APPLICATION OF SYSTEMS FACTORIAL TECHNOLOGY AND HIERARCHICAL BAYESIAN MODELING TO CHUNKING IN WORKING MEMORY

A Dissertation presented to the Faculty of the Graduate School at the University of Missouri – Columbia

In Partial Fulfilment of the Requirements for the Degree

Doctor of Psychology

by

JONATHAN THIELE

Dr. Jeffrey Rouder, Thesis Supervisor

July 2015
The undersigned, appointed by the dean of the Graduate School, have examined the
dissertation entitled

APPLICATION OF SYSTEMS FACTORIAL TECHNOLOGY AND HIERARCHICAL
BAYESIAN MODELING TO CHUNKING IN WORKING MEMORY

presented by Jonathan Thiele,
a candidate for the degree of Doctor of Psychology,
and hereby certify that, in their opinion, it is worthy of acceptance.

---------------------------------
Professor Nelson Cowan

---------------------------------
Professor Clintin Stober

---------------------------------
Professor Paul Speckman

---------------------------------
Professor Jeffrey Rouder
DEDICATION

I wish to thank my friends and family for the support they have provided to me every step of the way. In particular, I wish to thank my Mother, Debra Thiele, for everything she has done for me and I wish to dedicate this work to her in her memory.
I wish to acknowledge my advisor, Dr. Jeff Rouder, for his help in developing this approach, refining the analysis, and composing this final paper, in addition to all of the training he has given me in the years leading up to this paper. This project would most likely have never occurred if not for his influence.

Additionally, I wish to thank Dr. Paul Speckman for his advise that led to the development of the second statistical framework discussed in this paper. If not for his suggestion, I would probably still be debugging the programs that I use for analysis. I would also like to thank Dr. Nelson Cowan, Dr. Tim Ricker, and April Swagman for their opinions and insights into the experimental method that I employed and into the interpretation of my results.

Further thanks go to Dr. Houpt of Wright State University and Dr. Fific of Grand Valley State University for providing additional motivation for finishing this research as well as to visiting graduate student Julia Haaf, undergraduate research assistants Ryan Bahr, Tony Sun, Emily Lind, David Barnes, Elliot Cade, and Dennis Magdato, and my Father, Ivan Thiele, and Grandmother, Betty Myers, for listening to my explanations and helping me practice explaining the ideas contained in this volume.
Contents

Acknowledgements ......................................................... ii
List of Figures ........................................................... vii
List of Tables ............................................................... xiv
Abstract ............................................................................. xviii

1 Introduction ................................................................. 1
  1.1 Processing Architectures ............................................. 3
  1.2 Systems Factorial Technology ...................................... 6
  1.3 Simplification of the Method ........................................ 7
  1.4 Prior Extensions of SFT to WM ................................. 9
  1.5 Chunking as Coactive Processing ............................... 9
  1.6 Change-Detection Experiments ................................. 9

2 Details of Analysis ........................................................ 13
  2.1 Log-Normal Framework ............................................ 13
  2.2 Normal Framework .................................................. 15
  2.3 Bayes Factors .......................................................... 17
  2.4 Savage-Dickey ........................................................... 21
  2.5 Conditional Marginal Density Estimation ..................... 23
  2.6 Full Conditional Posteriors in the Normal Framework .... 23
      2.6.1 Serial Model ................................................... 26
2.6.2 General Model .................................................. 26
2.6.3 Parallel Model .................................................. 27
2.6.4 Coactive Model ................................................ 27
2.6.5 Joint Distribution and Posterior Derivation of Common Parameters 27
2.6.6 Specific Posterior Derivations ............................. 35
2.7 Density Derivations ............................................. 41
2.8 Issues ......................................................... 42

3 Additional Frameworks ........................................ 44
3.1 Ex-Gaussian Framework ..................................... 44
3.2 Gamma Framework ............................................ 46
  3.2.1 Serial Model ................................................. 48
  3.2.2 General Model .............................................. 48
  3.2.3 Parallel Model .............................................. 49
  3.2.4 Coactive Model ............................................. 49
  3.2.5 Joint Distribution and Posterior Derivation of Common Parameters 49
  3.2.6 Specific Posterior Derivations ......................... 51
3.3 Density Derivations ........................................ 54

4 Methods and Results .......................................... 56
4.1 Method Summary ............................................. 56
  4.1.1 Stimuli Used ................................................ 56
  4.1.2 General Method .......................................... 57
  4.1.3 Predictions ............................................... 60
4.2 Analysis Summary ........................................... 60
  4.2.1 Prior Settings ............................................. 60
  4.2.2 Chain Runtimes and Output ............................ 63
4.2.3 Simulation Analyses ........................................... 65
4.2.4 Other Analytic Approaches ................................. 66
4.3 Changes from the Original Proposal ....................... 67
4.4 Experiment 1: Screwhead Perceptual Task ............... 70
  4.4.1 Participants .................................................... 70
  4.4.2 Data Cleaning .................................................. 70
  4.4.3 Stimuli ......................................................... 70
  4.4.4 Procedure ..................................................... 71
  4.4.5 Results ......................................................... 71
4.5 Experiment 2: Screwhead Memory Task ................... 72
  4.5.1 Participants .................................................... 72
  4.5.2 Data Cleaning .................................................. 74
  4.5.3 Stimuli ......................................................... 74
  4.5.4 Procedure ..................................................... 75
  4.5.5 Results ......................................................... 75
4.6 Experiment 3: 2-Digit Memory Task ....................... 77
  4.6.1 Participants .................................................... 77
  4.6.2 Data Cleaning .................................................. 78
  4.6.3 Stimuli ......................................................... 78
  4.6.4 Procedure ..................................................... 78
  4.6.5 Results ......................................................... 78
4.7 Experiment 4: 2-Digit Perception Task .................... 81
  4.7.1 Participants .................................................... 81
  4.7.2 Data Cleaning .................................................. 81
  4.7.3 Stimuli ......................................................... 81
  4.7.4 Procedure ..................................................... 81
List of Figures

1.1 A graphical depiction of the three processing architectures (from top to bottom): Serial processing, Parallel processing, and Coactive processing. . . 4

1.2 Example CDFs for reaction times for the two different single-signal trials (black and red) and the CDF of the minimum of the two single-signal trial reaction times, which would correspond to the CDF of the double-signal trials under parallel processing. . . . . . . . . . . . . . . . . . . . . . . . . . 5

1.3 A. Examples of the standard (study) and sample (test) nested-squares stimuli that would be used in the proposed change-detection tasks along with their associated responses. B. The crosstable used to describe Systems Factorial Technology adapted to show examples of the sample stimuli that would be used in the change-detection variant of this task. The dashed lines separate the stimuli associated with a YES response above and to the right of the dashed lines. . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2.1 Prior (solid black) and Posterior (dashed green) densities of a parameter $\theta$ for model $M_1$. The Savage-Dickey ratios for the specified restrictions are also provided. A. For $\theta_0 = .3$, the likelihood of $\theta$ is greater under the prior than under the posterior, suggesting that the restriction is not well-supported. B. For $\theta_0 = 2$, the likelihood of $\theta$ is greater under the posterior than under the prior, providing evidence favoring the restriction. . . . . . 22
4.1 Examples of the two stimuli used in this research: Screwhead stimuli (left), which vary in size and the orientation of the dividing line, and Digit Pairs (right), which vary from 1 through 9 for each digit.

4.2 General Trial Schematic of the Perceptual experiment trials (with Screwhead stimuli for an example). Participants see a fixation cross at the center before seeing both stimuli, which remain presented until the participant provides a response.

4.3 General Trial Schematic of the Memory experiment trials (with Screwhead stimuli for an example). Participants see a fixation cross at the center before seeing one stimulus for a short interval (200 ms listed here, but might differ in specific experiments). After this interval, the screen is left blank for another waiting period (1000 ms listed here, but might differ in specific experiments) before the second stimulus is presented, with the second stimulus displayed until a response is provided.

4.4 Hyperpriors for the hierarchical mean (top-left) and variance (top-right) of the interaction parameters in units of seconds and squared seconds, respectively. The bottom plots describe the conditional and marginal densities of the hierarchical interaction mean parameter and the interaction parameters, respectively, (left) and the marginal prior for the interaction parameters (in seconds) when scaled (by 4 for the Normal framework or by $4\theta$ for the Gamma framework, both of which are assumed to be dimensionless constants) to match the scale of the surface MIC estimates.

4.5 An example of how perceptual contrast effects could negatively impact any attempt to use Systems Factorial methods with Nested Square stimuli.

4.6 Examples of the plaids stimuli and the magnitudes of change considered.
4.7 A. Surface-Normal plot for all four models. Serial model estimates (black) are included for completeness. General model estimates (red), Parallel model estimates (green) and Coactive model estimates (blue) all increase monotonically with the surface MIC estimates, albeit with substantial shrinkage.

B. Surface-Gamma plot for all four models.

C. Normal-Gamma plot for all four models.

D. Maximum Absolute Autocorrelation plot across all individual interaction parameters within each model. . . . . . . . . . . . . 73

4.8 A. Surface-Normal plot for all four models. Serial model estimates (black) are included for completeness. General model estimates (red), Parallel model estimates (green) and Coactive model estimates (blue) all increase monotonically with the surface MIC estimates, albeit with substantial shrinkage.

B. Surface-Gamma plot for all four models.

C. Normal-Gamma plot for all four models.

D. Maximum Absolute Autocorrelation plot across all individual interaction parameters within each model. . . . . . . . . . . . . 76

4.9 A. Surface-Normal plot for all four models. Serial model estimates (black) are included for completeness. General model estimates (red), Parallel model estimates (green) and Coactive model estimates (blue) all increase monotonically with the surface MIC estimates, albeit with substantial shrinkage.

