basic workflow for each model was: (1) separation into design cells; (2) maximum likelihood estimation; (3) dimension reduction; and (4) Bayesian model comparison.

1. *Separation into design cells***.** I separated individual response times into 106 design cells (53 participants x 2 trial types). In other words, I produced a factorial combination of 53 participants with 2 trial types each (correct, incorrect).
2. *Maximum likelihood estimation***.** I conducted maximum likelihood estimation to produce the 3 parameters for both congruent and incongruent trials. Thus, I had a collection of 6 parameter estimates for each participant. Note that the EZ-Diffusion model has specific equations that do not include the use of maximum likelihood estimation.
3. *Dimension reduction***.** I acquired 318 parameter estimates (106 design cells x 3 parameters), which were collapsed into 6 grand mean parameter estimates (3 congruent, 3 incongruent).

4. *Bayesian model comparison***.** I conducted a Bayesian paired samples *t*-test to compare each parameter set by trial type. Instead of using a traditional frequentist *t*-tests, I computed Bayesian *t*-tests for each parameter estimate (Rouder, Speckman, Sun, Morey, & Iverson, 2009). From this, I obtained a collection of Bayes Factors, which are likelihood ratios that provide a continuous measure of the extent to which the observed data is more likely to have occurred under one hypothesis than another (Kass & Raftery, 1995). This indexed the support for the two competing hypotheses:

(1) BF_{10} represents a Bayes factor in support of the alternative over the null, or

(2) BF_{01} represents a Bayes factor in support of the null over the alternative. This approach is useful because evidence can be found for either the alternative hypothesis or the null hypothesis, which is not something that the frequentist framework provides (Wagenmakers, 2007). Finally, for the EZ-Diffusion model, there was not a cleaning procedure of the response times because this diffusion model has two response boundaries for correct responses and incorrect responses. However, I followed the same workflow as for the other models.

CHAPTER III

RESULTS

Participants completed a total of 20,352 trials. We discarded 636 trials that contained an incorrect response (error rate = 3.125%). Further, we removed an additional 11 trials that were slower than 100 milliseconds and 71 trials that were longer than 2,000 milliseconds. This cleaning procedure resulted in retaining a total of 19,634 trials (96.47 % of original trials) for further analysis.

As previously mentioned, the general analysis plan included four steps. First, the response times were separated into design cells (53 participants x 2 trial types = 106 design cells). Second, each design cell had a distribution of response times, which were then given a three-parameter description via one of four different modeling strategies (ex-Gaussian, ex-Wald, shifted Wald, or EZ-diffusion). This resulted in 318 parameter estimates for each modeling strategy. For the third step, these estimates were collapsed along subjects, resulting in 3 grand mean parameter estimates for incongruent trials and 3 grand mean parameter estimates for congruent trials. Finally, I conducted a Bayesian paired-samples *t*-test (Rouder, Speckman, Sun, Morey, & Iversen, 2009) on each parameter pair to obtain Bayes factors (Kass & Raftery, 1995) for various hypotheses about the effects of the congruity manipulations. A Bayes factor is defined as the relative likelihood of the observed data under two competing hypotheses. That is, a Bayes factor provides a continuous measure of the extent to which the observed data is more likely to have occurred under either the alternative hypothesis or the null hypothesis. The notation BF_{10} represents the relative likelihood of the data under the alternative over the null, whereas BF_{01} represents the relative likelihood of the data under the null over the alternative.

For this study, Bayesian inference is beneficial because it allows for the ability to quantify evidence in support of either hypothesis. This gives us the ability to differentiate between “evidence of absence” and “absence of evidence”. In traditional hypothesis testing, failure to reject the null hypothesis does not provide support for a null effect. Such a failure to reject the null leads to ambiguity in the interpretation of the data; either there is no effect (evidence of absence) or there is an effect, but we did not detect it (i.e., absence of evidence, otherwise known as a Type II error). We just simply do not know. Further, traditional hypothesis testing does not give us any index of how well the alternative hypothesis predicts our observed data. In contrast, Bayes factors can be used to identify which hypothesis is likely correct. For example, a large BF_{01} is interpreted as evidence for the null hypothesis (i.e., evidence of absence). Overall, Bayesian inference provides the ability to quantify evidence for either hypothesis, a major benefit over traditional hypothesis testing.