B. Surface-Gamma plot for all four models.

C. Normal-Gamma plot for all four models.

D. Maximum Absolute Autocorrelation plot across all individual interaction parameters within each model. . . . . . . . . . . . . 80
4.10 **A.** Surface-Normal plot for all four models. Serial model estimates (black) are included for completeness. General model estimates (red), Parallel model estimates (green) and Coactive model estimates (blue) all increase monotonically with the surface MIC estimates, albeit with substantial shrinkage. **B.** Surface-Gamma plot for all four models. **C.** Normal-Gamma plot for all four models. **D.** Maximum Absolute Autocorrelation plot across all individual interaction parameters within each model.

A.1 **A.** Marginal mean accuracies (pre-cleaning) by condition in Experiment 1. The change in angle is indicated by the x-axis while the change in size is indicated by the color of the points, with blue corresponding to no change, green corresponding to a small change, and red corresponding to a large change. The four points corresponding to the double-change trials are further highlighted with black borders. **B.** Surface MIC estimates from Experiment 1, sorted by magnitude, with error bars that correspond to the 80% confidence intervals of the means. The points and error bars are colored with regard to sign: green for negative, red for positive, and blue for intervals containing zero. **C.** Marginal mean reaction times (post-cleaning) by condition in Experiment 1. The arrangement, coloration, and highlighting are as in part A. **D.** Posterior means of MIC estimates from the General model in the Normal framework for Experiment 1, sorted by the magnitude of the surface MIC estimates in part B. Error bars correspond to the 90% posterior credible interval.
A.2  **A.** Marginal mean accuracies (pre-cleaning) by condition in Experiment 2. The change in angle is indicated by the x-axis while the change in size is indicated by the color of the points, with blue corresponding to no change, green corresponding to a small change, and red corresponding to a large change. The four points corresponding to the double-change trials are further highlighted with black borders. **B.** Surface MIC estimates from Experiment 2, sorted by magnitude, with error bars that correspond to the 80% confidence intervals of the means. The points and error bars are colored with regard to sign: green for negative, red for positive, and blue for intervals containing zero. **C.** Marginal mean reaction times (post-cleaning) by condition in Experiment 2. The arrangement, coloration, and highlighting are as in part A. **D.** Posterior means of MIC estimates from the General model in the Normal framework for Experiment 2, sorted by the magnitude of the surface MIC estimates in part B. Error bars correspond to the 90% posterior credible interval.
A.3  A. Marginal mean accuracies (pre-cleaning) by condition in Experiment 3. The change in the second digit is indicated by the x-axis while the change in the first digit is indicated by the color of the points, with blue corresponding to no change, green corresponding to a small change, and red corresponding to a large change. The four points corresponding to the double-change trials are further highlighted with black borders. B. Surface MIC estimates from Experiment 3, sorted by magnitude, with error bars that correspond to the 80% confidence intervals of the means. The points and error bars are colored with regard to sign: green for negative, red for positive, and blue for intervals containing zero. C. Marginal mean reaction times (post-cleaning) by condition in Experiment 3. The arrangement, coloration, and highlighting are as in part A. D. Posterior means of MIC estimates from the General model in the Normal framework for Experiment 3, sorted by the magnitude of the surface MIC estimates in part B. Error bars correspond to the 90% posterior credible interval.
A.4  **A.** Marginal mean accuracies (pre-cleaning) by condition in Experiment 4. The change in the second digit is indicated by the x-axis while the change in the first digit is indicated by the color of the points, with blue corresponding to no change, green corresponding to a small change, and red corresponding to a large change. The four points corresponding to the double-change trials are further highlighted with black borders. **B.** Surface MIC estimates from Experiment 4, sorted by magnitude, with error bars that correspond to the 80% confidence intervals of the means. The points and error bars are colored with regard to sign: green for negative, red for positive, and blue for intervals containing zero. **C.** Marginal mean reaction times (post-cleaning) by condition in Experiment 4. The arrangement, coloration, and highlighting are as in part A. **D.** Posterior means of MIC estimates from the General model in the Normal framework for Experiment 4, sorted by the magnitude of the surface MIC estimates in part B. Error bars correspond to the 90% posterior credible interval.
List of Tables

1.1 Crosstable of the nine possible combinations of two signals that occur in a Systems Factorial Technology experiment, with "Ø" corresponding to no signal, "L" corresponding to a weak signal, and "H" corresponding to a strong signal. ................................. 6

1.2 The same crosstable from Table 1.1 color-coded to describe the OR response structure, with yes responses colored in green and no responses colored in red. ........................................... 7

1.3 The same crosstable from Table 1.1 color-coded to describe the AND response structure, with yes responses colored in green and no responses colored in red. ........................................... 7

1.4 The same crosstable from Table 1.1 color-coded to describe the terms used in calculating the MIC. The MIC is calculated by subtracting the sum of the purple terms from the sum of the blue terms. ......................... 8
4.1 Eight Simulated data sets, with the distributions used to generate the interaction parameters, and the Serial-Other Log (base 10) Bayes Factors, rounded to two decimal places, comparing the Normal-framework Serial model to the other three models: General, Parallel, and Coactive. Positive values constitute evidence in favor of the Serial model and negative values constitute evidence in favor of the opposing “other” model. Log Bayes Factors comparing any other pair of models (within a given framework) can be derived by taking the difference between two log Bayes Factors, with positive values of this difference suggesting evidence in favor of the “other” model corresponding to the second log Bayes Factor and negative values in favor of the “other” model corresponding to the first log Bayes Factor.

4.2 Number of participants classified as using Serial, Parallel, or Coactive processing according to the participant-wise ANOVA analysis, using both a raw .05 error rate and an experiment-wise .05 error rate determined by Bonferroni correction.

4.3 Serial-Other Log (base 10) Bayes Factors, rounded to two decimal places, comparing the Serial model to the other three models: General, Parallel, and Coactive. Positive values constitute evidence in favor of the Serial model and negative values constitute evidence in favor of the opposing model. Log Bayes Factors comparing any other pair of models (within a given framework) can be derived by taking the difference between two log Bayes Factors, with positive values of this difference suggesting evidence in favor of the model corresponding to the second log Bayes Factor and negative values in favor of model corresponding to the first log Bayes Factor.
4.4 Number of participants classified as using Serial, Parallel, or Coactive processing according to the participant-wise ANOVA analysis, using both a raw .05 error rate and an experiment-wise .05 error rate determined by Bonferroni correction. .................................................. 77

4.5 Serial-Other Log (base 10) Bayes Factors, rounded to two decimal places, comparing the Serial model to the other three models: General, Parallel, and Coactive. Positive values constitute evidence in favor of the Serial model and negative values constitute evidence in favor of the opposing model. Log Bayes Factors comparing any other pair of models (within a given framework) can be derived by taking the difference between two log Bayes Factors, with positive values of this difference suggesting evidence in favor of the model corresponding to the second log Bayes Factor and negative values in favor of model corresponding to the first log Bayes Factor. ................................................................. 77

4.6 Number of participants classified as using Serial, Parallel, or Coactive processing according to the participant-wise ANOVA analysis, using both a raw .05 error rate and an experiment-wise .05 error rate determined by Bonferroni correction. .................................................. 79
4.7 Serial-Other Log (base 10) Bayes Factors, rounded to two decimal places, comparing the Serial model to the other three models: General, Parallel, and Coactive. Positive values constitute evidence in favor of the Serial model and negative values constitute evidence in favor of the opposing model. Log Bayes Factors comparing any other pair of models (within a given framework) can be derived by taking the difference between two log Bayes Factors, with positive values of this difference suggesting evidence in favor of the model corresponding to the second log Bayes Factor and negative values in favor of model corresponding to the first log Bayes Factor.

4.8 Number of participants classified as using Serial, Parallel, or Coactive processing according to the participant-wise ANOVA analysis, using both a raw .05 error rate and an experiment-wise .05 error rate determined by Bonferroni correction.

4.9 Serial-Other Log (base 10) Bayes Factors, rounded to two decimal places, comparing the Serial model to the other three models: General, Parallel, and Coactive. Positive values constitute evidence in favor of the Serial model and negative values constitute evidence in favor of the opposing model. Log Bayes Factors comparing any other pair of models (within a given framework) can be derived by taking the difference between two log Bayes Factors, with positive values of this difference suggesting evidence in favor of the model corresponding to the second log Bayes Factor and negative values in favor of model corresponding to the first log Bayes Factor.
Although there is substantial evidence supporting the existence of chunking in Working Memory, much of it is dependent on a small range of experimental methods, many of which rely on accuracy measures. Consequently, I feel there is a need to develop a new method for studying chunking. In this paper, I propose one such method, adapted from a perceptual research method known as Systems Factorial Technology, to determine whether a given pair of features is chunked or maintained as separate objects. In addition, I test this method on two pairs of stimuli through four experiments, testing whether the features are processed together or separately in both perception and memory. The current research finds no evidence for chunking in either feature pair considered.
Chapter 1

Introduction

Working memory is the subset of information of which a person is consciously aware at a given point in time. The most critical aspect of this subset is that it can only hold a limited amount of information. One possible way to circumvent this limit is to combine several pieces of information into a single piece of information, a process referred to as chunking, with the largest possible result of chunking that a participant can still understand being called a chunk. This process, first described by Miller (1956), allows a person to hold more information than normal both by maintaining a larger-than-normal amount of information in a given portion of working memory and by freeing up other resources that could be used to maintain other pieces of information.

A commonly-used example of chunking is the asking participants to remember a long string of letters, much like in the original Sperling (1960) experiments, and presenting them with the string “FBICIAIRS”, which consists of the three substrings “FBI”, “CIA”, and “IRS”. Where, in most other cases, participants would have difficulty in remembering the full string, participants can use familiar substrings like these instead of remembering the individual letters.

There is substantial evidence supporting the existence of chunking, such as differences in overall accuracies between strings or between other multi-featured stimuli (Luck & Vo-
correlation in errors in certain chunked items (Vul & Rich, 2010), reductions in reaction times in recalling other portions of chunks (Koch & Hoffmann, 2000), and gestalt-like relations in the order of recollection of information (Chase & Simon, 1973; Gobet & Simon, 1998). Additionally, although not necessarily direct evidence for chunking, everyday experiences of remembering one detail after remembering other details (that is, remembering contingent on prior memories) also supports the notion that information is (or, at least, can be chunked).

This body of evidence strongly supports the notion that people can chunk items and describes several methods for determining when people do or do not chunk items. However, evidence for specific cases of chunking or for types of information that are consistently chunked by people in general (or by certain subpopulations) is still relatively sparse or even non-existent. As such, providing additional evidence and analytic methods would only serve to improve upon past and current understanding of chunking in humans. Consequently, I suggest supplementing the current evidence with evidence from a new method to allow for even more robust conclusions. Specifically, I suggest trying to determine if chunking exhibits a specific reaction-time signature and, if so, how to detect when that signature occurs.

An example of such a signature comes from a research method in the perception literature known as Systems Factorial Technology. In this paper, I describe Systems Factorial Technology and how I intend to use it to find an alternate signature of chunking in visual working memory. I will describe Systems Factorial Technology in more detail, including the processing architectures, the order in which information is analyzed, in the next section and will describe how I intend to apply it to working memory research in the two subsequent sections. I will then describe the application of Bayesian Hierarchical modelling to this method and how it allows for both more efficient data use and more general conclusions, along with the two models actually used. Finally, I will describe, in detail, four
experiments, the results of the analysis intended to identify reaction-time-based signature of chunking if such a signal exists, and the interpretations and conclusions I have come up with over the course of this work along with the weaknesses of this approach and several considerations for future applications of this work.

1.1 Processing Architectures

Consider a simple detection task in which participants are presented with two light sources, one in each eye. In this experiment, participants are prompted to give one of two responses: one, called a *yes* response, if at least one light is on, and another, called a *no* response, if no light is present. Note that participants are told to give yes responses both when only one light is on and when both lights are on. This double-mapping leaves room for an interesting question: how does the reaction time on trials with two lights, referred to here as the *double-signal reaction time*, compare the reaction time on trials with only one light, referred to here as the *single-signal reaction time*? It turns out that this depends on what processing architecture, the type and order of processing steps, is used to produce the response.

For the current example, there are three processing architectures that we are considering: *serial processing*, where signals are processed one at a time, *parallel processing*, where signals are processed simultaneously, and *coactive processing*, where signals are combined into a single signal before being processed. A visual depiction of these processing architectures is shown in Figure 1.1.

Additionally, both serial and parallel processing can be further divided into *self-terminating* and *exhaustive* processing. In self-terminating processes, the relevant decision can be made before both pieces of information are processed, whereas both pieces must be processed before any decision can be made in exhaustive processing. The above depictions of serial and parallel processing both assume exhaustive processing and, although they are not inherently impossible in regards to this method, they are being left out for reasons that will soon
Each of these processing architectures makes a specific prediction about the relation between single-signal and double-signal reaction times. Both parallel and coactive processing predict that the double-signal reaction times will tend to be significantly faster than the single-signal reaction times, whereas serial processing does not predict this speeding. Although both coactive and parallel processing produce speeded times for the double-signal, there is an intuitive way of using the RT data to distinguish between them: there is a limit to how much faster double-signal times are than single-signal times when there is no shared information, such as in parallel processing. This limit is known as the race-model equality, and the cumulative distribution function (CDF) of the double-signal trials is given by:

\[ F(t) = 1 - \prod_{k=1}^{n} [1 - F_k(t)] \]  

(1.1)
Figure 1.2: Example CDFs for reaction times for the two different single-signal trials (black and red) and the CDF of the minimum of the two single-signal trial reaction times, which would correspond to the CDF of the double-signal trials under parallel processing.

This equality features prominently in the work of Miller (1982) and Burbeck and Luce (1982).

If the CDF for the double-signal trials are larger in value (that is, the distributions are more speeded), then the parallel race cannot have occurred, and some coactivation is implied (Figure 1.2).
Table 1.1: Crosstable of the nine possible combinations of two signals that occur in a Systems Factorial Technology experiment, with "Ø" corresponding to no signal, "L" corresponding to a weak signal, and "H" corresponding to a strong signal.

<table>
<thead>
<tr>
<th></th>
<th>HØ</th>
<th>HL</th>
<th>HH</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>LØ</td>
<td>LL</td>
<td>LH</td>
</tr>
<tr>
<td>Ø</td>
<td>ØØ</td>
<td>ØL</td>
<td>ØH</td>
</tr>
<tr>
<td>Ø</td>
<td>L</td>
<td>H</td>
<td></td>
</tr>
</tbody>
</table>

1.2 Systems Factorial Technology

These differences in processing architectures can be further highlighted with the addition of two main methodological elements, which are the main points implemented by Systems Factorial Technology (SFT) in the original Townsend & Nozawa (1995) paper and in subsequent SFT research.

**Element 1:** Multiple Signal Strengths. The Systems Factorial Technology paradigm relies on the combination of factorial manipulation of signals and differences in reaction times described earlier and multiple signal strengths. This combination produces a task with nine possible types of trials, which are arranged in Table 1.1, and allow for a more detailed analysis of processing architectures through the joint analysis of multiple CDFs. The details of the CDF-based analysis is left for interested readers (Fific, Nosofsky, & Townsend, 2008), suffice to say that each of the five possible architectures considered by SFT has a unique pattern within this analysis.

**Element 2:** Conjunctive Response Structures. There are two types of response structures that participants can be asked to provide in this method: the OR response structure described earlier, where participants are asked to give a yes response whenever any signal is present and a no response otherwise, and the AND response structure, where participants are asked to give a yes response whenever both signals are present and a no response otherwise. The OR and AND response structures are illustrated in Table 1.2 and Table 1.3, respectively, with yes responses in green and no responses in red. The original experiments
Table 1.2: The same crosstable from Table 1.1 color-coded to describe the OR response structure, with yes responses colored in green and no responses colored in red.

Table 1.3: The same crosstable from Table 1.1 color-coded to describe the AND response structure, with yes responses colored in green and no responses colored in red.

were already based on the OR response structure, but the AND response structure allows for the study of additional response types.

1.3 Simplification of the Method

One of the drawbacks of the CDF method in Eq 1.1 and the CDF methods described above in typical SFT analysis is that researchers must accurately estimate CDFs, which can requires large samples from steady-state conditions without fatigue or learning. This is problematic, as it requires participants, likely trained, that are willing to repeat a simple task for several hours, which most people are not willing to do. As such, it is desirable to find a method that does not require as much data. Fortunately, such an approach does exist within SFT. Indeed, if participants are required to make AND type responses, then the total number of possible processing architectures considered decreases because AND type responses necessarily require exhaustive processing. This restriction to exhaustive processing allows for the use of other, simpler, statistics as they become diagnostic.

In this case, the primary statistic of interest is called the Mean Interaction Contrast
Table 1.4: The same crosstable from Table 1.1 color-coded to describe the terms used in calculating the MIC. The MIC is calculated by subtracting the sum of the purple terms from the sum of the blue terms.

(MIC), which is calculated from the reaction times in the YES response conditions from the AND response structure. The formula for the Mean Interaction Contrast is:

$$MIC = (RT_{HH} + RT_{LL}) - (RT_{LH} + RT_{HL})$$

where $RT_{HH}$ is the mean reaction time observed in the double-strong signal (or HH) trials, with the other three terms defined similarly. As said above, the MIC is fully indicative of the processing architecture used in AND responses, with values less than 0 indicating serial processing, values equal to 0 indicating parallel processing, and values greater than 0 indicating coactive processing. However, these indications are contingent on one other important condition: a specific ordering of reaction times referred to as selective influence. In selective influence, the reaction times from the double-strong signal trials must be faster than the reaction times observed in either of the mixed-strength signal (HL and LH) trials which, in turn, must both be faster than reaction times observed in the double-weak signal (LL) trials.

For the purposes of this research, selective influence is being assumed, as testing for selective influence requires CDF estimation which, as stated earlier, requires more data than will be available in this project. This is a weakness of the current research and will be expanded on along with other weaknesses in the last chapter.
1.4 Prior Extensions of SFT to WM

The proposed extension to working memory research is not necessarily new. Prior research by Yang (2011), Yang, Chang, & Wu (2013), and Donkin et al. (2013) have all attempted to apply Systems Factorial Technology to working memory research. However, this application to working memory has largely been focused on estimating working memory capacity, or channel capacity as referred to in the above corpus. This paper will concern with a different approach, both in the topic of interest, how to detect chunking, and in the specific experimental and analytic methods being used.

1.5 Chunking as Coactive Processing

Up to now, this proposal has focused on the Systems Factorial Technology and its ability to distinguish coactive processing, the compression of two pieces of information into one piece, from other forms of processing that preserve the number of pieces of information. Because chunking can, itself, be defined as the combination of $n$ pieces of information into $m$ pieces of information, with $n > m$, it seems straightforward to assume that evidence for coactive processing would constitute evidence for chunking. The main difficulty in asserting this assumption is constructing the proper paradigm in which to test it. In the next section, I describe my plan for adapting Systems Factorial Technology to Working Memory research, which will be followed by a detailed description of the exact experiments that will follow this implementation.

1.6 Change-Detection Experiments

Because the Systems Factorial Technology paradigm is generally used with detection tasks, the application to change-detection is straightforward. In fact, because of the research
mentioned above, this application to change-detection tasks is not even particularly new. However, for the sake of clarity, I will relate this application to the original light-detection example described earlier.