One potential limitation of Bayes factors is that there is no universal standard to determine when there is sufficient evidence. Therefore, Bayesian inference has a set of guidelines to follow, rather than the strict thresholds of traditional inferential statistics. Research guidelines are simply suggestions, not rules (Navarro, Pitt, & Myung, 2004). Jeffreys (1961) proposed the following guidelines:

Table 1

Guidelines for Bayes factors

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

Mean Response Times

First, I analyzed the effect of the congruity manipulation on mean response times. As expected, I found evidence for the size congruity effect on mean response times in the physical comparison task (see Figure 6). The mean of the response time distribution for incongruent trials (M = 620.1, SD = 114.6) was shifted rightward compared to the distribution for congruent (M = *554.3*, SD = 104.4) trials, indicating that incongruent trials required more processing time than congruent trials. This is confirmed by a Bayes factor of $BF_{10} = 2.078 \times 10^{17}$, indicating that the observed data are approximately 2×10^{17} times more likely under the alternative hypothesis than the null hypothesis. In all, we see substantial evidence of a congruity-related increase in mean response times.

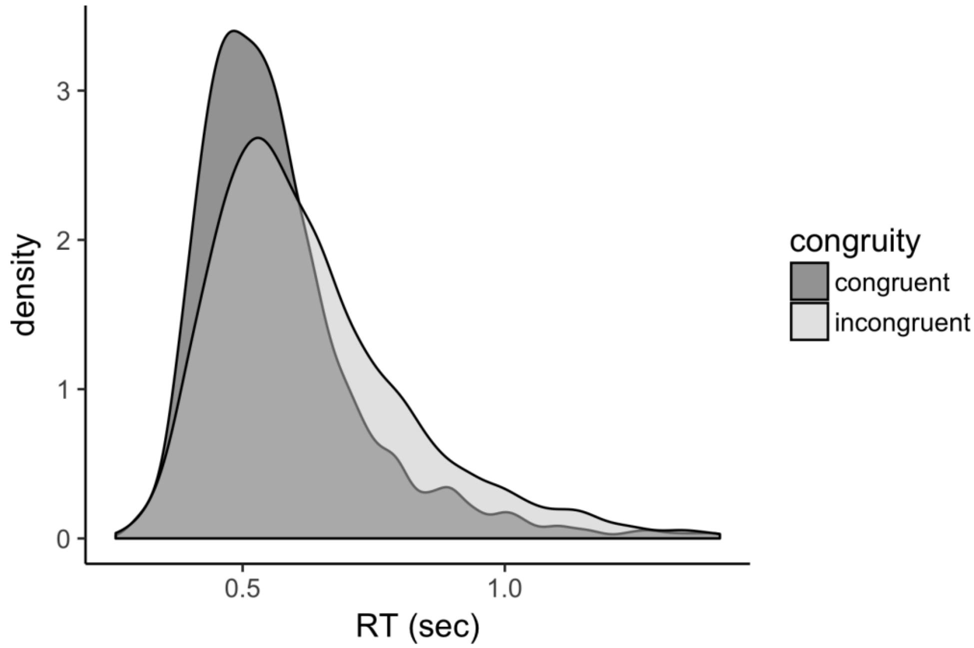

Figure 6. Distributions of response times (in seconds) as a function of congruity (congruent versus incongruent).

Next, I attempted to more fully describe the effects of physical-numerical congruity on the *distributions* of response times. To this end, I fit the distributions in each design cell with a class of mathematical models. As mentioned before, I collected 3 congruent and 3 incongruent parameter estimates for each mathematical model. Through Bayesian *t*-tests, I computed Bayes factors for the competing hypotheses about each parameter set.

Ex-Gaussian Modeling

The mean values for the mean parameter μ were larger for incongruent trials (M = 443.6) than for congruent trials (M = 425.3), BF_{10} = 910.9. This implies that the normal component of the curve was shifted more to the right for incongruent trials than congruent trials.

The mean values for the standard deviation of the normal component, σ, were larger for incongruent trials (*M* = 57.09) than for congruent trials (*M* = 46.24), BF_{10} = 307.3. This indicates that the distribution of response times for congruent trials were more concentrated around the mean compared to incongruent trials.

For the tail parameter, τ, the mean rate of the exponentially distributed tail was larger for the incongruent trials (*M* = 176.6) than for congruent trials (*M* = 128.9), BF_{10} = 1.21 x 10^{15}. This indicates that incongruent trials have a thicker tail than congruent trials.

Ex-Wald Modeling

Previous work with accumulator models (e.g., Faulkenberry et al., 2019) allows specific directional hypotheses for each parameter of the accumulator models. Based on this past work, we predict that rate of information uptake will be faster for congruent trials than for incongruent trials. So, we expect that mean drift rates will be larger for congruent trials than incongruent trials (see Figure 7.A). Regarding response threshold, we expect congruent trials to require less information before triggering a decision; thus, we expect the mean response threshold to be smaller for congruent trials than for incongruent trials (see figure 7.B). Finally, the late interaction account of the size congruity effect leads us to predict that nondecision parameters should be invariant between congruity conditions. Thus, for nondecision time, we expect no difference between congruent and incongruent trials (see Figure 7.C).