The original intuition on light detection with sources in each eye can be similarly applied to a common change-detection task (Luck & Vogel, 1997; Wheeler & Treisman, 2002) that uses pairs of squares, one over top of the other, referred to here as nested squares. An example of this stimulus can be seen on the left side of Figure 1.3A. Just as participants might see no lights, one light, or two lights in the light-detection task, participants might be tested with a nested square that doesn’t change at all, a nested square with one color different from study, or a nested square with both colors different from study (see Figure 1.3A.). Similarly, the dim and bright lights used in the original experiment correspond to small and large changes in the color of single squares (see Figure 1.3B.). Finally, change-detection tasks typically use a binary response which are also used in Systems Factorial Technology. As a result, the application of Systems Factorial Technology to change-detection tasks seems fairly straight-forward.

The primary difficulty with this application is finding the specific experimental parameters needed to allow the interpretation to inform memory research, as Systems Factorial Technology is normally used in perceptual tasks and any similarity it has with working memory research does not inherently guarantee that the comparison will hold in all cases. In particular, the presence of coactive processing in change-detection tasks doesn’t immediately imply that chunking occurred, because the features might naturally be coactively processed outside of memory. However, if stimuli that are normally processed serially or in parallel in perceptual tasks were later processed coactively in an otherwise identical memory task, a signature that I will refer to here as the forced-funneling signature, then it would be appropriate to claim that chunking has occurred.

Forced-funneling would then require both a test of processing in memory and a second
test of processing in perception. Thankfully, there is a straightforward way to test perception that requires only a slight modification of the change-detection method: displaying the two stimuli concurrently instead of staggering them. This new task, referred to here as a direct comparison task, follows most of the same logic as the original change-detection task but doesn’t require a participant to rely on memory to make a correct decision, as both stimuli are fully available to them. And, because it is a minimal modification, the same analytic approaches used for the original change-detection task can be applied here.

All of these methods can be applied to data from individual participants to determine how a single individual processes information. Indeed, most prior instances of this method have done precisely this, relying on contrast analysis in ANOVA to characterize each participant’s processing architecture. However, the current goal for this project is to go beyond characterizing individuals and attempt to infer the general processing structure that is used by a larger population. In order to do this, additional work must be done to connect the data provided by this experiment to an inferential system. Specifically, statistical models of reaction time incorporating an parameter structure including some form of the interaction contrast that may or may not be restricted to a given range and a model comparison metric must be developed to allow for such inference. In this paper, we accomplish this by applying a 2-by-2 ANOVA-like structure to reaction times and construct models of the parameters that allow for various restrictions on the interaction term. Furthermore, by assuming that the parameters in this structure reflect subject-specific, random effects instead of fixed effects, we can model individual differences and generalize the results to a larger population instead of documenting a few specific participants.

To accomplish this, a Random-Effects Bayesian Hierarchical Model will be used to infer if people generally use the same architecture to process this information or if the variation in architecture is consistent enough to suggest the need to consider individual differences. The implementation details are described in the next chapter.
Figure 1.3: A. Examples of the standard (study) and sample (test) nested-squares stimuli that would be used in the proposed change-detection tasks along with their associated responses. B. The crosstable used to describe Systems Factorial Technology adapted to show examples of the sample stimuli that would be used in the change-detection variant of this task. The dashed lines separate the stimuli associated with a YES response above and to the right of the dashed lines.
Chapter 2

Details of Analysis

In this chapter I will cover the details that led to the first analytic framework employed in this project and how they relate to the goal of inferring processing architecture from a sample of participants.

2.1 Log-Normal Framework

The first framework that I considered to make the theory-data connection is the Log-Normal framework. In this framework, reaction time is assumed to follow a Log-Normal distribution. Specifically, with shift parameters $\psi_i$, scale parameters $\mu_{ijk}$, and shape parameter $\sigma^2$

$$y_{ijkl} \sim \text{Log-Normal} \left( \mu_{ijk}, \sigma^2, \psi_i \right)$$

(2.1)

for participant $i = 1, \ldots I$, for change levels $j = 1, 2$ and $k = 1, 2$ for the first and second features, respectively, with 1 corresponding to small changes (L) and 2 corresponding to large changes (H), and for repetition $l = 1, \ldots, L_{ijk}$ of the $jk^{th}$ condition for participant $i$. These subscripts will be listed as such throughout the paper. Additionally, it is important to
note that the total number of iterations per participant-condition combination is not necessarily equal due to data cleaning procedures and other potential unforeseen issues with data collection.

The above formulation is equivalent to the following

\[
\log (y_{ijkl} - \psi) \sim \text{Normal} (\mu_{ijk}, \sigma^2)
\]

framework was originally considered because the Log-Normal distribution has been used to model reaction time data by other researchers (Rouder et al., 2003; Rouder et al., 2008) and is fairly well-characterized.

However, this distribution does not allow for easy establishment of sign-constraint on the MIC parameters and, thus, it is not appropriate for our purposes. Specifically, the within-condition means under the Log-Normal assumption are equal to \(\exp \mu_{ijk} + \frac{\sigma^2}{2}\). Given this, the formula for the MICs is a sum of exponential terms

\[
MIC_i = \exp \left( \mu_{i11} + \frac{\sigma^2}{2} \right) + \exp \left( \mu_{i22} + \frac{\sigma^2}{2} \right) - \exp \left( \mu_{i21} + \frac{\sigma^2}{2} \right) - \exp \left( \mu_{i12} + \frac{\sigma^2}{2} \right)
\]

Unfortunately, there is no simple way to place sign constraint on sums of exponential terms. Because of this inability to place constraint, this framework was only considered for distribution estimation and initial program development and not for model comparison. As such, no Bayes Factors were calculated with this framework, although future development with this framework is being considered. To continue with this line of research, further development was performed with a second, somewhat similar framework: the Normal Framework.
2.2 Normal Framework

The second framework I attempted to use for this analysis was the 'Normal' framework, which specifies the following:

\[ y_{ijkl} \sim \text{Normal}(\mu_{ijk}, \sigma^2) \]  

(2.3)

for participant \( i = 1, \ldots, I \), change levels \( j = 1, 2 \) and \( k = 1, 2 \) for the first and second features, respectively, and repetition \( l = 1, \ldots, L_{ijk} \) of the \( jk^{th} \) condition for participant \( i \) and for mean parameters \( \mu_{ijk} \) and scale parameter \( \sigma^2 \).

This framework is being considered for two main reasons: mathematical convenience and similarity to the analyses used in previous SFT applications. First, the Normal assumption affords a lot of simplicity to the analysis by allowing reliance on a well-studied ANOVA-like approach (conjugate priors and other mathematical conveniences). Second, applying an ANOVA-like approach, such as the one described subsequently, resembles the ANOVAs used by several other researchers in their applications of SFT (Fific, Nosofsky, & Townsend, 2008; Yang, 2011; Yang, Chang, & Wu, 2013; Donkin et al., 2013).

The mean parameters, \( \mu_{ijk} \), reflect the \( jk^{th} \) condition means for each participant. Consequently, the MIC is calculated as follows:

\[ MIC_i = \mu_{i11} + \mu_{i22} - \mu_{i21} - \mu_{i12} \]  

(2.4)

If we use an ANOVA-like model on \( \mu_{ijk} \), where

\[ \mu_{ijk} = \eta_i + \alpha_i(-1)^{i+j+1} + \beta_i(-1)^{k+1} + \gamma_i(-1)^{i+k} \]  

(2.5)
then the MIC above simplifies to

\[
MIC_i = \left( \eta_i + \alpha_i (-1)^{1+1} + \beta_i (-1)^{1+1} + \gamma_i (-1)^{1+1} \right)
+ \left( \eta_i + \alpha_i (-1)^{2+1} + \beta_i (-1)^{2+1} + \gamma_i (-1)^{2+2} \right)
- \left( \eta_i + \alpha_i (-1)^{2+1} + \beta_i (-1)^{1+1} + \gamma_i (-1)^{2+1} \right)
- \left( \eta_i + \alpha_i (-1)^{1+1} + \beta_i (-1)^{2+1} + \gamma_i (-1)^{1+2} \right)

= (\eta_i + \alpha_i + \beta_i + \gamma_i) + (\eta_i - \alpha_i - \beta_i - \gamma)
- (\eta_i - \alpha_i + \beta_i - \gamma_i) - (\eta_i + \alpha_i - \beta_i - \gamma_i)

= \eta_i + \alpha_i + \beta_i + \gamma_i + \eta_i - \alpha_i - \beta_i + \gamma_i
- \eta_i + \alpha_i - \beta_i + \gamma_i - \eta_i - \alpha_i + \beta_i + \gamma_i

= \eta_i + \eta_i - \eta_i + \eta_i + \alpha_i - \alpha_i + \alpha_i - \alpha_i
+ \beta_i - \beta_i - \beta_i + \gamma_i + \gamma_i + \gamma_i

= 4\gamma_i
\]

Consequently, use of the ANOVA-like model, with normally-distributed grand mean, main effect, and interaction terms, allows for the calculation of a general model in this framework that produces an easy-to-compute Bayes Factor comparing the serial and general models with the Savage-Dickey ratio. By using truncated-priors, sign constraint can be applied to the interaction term, which allows for similar calculation of Bayes Factors comparing the serial model to both the parallel and coactive models.

Although this framework suffices for data analysis and inference, it does suffer from a couple of problems. Specifically, it does not adequately describe reaction-time data because it does not account for two well-known characteristics of such data: positive skew and, more importantly, non-independence of the mean and variance. These issues will be covered in more detail later, but they are mentioned now to describe what will be an integral
motivation for the development of later frameworks.

Now that we’ve established a basic framework that allows for the statistical description of data, we need a way to connect these descriptions to the four models described in Chapter 1 in order to determine which one is most representative of the data. This connection will allow us to use experimental data to identify which architecture most accurately reflects population behavior.

2.3 Bayes Factors

For my research, I will be using Bayes Factors to compare the four models of interest and to determine the best performing model with respect to the data. Bayes Factors are ratios of the probabilities of observing data given two different models. These ratios describe the amount of evidence in favor the first model, relative to the second model, that is provided by the observed data. To do so, Bayes Factors rely on a fundamental rule in probability, Bayes’ Rule, which states that, for events $\mathcal{A}$ and $\mathcal{B}$:

$$
\Pr(\mathcal{B}|\mathcal{A}) = \frac{\Pr(\mathcal{A}|\mathcal{B}) \Pr(\mathcal{B})}{\Pr(\mathcal{A})} \quad (2.6)
$$

Using this rule and a set of models $\mathcal{M}$, we can similarly define the probabilities of one model $\mathcal{M}_i \in \mathcal{M}$ given some data $\mathcal{D}$, or the posterior probability of the model as:

$$
\Pr(\mathcal{M}_i|\mathcal{D}) = \frac{\Pr(\mathcal{D}|\mathcal{M}_i) \Pr(\mathcal{M}_i)}{\Pr(\mathcal{D})} \quad (2.7)
$$

where $\Pr(\mathcal{D}|\mathcal{M}_i)$ is the probability of the data assuming the model is true, $\Pr(\mathcal{M}_i)$ is the prior probability of model $\mathcal{M}_i$ being true, and $\Pr(\mathcal{D})$ is defined as follows:

$$
\Pr(\mathcal{D}) = \sum_{\mathcal{M}_i \in \mathcal{M}} \Pr(\mathcal{D}|\mathcal{M}_i) \Pr(\mathcal{M}_i) \quad (2.8)
$$
This posterior probability characterizes the evidence in favor of the model relative to the whole set of models. However, an absolute measure like this is generally hard to calculate due to the large number of models being considered. A preferable measure would allow for estimation from a smaller subset of these models but maintain the relative evidence between any pair of models being considered. One such measure becomes apparent when you compare the posterior probabilities of two different models, listed here as $M_1$ and $M_2$. Specifically, if we take the ratio of the posterior probabilities to determine the posterior odds, or the odds of one model relative to the other, defined as

$$\frac{\Pr(M_1|D)}{\Pr(M_2|D)} = \frac{\Pr(D|M_1)\Pr(M_1)}{\Pr(D|M_2)\Pr(M_2)}$$

(2.9)

then we can calculate the evidence in favor of arbitrary pairs of models without needing to consider the other models, because we can assume that the $\Pr(D)$ terms in the numerator and denominator are equivalent either by assuming only the two models being compared in the posterior odds or by considering the same overall set of models. In either case, because these terms are equivalent, they cancel out and calculating them is no longer needed, leaving the posterior odds equal to

$$\frac{\Pr(M_1|D)}{\Pr(M_2|D)} = \frac{\Pr(D|M_1)\Pr(M_1)}{\Pr(D|M_2)\Pr(M_2)}$$

(2.10)

These odds can be broken down further into two main components:

$$\frac{\Pr(M_1|D)}{\Pr(M_2|D)} = \frac{\Pr(D|M_1)}{\Pr(D|M_2)} \times \frac{\Pr(M_1)}{\Pr(M_2)}$$

(2.11)

$$= B_{12} \times OR_{\text{prior}}$$

(2.12)

These components are the Bayes Factor ($B_{12}$) and the prior odds ($OR_{\text{prior}}$), respectively. The Bayes Factor is calculated from the data, whereas the prior odds are always specified by
the researcher interpreting the results and reflect the researcher’s prior beliefs in the relative evidence for each model. As such, the resulting posterior odds will always depend on the individual interpreting the information. However, because the prior odds are a subjective component supplied by the reader, the full posterior odds generally cannot be specified, even with full data. This does not mean that they are not important, rather that the exact value depends on the person interpreting the results. Indeed, this subjectivity itself suggests that the full posterior odds are not of primary interest, as specifying the posterior odds for even a small subset of potential readers would likely be an exercise in wasted effort. Instead, we will focus primarily on the Bayes Factor, as it is fully determined by the models being specified and the data observed.

The Bayes Factor possesses a few beneficial properties, in addition to the above generality, that makes it well suited to this research. First, it has an inherent meta-analytic property that allows for easy updating in the face of new evidence. Specifically, the posterior odds after observing one data set are, in turn, used as the new prior odds for subsequent analyses within the same model space, providing one more reason to focus on the Bayes Factor over the prior or posterior odds themselves. This property means that stopping rules generally have no direct impact on the conclusion being made, provided that the data are not re-analyzed or otherwise mistreated (Unclear).

Second, because the Bayes Factors are relative to the two models being considered, they do not change if more or less models are being considered overall. That is, the relative evidence for one model over the other does not change when new models are considered.

Third, the Bayes Factor is transitive. If you wish to compare three different models, \( M_1 \), \( M_2 \), and \( M_3 \), but only have the Bayes Factors comparing \( M_1 \) to \( M_2 \) (\( B_{12} \)) and comparing \( M_1 \) to \( M_3 \) (\( B_{13} \)), then you can directly compare models \( M_2 \) and \( M_3 \) with the following
\[ B_{23} = \frac{\Pr(\mathcal{D}|\mathcal{M}_2)}{\Pr(\mathcal{D}|\mathcal{M}_3)} \]  \tag{2.13}

\[ = \frac{\Pr(\mathcal{D}|\mathcal{M}_1)}{\Pr(\mathcal{D}|\mathcal{M}_3)} \times \frac{\Pr(\mathcal{D}|\mathcal{M}_2)}{\Pr(\mathcal{D}|\mathcal{M}_1)} \]  \tag{2.14}

\[ = \frac{\Pr(\mathcal{D}|\mathcal{M}_1)}{\Pr(\mathcal{D}|\mathcal{M}_3)} \times \frac{\Pr(\mathcal{D}|\mathcal{M}_2)}{\Pr(\mathcal{D}|\mathcal{M}_1)} \]  \tag{2.15}

\[ = \frac{B_{13}}{B_{12}} \]  \tag{2.16}

This last point is especially important as it allows the calculation of the full set of \( m \) Bayes Factors between each possible pair of \( m \) models from a subset of the Bayes Factors (provided that at least \( m - 1 \) Bayes Factors are calculated). By determining the remaining Bayes Factors from the \( m - 1 \) calculated Bayes Factors, less time is needed for determining the full set of Bayes Factors, which would otherwise take a fair amount of time to determine in full.

Derivation up to proportionality is a quick way to estimate \( \Pr(\mathcal{M}_1|\mathcal{D}) \) by estimating \( \Pr(\mathcal{D}|\mathcal{M}_1)\Pr(\mathcal{M}_1) \) and ignoring the denominators detailed above in Equations ?? and ???. This can be done because, as stated earlier, the denominators are constant with respect to a given data set. Additionally, all of the frameworks we consider are of a simple enough functional form that exact derivation is generally never needed, especially due to the use of the Savage-Dickey ratio.

Calculation of Bayes Factors is often an involved process, typically requiring a fair amount of numerical sampling or exact algebraic derivation. This is not always the case and, in some simple, well-studied models, the Bayes Factor can be derived without needing to rely on numeric sampling methods. Systems Factorial Technology, which my research is based on, is not a simple, well-studied model and, consequently, will at least require numerical sampling. Fortunately, there are a couple of simple methods that we can use that
do not require anything beyond numerical sampling: derivation up to proportionality and a Bayes Factor estimate known as the Savage-Dickey ratio.

Additionally, if some Bayes Factors can be calculated more easily than others, then priority can be given to the easier-to-determine Bayes Factors, saving additional time and computational resources. This reliance on a simpler derivation is part of my application and will be described in the next section.

2.4 Savage-Dickey

The Savage-Dickey ratio, originally developed by Savage (1972) and Dickey and Lientz (1970), is a simple way to calculate the Bayes Factor comparing two models where one model is a constrained version of the other model, assuming other conditions are met. For models $\mathcal{M}_0$ and $\mathcal{M}_1$ with common parameters $\theta$ and $\phi$ where, in model $\mathcal{M}_0$, $\theta = \theta_0$, the Savage-Dickey ratio is defined as

\[
SDR = \frac{p(\theta = \theta_0 | \mathbf{y})}{p(\theta = \theta_0)} \tag{2.17}
\]

where the numerator is the posterior density of $\theta$ in model $\mathcal{M}_1$ evaluated at $\theta = \theta_0$ and the denominator is the prior density of $\theta$ in model $\mathcal{M}_1$ evaluated at the same restriction.

Figure 2.1 illustrates this ratio for a single-dimension $\theta$ when compared to models where $\theta_0 = .3$ (Figure 2.1A) and where $\theta_0 = 2$ (Figure 2.1B).

It is important to note that the Savage-Dickey ratio is only equal to the Bayes Factor if the prior densities for each model are not dependent on the specific parameter of interest. That is

\[
p_0(\phi) = p_1(\phi | \theta = \theta_0) \tag{2.18}
\]
Figure 2.1: Prior (solid black) and Posterior (dashed green) densities of a parameter $\theta$ for model $\mathcal{M}_1$. The Savage-Dickey ratios for the specified restrictions are also provided. 

A. For $\theta_0 = .3$, the likelihood of $\theta$ is greater under the prior than under the posterior, suggesting that the restriction is not well-supported. 