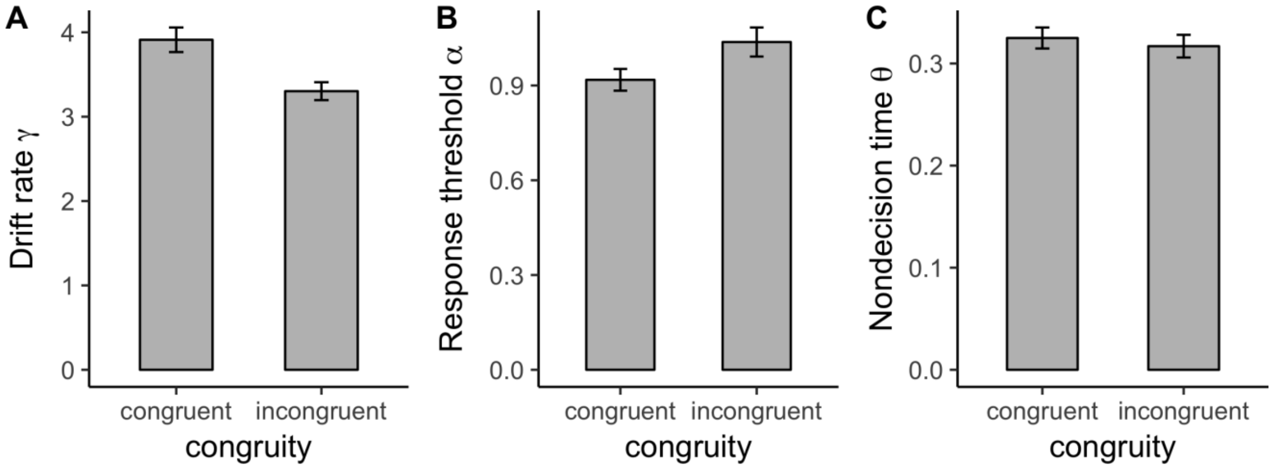

Figure 7. Example of predictions for accumulator models of each parameter as a function of congruity (congruent versus incongruent). Figure from "A shifted Wald Decomposition of the numerical size-congruity effect: Support for a late interaction account." by Faulkenberry, T., Vick, A., & Bowman, K., 2018, *Polish Psychological Bulletin 49*(4):391-397. doi: 10.24425/119507 p. 9. Reprinted with permission.

First, we consider the Ex-Wald drift rate. We found that the mean drift rate γ was larger for congruent trials (*M* = 0.4495) than for incongruent trials (M = 0.3808), BF_{10} = 147.5. This indicates that the rate of information uptake was faster on congruent trials than incongruent trials.

The mean response threshold for congruent trials (*M* = 187.9) was *not* smaller than the mean response threshold for incongruent trials (*M* = 164.5), BF_{01} = 28.78. This indicates support for a null effect of congruity on response threshold, as the observed data were approximately 28 times more likely under the null hypothesis than the alternative hypothesis.

Finally, for the tail parameter of the ex-Wald, the mean for congruent trials (*M* = 123.4) was smaller than the mean for incongruent trials (*M* = 163.8), $BF_{10} = 1.23 \times 10^6$.

This indicates that for the tail parameter there is a difference between congruent and incongruent trials. Previous work on the effects of congruity on nondecision components was done in the context of shifts rather than convolution with an exponential tail. Thus, we did not have solid predictions about the tail component. As such, the results we present here are purely exploratory rather than confirmatory.

Shifted Wald Modeling

The mean drift rate γ was larger for congruent trials (M = 0.13) than for incongruent trials (M = 0.11), BF_{10} = 1.97 x 10^6. This indicates that the rate of information accumulation from congruent trials was faster compared to incongruent trials.

The mean response threshold α was only slightly smaller for congruent trials (M = 39.68) than for incongruent trials (M = 41.28), a difference which was not different from 0, BF_{01} = 3.657. That is, the observed data were approximately 4 times more likely under the null hypothesis than the alternative hypothesis, giving us positive evidence for a null effect of congruity on response threshold.

Critically, for the mean nondecision time for congruent trials (M = 252.1) was not less than the mean nondecision time for incongruent trials (M = 238). This null effect was indexed by a Bayes factor of BF_{01} = 14.73, indicating that the observed data were approximately 15 times more likely under the null hypothesis than the alternative hypothesis. This gives us positive evidence for null effect of congruity on nondecision time, thus supporting a late interaction account of the size congruity effect.

EZ-Diffusion Modeling

This model follows the same directional hypotheses as the accumulator models. However, unlike the previous accumulator models, the EZ-Diffusion model does not utilize maximum likelihood estimation for model fitting. As described earlier, this model uses a series of equations to estimate drift rate, response threshold, and nondecision time. From these estimates, I utilized Bayesian *t*-tests to compare the mean parameters for incongruent trials and congruent trials.

The mean drift rate γ was larger for congruent trials (*M* = 0.01102) than for incongruent trials (*M* = 0.009511), $BF_{10} = 1.31 \times 10^{11}$. This indicates that the rate of information accumulation for congruent trials was larger than for congruent trials.

The mean response threshold α was smaller for congruent trials (*M* = 5.489) than for incongruent trials (*M* = 6.283), $BF_{10} = 3.21 \times 10^{12}$. This indicates that participants required more information before making a decision on incongruent trials.

Finally, the mean nondecision time for congruent trials (*M* = 296.2) was not smaller than the mean nondecision time for incongruent trials (*M* = 282). This null effect was indexed by a Bayes Factor of $BF_{01} = 20.41$, indicating that the observed data were approximately 20 times more likely under the null hypothesis than the alternative hypothesis. This strong evidence for a null effect of congruity on nondecision time supports a late interaction account of the size congruity effect.

Chapter IV

DISCUSSION

The purpose of the present study was to use mathematical modeling of response times to compare two theoretical models of the size-congruity effect. This size-congruity effect is a classic phenomenon in numerical cognition (Besner & Coltheart, 1979; Henik & Tzelgov, 1982) that is similar to the well-known Stroop effect. When participants are presented with two numbers of differing physical size and are asked to choose the physically-larger digit from the pair, an interference effect occurs. People's responses are quite fast when the pair is *congruent* – that is, when the physically-larger number also has the larger number magnitude. For example, when choosing the physically-largest from 2 versus 8, the physically-larger digit also has the larger numerical magnitude. However, when pairs are *incongruent*, responses are markedly slowed down. In this case, the physically-larger digit has the smaller numerical magnitude (i.e., 2 versus 8). Here, the comparisons based on physical size and numerical magnitude lead to two different decisions, thus reflecting an interference effect. This is remarkable, as the task only requires participants to choose the physically-larger stimulus. Conceivably, this task could be completed by completely ignoring the semantic information of numerical magnitude and simply focusing on the visual template of the stimuli. However, the presence of the size-congruity effect is strong evidence that people simply cannot ignore this numerical information. Thus, the size-congruity effect is a classic empirical marker of the automatic processing of numerical magnitude (Henik & Tzelgov, 1982).

The size-congruity effect and its tie to automatic processing is of considerable interest in a wide array of studies in mathematical cognition. However, my interest in this phenomenon goes beyond focusing on the fact that a simple congruity manipulation leads to shorter response times for congruent trials than incongruent trials. Though the effect has been demonstrated many times over the past 40 years, there is little agreement over *why* the effect occurs. Early attempts to uncover the cognitive mechanisms behind the size-congruity effect led Santens and Verguts (2011) to propose two classes of explanations for the interference. The first of these is an *early interaction account*, where the interference occurs at early stages of processing (i.e., encoding) and reflects a shared representation of numerical and physical size (e.g., the ATOM theory of Walsh, 2003; Townsend & Ashby, 1983). As an alternative, other researchers have proposed theories that Santens and Verguts classified as *late interaction accounts* (Santens & Verguts, 2011; Faulkenberry et al., 2016; 2019; Schwarz & Heinze, 1998). In these accounts, the interference reflects dynamic competition between parallel and partially active responses during a decision. As such, these accounts propose that the interference exhibited in the size-congruity effect stems not from early shared representations at encoding, but rather from interaction of competing decisions *during* the response. The broader purpose of my thesis was to use mathematical modeling to disentangle these two competing accounts of the size-congruity effect.