B. For $\theta_0 = 2$, the likelihood of $\theta$ is greater under the posterior than under the prior, providing evidence favoring the restriction.
This is generally the case with the models below and, as such, will not interfere with our calculation of the Savage-Dickey ratio. The actual estimation of these ratios will be done with Conditional Marginal Density Estimation, which is described subsequently, and which provides the final link between the data and the models.

2.5 Conditional Marginal Density Estimation

Conditional Marginal Density Estimation (CMDE) is an estimation method developed by Gelfand and Smith (1990) that takes advantage of information in and MCMC chain to generate estimates of the marginal density. Specifically, ”sampling” the density of the marginal posterior distribution of the parameter of interest, using the chain estimates of the parameters to define the posterior distribution as needed. However, this sampling requires a known normalizing constant and, as such, frequently involves a fair amount of additional computation. Fortunately, because we used a conjugate prior for the interaction parameter in the Normal framework, the full functional form and normalizing constant for the posterior distribution is always known, but this is not generally the case and xxxx.

In addition to directly connecting the data to the density estimates used to calculate the Bayes Factors, CMDE also outperforms other known density estimation methods, including normal approximation and log-spline estimation (Morey, Rouder, Pratte, & Speckman, 2011).

2.6 Full Conditional Posteriors in the Normal Framework

By combining all of the above techniques and assumptions, we can use Bayesian estimation with Systems Factorial Technology to determine whether a group of people are best described as using serial, parallel, or coactive processing in a given stimulus or if there is no single common strategy used by all participants, which will be referred to throughout as the
general model. In the case of the general model, a Bayes Factor is relatively easy to derive: calculate the Savage-Dickey ratio comparing some prior over all possible real-number values of the MIC (for the general model) to an MIC of zero (for the serial model). As for the parallel and coactive models, only a small adjustment to the prior is needed: constraining the parameter space to being less than or equal to zero (parallel) or greater than or equal to zero (coactive). This restriction, referred to as sign constraint throughout, can be done with truncation. In general terms, truncating a distribution $f$ from above at some value $t$ (thus, only permitting the variable to take values no more than $t$) can be done with the following formula

$$g_{\leq t}(x) = \frac{f(x)}{F(t)} \mathcal{I}(x \leq t)$$

(2.19)

where $g$ is the truncated distribution of interest, $F$ is the cumulative density function for distribution $f$, and $\mathcal{I}(x \leq t)$ is an indicator function that equals 1 when $x \leq t$ and 0 otherwise. Similarly, truncating from below at $t$ can be done with

$$g_{\geq t}(x) = \frac{f(x)}{1 - F(t)} \mathcal{I}(x \geq t)$$

(2.20)

By applying these truncations to the original, general-model prior, with $t = 0$ in both cases, priors corresponding to the parallel and coactive models can be generated, allowing for similar Savage-Dickey ratio calculations to be made comparing the parallel and coactive models to the serial model. All of these Bayes Factors can, in turn, be used to derive other Bayes Factors to compare the general model to the parallel and coactive models or to compare the parallel and coactive models to each other, as

$$\frac{B_{12}}{B_{13}} = \frac{\Pr(D|M_1)}{\Pr(D|M_2)} \frac{\Pr(D|M_1)}{\Pr(D|M_3)}$$

(2.21)
\[ \begin{align*}
\Pr(D|M_1) & = \frac{\Pr(D|M_1)}{\Pr(D|M_2)} \times \left( \frac{\Pr(D|M_1)}{\Pr(D|M_3)} \right)^{-1} \quad (2.22) \\
\Pr(D|M_1) & = \frac{\Pr(D|M_1)}{\Pr(D|M_2)} \times \frac{\Pr(D|M_3)}{\Pr(D|M_1)} \quad (2.23) \\
\Pr(D|M_3) & = \frac{\Pr(D|M_1)}{\Pr(D|M_2)} \quad (2.24) \\
\Pr(D|M_3) & = B_{32} \quad (2.25)
\end{align*} \]

With the above formulas in mind, the full derivation for the models within the normal framework is as follows:

For participant \( i = 1, \ldots, I \), first stimulus change level \( j = 1, 2 \), second stimulus change level \( k = 1, 2 \), and iteration \( l = 1, \ldots, L_{ijk} \), the common framework assumptions are as follows:

\( y_{ijkl} \sim \text{Normal}(\mu_{ijk}, \sigma^2) \)

\[ \mu_{ijk} = \eta_i + \alpha_i (-1)^j + \beta_i (-1)^k + \gamma_i (-1)^{j+k} \]

\( \eta_i \sim \text{Normal}(\nu_\eta, \delta_\eta) \)

\( \alpha_i \sim \text{Normal}(\nu_\alpha, \delta_\alpha) \)

\( \beta_i \sim \text{Normal}(\nu_\beta, \delta_\beta) \)

\( \sigma^2 \sim \text{Inverse-Gamma}(a_\sigma, b_\sigma) \)

\( \nu_\eta \sim \text{Normal}(\phi_\eta, \theta_\eta) \)

\( \nu_\alpha \sim \text{Normal}(\phi_\alpha, \theta_\alpha) \)

\( \nu_\beta \sim \text{Normal}(\phi_\beta, \theta_\beta) \)

\( \delta_\eta \sim \text{Inverse-Gamma}(a_\eta, b_\eta) \)

\( \delta_\alpha \sim \text{Inverse-Gamma}(a_\alpha, b_\alpha) \)

\( \delta_\beta \sim \text{Inverse-Gamma}(a_\beta, b_\beta) \)
The specific assumptions are as follows:

### 2.6.1 Serial Model

Because the MIC predicted by serial processing is zero, this model assumes that all interaction parameters, $\gamma_i$ are exactly zero. Because these parameters are all constant, there is no need for any further hierarchical specification in this model.

### 2.6.2 General Model

The general model places minimal constraint on the possible range of the interaction parameters. For the purposes of this current research, we will assume that the interaction parameters each follow a normal distribution with a hierarchical structure described as follows:

\[
\gamma_i \sim \text{Normal}(\nu_\gamma, \delta_\gamma)
\]

\[
\nu_\gamma \sim \text{Normal}(\phi_\gamma, \theta_\gamma)
\]

\[
\delta_\gamma \sim \text{Inverse-Gamma}(a_\gamma, b_\gamma)
\]

where $\phi_\gamma$, $\theta_\gamma$, $a_\gamma$, and $b_\gamma$ are hyperparameters that are specified before analysis. It should be noted that the assumptions in this model are very similar to those in the next two models. The only major change is in the distribution of the interaction parameter itself, with the hyperpriors on $\nu_\gamma$ and $\delta_\gamma$ remaining the same. As such, only the change to the distribution to $\gamma_i$ will be noted.
2.6.3 Parallel Model

Because parallel processing predicts negative MICs, this model assumes that all interaction parameters must be negative. This is accomplished by the following restriction:

\[ \gamma_i \sim \text{Truncated-Normal}\left(\nu, \delta, (-\infty, 0]\right) \]

with \((-\infty, 0]\) being the support set for these truncated distributions. This limitation to the support set constitutes the sign constraint needed for this model and only allows interaction parameters to equal zero in order to allow for Savage-Dickey estimation.

2.6.4 Coactive Model

Because coactive processing predicts positive MICs, this model assumes that all interaction parameters must be positive. This is accomplished by the following restriction:

\[ \gamma_i \sim \text{Truncated-Normal}\left(\nu, \delta, [0, \infty)\right) \]

with \([0, \infty)\) being the support set for these truncated distributions. This limitation to the support set constitutes the sign constraint needed for this model and only allows interaction parameters to equal zero in order to allow for Savage-Dickey estimation.

2.6.5 Joint Distribution and Posterior Derivation of Common Parameters

Because the common parameters are independent of the specific parameters and only depend on them when they are assumed to be known, the priors for the model-specific parameters are effectively normalizing constants that can be dropped during derivation up to proportionality. As such, those distributions are of no interest to the current joint analysis.
and will be omitted. Consequently, the joint distribution of all common parameters is as follows:

\[
\begin{align*}
    f(Y, \bar{\eta}, \bar{\alpha}, \bar{\beta}, \bar{\gamma}, \sigma^2, \nu_\eta, \nu_\alpha, \nu_\beta, \nu_\gamma, \delta_\eta, \delta_\alpha, \delta_\beta, \delta_\gamma) \\
    \propto \prod_{i,j,k} \prod_{l} (2\pi \sigma^2)^{-\frac{1}{2}} \exp \left( -\frac{(y_{ijkl} - \mu_{ijkl})^2}{2\sigma^2} \right) \times \prod_{i} (2\pi \delta_\eta)^{-\frac{1}{2}} \exp \left( -\frac{(\eta_i - \nu_\eta)^2}{2\delta_\eta} \right) \\
    \times \prod_{i} (2\pi \delta_\alpha)^{-\frac{1}{2}} \exp \left( -\frac{(\alpha_i - \nu_\alpha)^2}{2\delta_\alpha} \right) \times \prod_{i} (2\pi \delta_\beta)^{-\frac{1}{2}} \exp \left( -\frac{(\beta_i - \nu_\beta)^2}{2\delta_\beta} \right) \\
    \times \frac{b_{\sigma}^{a_\sigma}}{\Gamma(a_\sigma)} (\sigma^2)^{-a_\sigma - 1} \exp \left( -\frac{b_{\sigma}}{\sigma^2} \right) \times \frac{b_{\alpha}^{a_\alpha}}{\Gamma(a_\alpha)} (\alpha^2)^{-a_\alpha - 1} \exp \left( -\frac{b_{\alpha}}{\alpha^2} \right) \times \frac{b_{\beta}^{a_\beta}}{\Gamma(a_\beta)} (\beta^2)^{-a_\beta - 1} \exp \left( -\frac{b_{\beta}}{\beta^2} \right) \\
    \times \frac{b_{\eta}^{a_\eta}}{\Gamma(a_\eta)} (\eta^2)^{-a_\eta - 1} \exp \left( -\frac{b_{\eta}}{\eta^2} \right) \\
    \times (2\pi \sigma^2)^{-\frac{\Sigma_i \Sigma_j \Sigma_k \Sigma_l}{2}} \exp \left( -\frac{\Sigma_i (\eta_i - \nu_\eta)^2}{2\delta_\eta} \right) \times (2\pi \delta_\alpha)^{-\frac{\Sigma_i (\alpha_i - \nu_\alpha)^2}{2\delta_\alpha}} \\
    \times (2\pi \delta_\beta)^{-\frac{\Sigma_i (\beta_i - \nu_\beta)^2}{2\delta_\beta}} \times (2\pi \theta_\eta)^{-\frac{(\nu_\eta - \phi_\eta)^2}{2\theta_\eta}} \times (2\pi \theta_\alpha)^{-\frac{(\nu_\alpha - \phi_\alpha)^2}{2\theta_\alpha}} \\
    \times (2\pi \theta_\beta)^{-\frac{(\nu_\beta - \phi_\beta)^2}{2\theta_\beta}} \times \frac{b_{\alpha}^{a_\alpha}}{\Gamma(a_\alpha)} (\alpha^2)^{-a_\alpha - 1} \exp \left( -\frac{b_{\alpha}}{\alpha^2} \right) \\
    \times \frac{b_{\beta}^{a_\beta}}{\Gamma(a_\beta)} (\beta^2)^{-a_\beta - 1} \exp \left( -\frac{b_{\beta}}{\beta^2} \right) \\
\end{align*}
\]