To investigate these competing explanations for the size-congruity effect, I used a class of response time models. A classic model for response times is the ex-Gaussian distribution, chosen because of its pronounced positive skew. This distribution contains a normal component combined with an exponential distribution, resulting in the similar skewed appearance typical with response time distributions. This model contains three parameters which various researchers have claimed to be reflective of various underlying cognitive processes (e.g., Hohle, 1965). However, these parameters should be viewed as solely a descriptive tool (Schwarz, 2001; Matzke & Wagenmakers, 2009). To gain further insight on cognitive processes, I utilized two models based on a Wald distribution, which is the distribution of the time required for a continuous diffusion process with positive drift to hit a fixed boundary. The Wald distribution provides a mathematical way to represent decision-making as a collection of information accumulation processes (Link, 1975). Specifically, I used two modifications of the Wald distribution for this study – the ex-Wald and the shifted Wald – each chosen to more accurately reflect properties of response time distributions. The ex-Wald is an exponentially modified Wald distribution; the convolution of these two distributions provides a rightward shift so that the distribution begins at zero (Schwarz, 2001; Heathcote, 2004). It is important that these models do not begin at zero because it is unrealistic for response times to be zero (Luce, 1986). As such, the ex-Wald is more theoretically sound than the ex-Gaussian model, but it is not widely used in the literature, potentially because of its complexity. Therefore, I also used the shifted Wald, which mathematically easier to represent than the ex-Wald because it is a constant shift of the entire distribution rather than an averaging of two curves (Anders, Alario & Van Maanen, 2016).

From each of these models, I produced parameter estimates for each participant in each condition. To calculate each of these parameters for each of the design cells, I used maximum likelihood estimation to calculate the most likely parameter values given the observed data (Myung, 2003). In total, I collected 6 parameter estimates for each participant for each of the three response time models (ex-Gaussian, ex-Wald, and shifted Wald). In total, I collected 318 parameter estimates per model, and this collection of parameter estimates was then collapsed into 6 grand means (3 congruent, 3 incongruent). Finally, I compared specific hypotheses about the effects of the congruity manipulation on each parameter with Bayes factors, which index the relative likelihood of one hypothesis over the other.

This workflow was effective for the first three mathematical models in this study. Yet, the ex-Gaussian and the accumulator models can only fit correct responses. The size congruity task yields the possibility of correct responses and incorrect responses. While the error rate for this experiment was relatively low (error rate = 3.125%) there could be rich information about how people process numbers within the error rate portion of the data. One solution to "throwing away" potentially beneficial data is to use a *dual-boundary* diffusion model (Ratcliff, 1978). Thus, I chose to also use the EZ-diffusion model (Wagenmakers et al., 2007). This model is fundamentally different than the other models because it does not utilize maximum likelihood estimation. Instead, the EZ-diffusion model is fit with a set of relatively simple mathematical equations applied to the descriptive statistics of the observed data (i.e., mean and variance of the response times and the proportion of correct trials). Like the shifted Wald, the parameters of the EZ-diffusion model are boundary separation, drift rate, and nondecision time. These parameters are known to index the underlying cognitive processes of decision making (Matzke & Wagenmakers, 2009).

Each of the mathematical models were used to decompose 53 participants response times into a collection of parameter estimates. Note that the overall goal of collecting the parameters was to provide evidence for either an early interaction account (i.e., interference occurs at the *encoding* stage) or a late interaction account (i.e., interference occurs at the *decision* stage). Thus, either the interference occurs in the decision components (e.g., drift rate, response threshold) or in the nondecision components (e.g., nondecision time).

Interpretation of Results

As I expected, I found a large effect of physical-numerical congruity on mean response times. Congruent trials were processed faster than incongruent trials. Beyond this, I was more concerned with describing the effects of physical-numerical congruity on the entire *shape* of the distribution, rather than focusing simply on the mean response time.

For the ex-Gaussian distribution, I found an effect of congruity for all three parameters. The mean parameter μ was smaller for congruent trials than incongruent trials. This indicates that the normal component of the curve is more leftward for congruent trials than incongruent trials. The standard deviation parameter σ was smaller for congruent trials than incongruent trials, showing that there was less deviation around the mean for congruent trials than incongruent trials. Finally, the tail parameter τ was also smaller for congruent trials than incongruent trials. This indicates that the tail of distribution for congruent trials was thinner than the tail of the distribution for incongruent trials. Thus, for both the congruent trials and incongruent trials, the distribution of response times had a higher density on the left that gradually decreased over time with a long positive tail on the right. However, the congruent trials were positioned more leftward than incongruent trials (see Figure 6).

For the ex-Wald, the shifted Wald, and the EZ-Diffusion models, I found a consistent effect of the congruity manipulation on drift rate γ, where congruent trials have a faster rate of information uptake than incongruent trials. However, I found some inconsistency in my results for response threshold α. That is, that all three models did not have effects on response threshold. Recall that this parameter indexes the amount of information accumulation required for a response. Previous work has found that that the amount of information required for a response is lower for congruent trials is than for incongruent trials. However, my outcomes for response threshold were not the same across all of the mathematical models in this study. For the shifted Wald, I found evidence for a null effect of the congruity manipulation on response threshold. Yet, for both the ex-Wald and the EZ-diffusion model, I found the expected negative effect of congruity for response threshold α, where congruent trials required less information to be gathered before making a decision than incongruent trials. In all, these results support the notion that the interference of the size-congruity effect occurs in the decision-related parameters. However, does it also occur in the nondecision parameters?

It turns out that the answer is "no." For the shifted Wald and EZ-Diffusion models I found positive evidence for a null effect of the congruity manipulation on nondecision time θ. That is, our data were many times more likely under a hypothesis of no congruity effect than they were under a hypothesis where incongruent trials increase the time required for nondecision-related processing. As previously mentioned, the parameter of nondecision time indexes encoding processes and motor movement of reaction time, not the decision-related components of response time. Thus, these data support a late interaction account of the size congruity effect.

In all, I found evidence that the locus of the size-congruity effect occurs in the decision-related stages of processing and not in the encoding-stages. This conclusion is in line with several other studies (Faulkenberry et al., 2016; Sobel et al., 2016, 2017). These data provide further evidence for the late interaction account that was proposed by Santens and Verguts (2011). Furthermore, the present study is novel in its use of mathematical modeling to more fully describe response time distributions. It is a common practice in psychological science to simply use the mean and standard deviation to make inferences about participants' cognitive abilities. However, by collapsing data into simply the mean and the standard deviation, there is rich information lost. On the other hand, mathematical models allow for the entire shape of the distribution to be utilized.

One of the main limitations of my results is the inconsistency in the observed congruity effects on response threshold – particularly the surprising null effect in the shifted Wald model. Based on the literature, I predicted that the response threshold for congruent trials would be less than the response threshold for incongruent trials because the information that is needed to be accumulated for congruent trials should be less than that for incongruent trials. While this is not the result that I expected, the evidence for this null effect was relatively small. This unexpected finding has the potential to be further investigated. Additionally, there is a debate in the field about whether these parameters accurately reflect the underlying cognitive processes. It is clear for the ex-Gaussian that we should avoid the temptation of interpretation (Schwarz, 2001; Matzke & Wagenmakers, 2009), but for the Ex-Wald and shifted Wald models, there is evidence that the parameters can be interpreted as specific processing stages of the decision-making process (Schwarz, 2001; Anders, Alario & Van Maanen, 2016). Furthermore, the EZ-diffusion model is highly credible as indexing the underlying cognitive functions because it is based on Ratcliff's diffusion model (Matzke & Wagenmakers, 2007;

Ratcliff, Thapar, & McKoon, 2001; Ratcliff & Rouder, 2000; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008; Voss, Rothermund, & Voss, 2004). I was able to use the EZ-Diffusion model to confirm my predictions. For the most part, the results from the EZ-diffusion model matched with those from the ex-Wald and the shifted Wald. The similarities of these results between the accumulator models and the diffusion model provides further validation of the use of maximum likelihood estimation as a tool to uncover information about cognitive processing.

Another broad implication of this study is that it provides evidence about the nature of processing numerical values. As previously mentioned, the instructions for the size-congruity task were to only select the physically larger symbol. If participants were able to ignore numerical magnitude and only process physical size, then congruent and incongruent trials would have the same response time. However, it is clear that incongruent trials require more time to process than congruent trials. This is due to the cognitive interference that numerical value has on physical size (Besner & Coltheart, 1979). In other words, I have provided further evidence that processing numerical magnitude, even when a task does not require it, is unavoidable.

One future extension of this study would include the use of other mathematical models to investigate the locus of interference in the size-congruity effect. For example, I adapted the functions used to fit the shifted Wald from Farrell and Lewandowsky (2018). An alternative method of fitting the shifted Wald comes from Anders et al. (2016). It would be interesting to compare the outcomes between the two methods to see if there is a difference between them. Ideally, these two methods would lead to the same outcome, but this would need to be further investigated.

Another future study would include investigating this interference in children. While many studies have investigated cognitive interference in children (Ashkenazi, Mark-Zigdon, Henik, 2009; Ansari, Fugelsang, Dhital, & Venkatraman, 2006; Hervey et al., 2006), I would like to implement the use of mathematical modeling of response times. The size-congruity task is simple enough that elementary school children could participate. This would be a worthwhile project because I have two opposite predictions. The first would be that children exhibit more of a congruity effect because they would require more response time on incongruent trials than congruent trials due to their limited experience with numbers. On the other hand, children may exhibit less of a congruity effect (if any) because their limited experience would lead to following the task instruction of identifying the *physically* larger number better than adults. That is, children might be resistant to automatic processing.