This result is proportional because the distributions for \(\gamma, \nu_\gamma, \) and \(\delta_\gamma\) are not included in
the above formulation. This is because those distributions do not contain any factors of the above common parameters and, consequently, are of constant value with respect to them. Thus, using the above joint distribution to solve the posterior distribution of any common parameter is equivalent to using the full joint for any given model. With that in mind, the posterior distributions for the common parameters are given as follows:

\[ \eta_i: \]

\[
f(\eta_i|\cdot) \propto \exp \left( -\frac{\sum_j \sum_k \sum_l \left( y_{ijkl} - \eta_i - \alpha_i(-1)^j - \beta_i(-1)^k - \gamma_i(-1)^{j+k} \right)^2}{2\sigma^2} \right) \times \exp \left( -\frac{(\eta_i - \nu_{\eta})^2}{2\delta_{\eta}} \right) \times \exp \left( -2y_{ijkl} \eta_i \right) + 2\eta_i \beta_i(-1)^k + 2\alpha_i \gamma_i(-1)^{j+k} + 2\alpha_i \gamma_i(-1)^{j+k} \Rightarrow f(\eta_i|\cdot) \propto \exp \left( -\frac{\sum_j \sum_k \sum_l \eta_i^2 - 2y_{ijkl} \eta_i + 2\eta_i \alpha_i(-1)^j + 2\eta_i \gamma_i(-1)^{j+k} + 2\eta_i \gamma_i(-1)^{j+k}}{2\sigma^2} \right) \times \exp \left( -\frac{\eta_i^2 - 2\eta_i \nu_{\eta}}{2\delta_{\eta}} \right) \exp \left( -\frac{L_i \eta_i^2 - 2\eta_i \left( \sum_j \sum_k \sum_l y_{ijkl} - \sum_j \alpha_i(-1)^j L_{ij} - \sum_k \beta_i(-1)^k L_{ik} - \sum_j \gamma_i(-1)^{j+k} L_{ijl} \right)}{2\sigma^2} \right) \times \exp \left( -\frac{\eta_i^2 - 2\eta_i \nu_{\eta}}{2\delta_{\eta}} \right) \]
for \( L_{ik} = \sum_j L_{ijk}, \ \ L_{ij} = \sum_k L_{ijk}, \) and \( L_{ii} = \sum_j \sum_k L_{ijk} \)

\[
\begin{align*}
&= \exp \left( -\frac{1}{2} \left[ \eta_i^2 \left( \frac{L_{ii}}{\sigma^2} + \frac{1}{\delta \eta} \right) - 2\eta_i \left( \frac{L_{ii} \bar{y}_{i j k l} - L_{i\alpha} \alpha_i - L_{i\beta} \beta_i - L_{i\gamma} \gamma_i}{\sigma^2} + \frac{\nu \eta}{\delta \eta} \right) \right] \right) \\
\end{align*}
\]

for \( L_{i\alpha} = L_{i2} - L_{i1}, \ \ L_{i\beta} = L_{i2} - L_{i1}, \) and \( L_{i\gamma} = L_{i22} + L_{i11} - L_{i12} - L_{i21} \)

\[
\begin{align*}
&= \exp \left( -\frac{1}{2} \left[ \frac{L_{ii}}{\sigma^2} + \frac{1}{\delta \eta} \right] \left[ \eta_i^2 - 2\eta_i \left( \frac{L_{ii} \bar{y}_{i j k l} - L_{i\alpha} \alpha_i - L_{i\beta} \beta_i - L_{i\gamma} \gamma_i}{\sigma^2} + \frac{\nu \eta}{\delta \eta} \right) \right] \right) \\
&= \exp \left( \frac{L_{ii}}{\sigma^2} + \frac{1}{\delta \eta} \right)^{-1} \\
\Rightarrow \eta_i \sim \text{Normal} \left( \frac{L_{ii} \bar{y}_{i j k l} - L_{i\alpha} \alpha_i - L_{i\beta} \beta_i - L_{i\gamma} \gamma_i}{\sigma^2} + \frac{\nu \eta}{\delta \eta}, \frac{L_{ii}}{\sigma^2} + \frac{1}{\delta \eta}^{-1} \right) \\
\end{align*}
\]

Ideally, \( L_{i\alpha}, L_{i\beta}, \) and \( L_{i\gamma} \) would all equal 0 and all non-data-related terms would fall out of the posterior, but these factors are being included in the event that one or more of these factorial manipulations isn’t equally represented, whether this failure to include is a result of data cleaning, poor experiment coding, or some unforeseen accident.

\( \alpha_i: \) Although the terms kept in the exp function are different, specifically

\[
\begin{align*}
\alpha_i^2 (-1)^{2j - 2y_{ijk}} \alpha_i(-1)^j + 2\eta_i \alpha_i(-1)^j + 2\alpha_i \beta_i(-1)^{j+k} + 2\alpha_i \gamma_i(-1)^{2j+k} \\
&= \alpha_i^2 - 2y_{ijk} \alpha_i(-1)^j + 2\eta_i \alpha_i(-1)^j + 2\alpha_i \beta_i(-1)^{j+k} + 2\alpha_i \gamma_i(-1)^{k} \\
\Rightarrow \sum_j \sum_k \sum_l \alpha_i^2 - 2y_{ijk} \alpha_i(-1)^j + 2\eta_i \alpha_i(-1)^j + 2\alpha_i \beta_i(-1)^{j+k} + 2\alpha_i \gamma_i(-1)^{k} \\
\end{align*}
\]
\[ L_i.\alpha_i^2 - 2\alpha_i \left( \sum_j \sum_k \sum_l y_{ijkl}(-1)^j - L_i\alpha_i\eta_i - L_i\gamma_i\beta_i - L_i\beta_i \right) \]

the derivation for \( f(\alpha_i|\cdot) \) is similar to that for \( f(\eta_i|\cdot) \). As such, only the posterior will be listed here as a space-saving measure.

\[
\alpha_i| \sim \text{Normal} \left( \left( \sum_j \sum_k \sum_l \bar{y}_{ijkl}(-1)^j - L_i\alpha_i\eta_i - L_i\gamma_i\beta_i - L_i\beta_\gamma \right) \left( \frac{L_i.}{\sigma^2} + \frac{1}{\delta \eta} \right)^{-1} , \left( \frac{L_i.}{\sigma^2} + \frac{1}{\delta \eta} \right)^{-1} \right)
\]

\( \beta_i \):

Similarly to \( \alpha_i \),

\[
\beta_i^2(-1)^{2k} - 2y_{ijkl}\beta_i(-1)^k + 2\eta_i\beta_i(-1)^k + 2\alpha_i\beta_i(-1)^{j+k} + 2\beta_i\gamma_i(-1)^{j+2k} = \beta_i^2 - 2y_{ijkl}\beta_i(-1)^k + 2\eta_i\beta_i(-1)^k + 2\alpha_i\beta_i(-1)^{j+k} + 2\beta_i\gamma_i(-1)^j
\]

\[
\sum_j \sum_k \sum_l \beta_i^2 - 2y_{ijkl}\beta_i(-1)^k + 2\eta_i\beta_i(-1)^k + 2\alpha_i\beta_i(-1)^{j+k} + 2\beta_i\gamma_i(-1)^j
\]

\[
L_i.\beta_i^2 - 2\beta \left( \sum_j \sum_k \sum_l y_{ijkl}(-1)^k - L_i\beta_i - L_i\gamma_i\alpha_i - L_i\beta_i \right)
\]

comprises the terms kept inside the exp function, with the remaining derivation of \( f(\beta_i|\cdot) \) being similar to the derivations of \( f(\eta_i|\cdot) \) and \( f(\alpha_i|\cdot) \). Thus, the posterior for \( \beta_i|\cdot \) is,

\[
\beta_i| \sim \text{Normal} \left( \left( \sum_j \sum_k \sum_l \bar{y}_{ijkl}(-1)^k - L_i\beta_i\eta_i - L_i\gamma_i\alpha_i - L_i\beta_i \right) \left( \frac{L_i.}{\sigma^2} + \frac{1}{\delta \eta} \right)^{-1} , \left( \frac{L_i.}{\sigma^2} + \frac{1}{\delta \eta} \right)^{-1} \right)
\]
\( \sigma^2 \):

\[
f(\sigma^2 | \cdot) \propto \left( \frac{\sum_i \sum_j \sum_k L_{ijk}}{\sigma^2} \right)^{-1} \exp \left( -\frac{\sum_i \sum_j \sum_k \sum_l (y_{ijkl} - \mu_{ijk})^2}{2\sigma^2} \right) \times \frac{b^a_{\sigma}}{\Gamma(a_{\sigma})} (\sigma^2)^{-a_{\sigma} - 1} \exp \left( -\frac{b_{\sigma}}{\sigma^2} \right) \propto \left( \sigma^2 \right)^{-\frac{L_{\cdot \cdot}}{2} - a_{\sigma} - 1} \exp \left( -\frac{1}{\sigma^2} \left[ \frac{\sum_i \sum_j \sum_k \sum_l (y_{ijkl} - \mu_{ijk})^2}{2} + b_{\sigma} \right] \right)
\]

for \( L_{\cdot \cdot} = \sum_i \sum_j \sum_k L_{ijk} \).