From a wider scope, the size-congruity effect is reflective of cognitive processing in general. This is just one of many experimental tasks that have the potential to reflect underlying cognitive processes. By learning how the process of interference occurs, psychologists could also learn how to inhibit the interference. The goal of cognitive psychology is to learn about how people process information and this thesis provides further insight toward this goal. Furthermore, this study is unique because it brings together the fields of psychology and mathematics. There is a major gap between the fields of mathematics and psychology, which is counter-intuitive because the field of psychological sciences has long utilized mathematical tools (e.g., statistical inference). Through the use of mathematical models, I provide evidence that the cognitive interference of the size-congruity event occurs in the decision-related stages. This approach could be used to investigate similar phenomena across the field.

REFERENCES

Algom, D., Dekel, A., & Pansky, A. (1996). The perception of number from the separability of the stimulus: The Stroop effect revisited. *Memory & Cognition, 24*(5), 557–572. doi:10.3758/bf03201083

Anders, R., Alario, F.-X., & Van Maanen, L. (2016). The shifted Wald distribution for response time data analysis. *Psychological Methods, 21*(3), 309–327. doi:10.1037/met0000066

Ansari, D., Fugelsang, J. A., Dhital, B., & Venkatraman, V. (2006). Dissociating response conflict from numerical magnitude processing in the brain: An event-related fMRI study. *NeuroImage, 32*(2), 799–805. doi:10.1016/j.neuroimage.2006.04.184

Ashkenazi, S., Mark-Zigdon, N., & Henik, A. (2009). Numerical distance effect in developmental dyscalculia. *Cognitive Development, 24*(4), 387–400. https://doi-org.zeus.tarleton.edu/10.1016/j.cogdev.2009.09.006

Atkinson, R. C., Holmgren, J. E., & Juola, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics, 6*(6), 321–326. doi:10.3758/bf03212784

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia, 17*(5), 467–472. doi:10.1016/0028-3932(79)90053-8

Borgmann, K., Fugelsang, J., Ansari, D., & Besner, D. (2011). Congruency proportion reveals asymmetric processing of irrelevant physical and numerical dimensions in the size congruity paradigm. *Canadian Journal of Experimental*

Psychology/Revue Canadienne de Psychologie Expérimentale, 65(2), 98–104. doi:10.1037/a0021145

Campbell, J. I. D., & Fugelsang, J. (2001). Strategy choice for arithmetic verification: effects of numerical surface form. *Cognition, 80*(3), B21–B30. doi:10.1016/s0010-0277(01)00115-9

Cohen Kadosh, R., & Henik, A. (2006). A common representation for semantic and physical properties. *Experimental Psychology, 53*(2), 87–94. doi:10.1027/1618-3169.53.2.87

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *The American Journal of Psychology, 105*(2), 239. doi:10.2307/1423029

Dawson, M. R. W. (1988). Fitting the ex-Gaussian equation to reaction time distributions. *Behavior Research Methods, Instruments, & Computers, 20*(1), 54–57. doi:10.3758/bf03202603

Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and Behavior*. New York, NY: Cambridge University Press. doi:10.1017/cbo9781316272503

Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review, 15*(6), 1209–1217. doi:10.3758/pbr.15.6.1209

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. doi:10.1016/j.actpsy.2015.11.010

Faulkenberry, T. J. (2017). A single-boundary accumulator model of response times in an addition verification task. *Frontiers in Psychology, 8* (1225). doi:10.3389/fpsyg.2017.01225

Faulkenberry, T. J., & Frampton, A. R., (2018) Mental arithmetic processes: Testing the independence of encoding and calculation. *Journal of Psychological Inquiry, 22*, 30-35.

Faulkenberry, T., Vick, A., & Bowman, K. (2018). A shifted Wald Decomposition of the numerical size-congruity effect: Support for a late interaction account. *Polish Psychological Bulletin 49*(4):391-397. doi: 10.24425/119507

Fitousi, D., & Algom, D. (2006). Size congruity effects with two-digit numbers: Expanding the number line? *Memory & Cognition, 34*(2), 445–457. doi:10.3758/bf03193421

Gardner, G. T. (1973). Evidence for independent parallel channels in tachistoscopic perception. *Cognitive Psychology, 4*(1), 130–155. doi:10.1016/0010-0285(73)90009-1

Heathcote, A., & Hayes, B. (2012). Diffusion versus linear ballistic accumulation: Different models for response time with different conclusions about psychological mechanisms? *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 66*(2), 125–136. doi:10.1037/a0028189

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. doi:10.3758/bf03202431

Hervey, A. S., Epstein, J. N., Curry, J. F., Tonev, S., Eugene Arnold, L., Keith Conners, C. & Hechtman, L. (2006). Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychology, 12*(2), 125–140. doi:10.1080/09297040500499081

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90(*430), 773. doi:10.2307/2291091

Kóbor, A., Takács, Á., Bryce, D., Szűcs, D., Honbolygó, F., Nagy, P., & Csépe, V. (2015). Children with ADHD show impairments in multiple stages of information processing in a Stroop task: An ERP study. *Developmental Neuropsychology, 40*(6), 329–347. doi:10.1080/87565641.2015.1086770

Krause, F., Bekkering, H., Pratt, J., & Lindemann, O. (2016). Interaction between numbers and size during visual search. *Psychological Research, 81*(3), 664–677. doi:10.1007/s00426-016-0771-4

Luce, R. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization.* New York, NY: Oxford University Press

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163–203. doi:10.1037/0033-2909.109.2.163

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. doi:10.3758/s13428-011-0168-7

Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16*(5), 798–817. doi:10.3758/pbr.16.5.798

Moyer, R. S., & Landauer, T. K. (1967). time required for judgements of numerical inequality. *Nature, 215*(5109), 1519–1520. doi:10.1038/2151519a0

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology, 49*(1), 47–84. doi:10.1016/j.cogpsych.2003.11.001

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108. doi:10.1037/0033-295x.85.2.59

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 127–140. doi:10.1037/0096-1523.26.1.127

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging, 16*(2), 323–341. doi:10.1037/0882-7974.16.2.323

Reike, D., & Schwarz, W. (2017). Exploring the origin of the number-size congruency effect: Sensitivity or response bias? *Attention, Perception, & Psychophysics, 79*(2), 383–388. doi:10.3758/s13414-016-1267-4

Rieger, T., & Miller, J. (2019). Are model parameters linked to processing stages? An empirical investigation for the ex-Gaussian, ex-Wald, and EZ diffusion models. *Psychological Research*. doi:10.1007/s00426-019-01176-4

Risko, E. F., Maloney, E. A., & Fugelsang, J. A. (2013). Paying attention to attention: evidence for an attentional contribution to the size congruity effect. *Attention, Perception, & Psychophysics, 75*(6), 1137–1147. doi:10.3758/s13414-013-0477-2

Rouder, J. N. (2015). The what, why, and how of born-open data. *Behavior Research Methods, 48*(3), 1062–1069. doi:10.3758/s13428-015-0630-z

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. doi:10.3758/pbr.16.2.225

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences, 105*(16), 5975–5979. doi:10.1073/pnas.0711295105

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology, 7*(2), 191–218. doi:10.1016/0022-2496(70)90044-1

Santens, S., & Verguts, T. (2011). The size congruity effect: Is bigger always more? *Cognition, 118*(1), 94–110. doi:10.1016/j.cognition.2010.10.014

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

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers, 33*(4), 457–469. doi:10.3758/bf03195403

Schwarz, W., & Eiselt, A. K. (2012). Numerical distance effects in visual search. *Attention, Perception, & Psychophysics, 74*(6), 1098–1103. doi:10.3758/s13414-012-0342-8

Sobel, K. V., Puri, A. M., & Faulkenberry, T. J. (2016). Bottom-up and top-down attentional contributions to the size congruity effect. *Attention, Perception, & Psychophysics, 78*(5), 1324–1336. doi:10.3758/s13414-016-1098-3

Sobel, K. V., Puri, A. M., Faulkenberry, T. J., & Dague, T. D. (2017). Visual search for conjunctions of physical and numerical size shows that they are processed independently. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 444–453. doi:10.1037/xhp0000323

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. doi:10.1037/h0054651

Szűcs, D., & Soltész, F. (2007). Event-related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. *Neuropsychologia, 45*(14), 3190–3202. doi:10.1016/j.neuropsychologia.2007.06.013

Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics, 10*(3), 161–163. doi:10.3758/bf03205778

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* Cambridge: Cambridge University Press.

Turconi, E., Campbell, J. I. D., & Seron, X. (2006). Numerical order and quantity processing in number comparison. *Cognition, 98*(3), 273-285. doi:10.1016/j.cognition.2004.12.002

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14*(5), 779–804. doi:10.3758/bf03194105

Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3–22. doi:10.3758/bf03194023

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. J*ournal of Memory and Language, 58*(1), 140–159. doi:10.1016/j.jml.2007.04.006

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*(11), 483–488. doi:10.1016/j.tics.2003.09.002

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*(3), 475–482. doi:10.1007/bf03395630

Vadillo, M.A., Garaizar, P. (2016) The effect of noise-induced variance on parameter recovery from reaction times. *BMC Bioinformatics 17*(147) doi:10.1186/s12859-016-0993-x

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32*(7), 1206–1220. doi:10.3758/bf03196893