\[
\Rightarrow \sigma^2 | \cdot \sim \text{Inverse-Gamma} \left( \frac{L_{\cdot \cdot}}{2} + a_{\sigma}, \frac{\sum_i \sum_j \sum_k \sum_l (y_{ijkl} - \mu_{ijk})^2}{2} + b_{\sigma} \right)
\]

For the given priors,

\[
\sigma^2 | \cdot \sim \text{Inverse-Gamma} \left( \frac{L_{\cdot \cdot}}{2} + 2, \frac{\sum_i \sum_j \sum_k \sum_l (y_{ijkl} - \mu_{ijk})^2}{2} + \frac{1}{4} \right)
\]

\( \nu_{\eta} \):

\[
f(\nu_{\eta} | \cdot) \propto \exp \left( -\frac{\sum_i (\eta_i - \nu_{\eta})^2}{2\delta_{\eta}} \right) \times \exp \left( -\frac{(\nu_{\eta} - \phi_{\eta})^2}{2\theta_{\eta}} \right)
\]

32
\[
\exp \left( -\frac{\sum_i (\eta_i^2 - 2\eta_i\nu + \nu^2)}{2\delta} \right) \times \exp \left( -\frac{\nu^2 - 2\nu\phi + \phi^2}{2\theta} \right) \\
\propto \exp \left( -\frac{1}{2} \left[ -2\nu \sum_i \eta_i + \nu^2 + \nu^2 - 2\nu\phi \right] \right) \\
= \exp \left( -\frac{1}{2} \nu^2 \left( \frac{I}{\delta} + \frac{1}{\theta} \right) - 2\nu \left( \frac{\nu}{\delta} + \frac{\phi}{\theta} \right) \right) \\
= \exp \left( -\frac{1}{2} \left[ \frac{I}{\delta} + \frac{1}{\theta} \right] \nu^2 - 2\nu \left( \frac{\nu}{\delta} + \frac{\phi}{\theta} \right) \left( \frac{I}{\delta} + \frac{1}{\theta} \right) \right) \\
\Rightarrow \nu_\eta \sim \text{Normal} \left( \frac{\nu}{\delta} + \frac{\phi}{\theta}, \left( \frac{I}{\delta} + \frac{1}{\theta} \right) \right) \\
\]

For the given priors,

\[
\nu_\eta \sim \text{Normal} \left( \frac{\nu}{\delta} + 2 \left( \frac{I}{\delta} + 1 \right)^{-1}, \left( \frac{I}{\delta} + 1 \right)^{-1} \right) \\
\]

\nu_\alpha:

The posterior for \nu_\alpha is derived similarly to that for \nu_\eta. As such, the derivation will not be included in the interest of saving space.

\[
\nu_\alpha \sim \text{Normal} \left( \frac{\nu}{\delta} + \frac{\phi}{\theta}, \left( \frac{I}{\delta} + \frac{1}{\theta} \right) \right) \\
\]

For the given priors,

\[
\nu_\alpha \sim \text{Normal} \left( \frac{\nu}{\delta} + 4 \left( \frac{I}{\delta} + 4 \right)^{-1}, \left( \frac{I}{\delta} + 4 \right)^{-1} \right) \\
\]
\( \nu_\beta: \)

The posterior for \( \nu_\beta \) is derived similarly to that for \( \nu_\eta \). As such, the derivation will not be included in the interest of saving space.

\[
\nu_\beta | \cdot \sim \text{Normal} \left( \left( \frac{I \nu_\beta}{\bar{\delta}_\beta} + \frac{\phi_\beta}{\theta_\beta} \right) \left( \frac{I}{\bar{\delta}_\beta} + \frac{1}{\theta_\beta} \right)^{-1}, \left( \frac{I}{\bar{\delta}_\beta} + \frac{1}{\theta_\beta} \right)^{-1} \right)
\]

For the given priors,

\[
\nu_\beta | \cdot \sim \text{Normal} \left( \left( \frac{I \nu_\beta}{\bar{\delta}_\beta} \right) \left( \frac{I}{\bar{\delta}_\beta} + 4 \right)^{-1}, \left( \frac{I}{\bar{\delta}_\beta} + 4 \right)^{-1} \right)
\]

\( \delta_\eta: \)

\[
f(\delta_\eta | \cdot) \propto \delta_\eta^{-\frac{1}{2}} \exp \left( -\frac{\sum_i (\eta_i - \nu_\eta)^2}{2\delta_\eta} \right) \delta_\eta^{-a_\eta-1} \exp \left( -\frac{b_\eta}{\delta_\eta} \right)
\]

\[
= \delta_\eta^{-\frac{1}{2}-a_\eta-1} \exp \left( -\frac{1}{\delta_\eta} \left[ \frac{\sum_i (\eta_i - \nu_\eta)^2}{2} + b_\eta \right] \right)
\]

\[\Rightarrow \delta_\eta | \cdot \sim \text{Inverse-Gamma} \left( \frac{I}{2} + a_\eta, \frac{\sum_i (\eta_i - \nu_\eta)^2}{2} + b_\eta \right)
\]

For the given priors,

\[\delta_\eta | \cdot \sim \text{Inverse-Gamma} \left( \frac{I}{2} + 2, \frac{\sum_i (\eta_i - \nu_\eta)^2}{2} + \frac{1}{4} \right)
\]

\( \delta_\alpha: \)
The posterior for $\delta_\alpha$ is derived similarly to that for $\delta_\eta$. As such, the derivation will not be included in the interest of saving space.

$$
\delta_\alpha \mid \cdot \sim \text{Inverse-Gamma} \left( \frac{I}{2} + a_\alpha, \frac{\sum_i (\alpha_i - \nu_\alpha)^2}{2} + b_\alpha \right)
$$

For the given priors,

$$
\delta_\alpha \mid \cdot \sim \text{Inverse-Gamma} \left( \frac{1}{2} + 2, \frac{\sum_i (\alpha_i - \nu_\alpha)^2}{2} + \frac{1}{4} \right)
$$

$\delta_\beta$:

The posterior for $\delta_\beta$ is derived similarly to that for $\delta_\eta$. As such, the derivation will not be included in the interest of saving space.

$$
\delta_\beta \mid \cdot \sim \text{Inverse-Gamma} \left( \frac{I}{2} + a_\beta, \frac{\sum_i (\beta_i - \nu_\beta)^2}{2} + b_\beta \right)
$$

For the given priors,

$$
\delta_\beta \mid \cdot \sim \text{Inverse-Gamma} \left( \frac{1}{2} + 2, \frac{\sum_i (\beta_i - \nu_\beta)^2}{2} + \frac{1}{4} \right)
$$

### 2.6.6 Specific Posterior Derivations

The model-specific posteriors are as follows.
Serial Model

In this model, all $\gamma_i$ are equal to 0. Consequently, there is no need for estimation of either $\nu_\gamma$ or $\delta_\gamma$ parameters, as they effectively do not exist for this model.

General Model

In this model, the prior distribution over each $\gamma_i$ parameter is similar to the priors for the $\eta_i$, $\alpha_i$, and $\beta_i$ parameters. As such, the posterior derivations for $\gamma_i$, $\nu_\gamma$, and $\delta_\gamma$ are similar to those for $\eta_i$, $\nu_\eta$, $\delta_\eta$, etc. and, consequently, are withheld in this section for space, with only the posteriors defined as follows.

$\gamma_i$:

\[
\gamma_i | \ldots \sim \text{Normal} \left( \frac{\sum_j \sum_k \sum_l \bar{y}_{ijkl}(-1)^{j+k} - L_i \gamma_i \eta_i - L_i \alpha_i \beta_i}{\sigma^2} + \frac{\nu_\gamma}{\delta_\gamma} \right) \\
\times \left( \frac{L_i}{\sigma^2 + \frac{1}{\delta_\gamma}} \right)^{-1}, \left( \frac{L_i}{\sigma^2 + \frac{1}{\delta_\gamma}} \right)^{-1}
\]

$\nu_\gamma$:

\[
\nu_\gamma | \ldots \sim \text{Normal} \left( \left( \frac{I_i \gamma + \phi_\gamma}{\delta_\gamma} \right) \left( \frac{I}{\delta_\gamma} + \frac{1}{\theta_\gamma} \right)^{-1}, \left( \frac{I}{\delta_\gamma} + \frac{1}{\theta_\gamma} \right)^{-1} \right)
\]
For the given priors,

\[ \nu_{\gamma} | \ldots \sim \text{Normal}\left( \frac{1}{\delta_{\gamma}}, \frac{1}{\delta_{\gamma} + 4}, \frac{1}{\delta_{\gamma} + 4} \right) \]

\[ \delta_{\gamma}: \]

\[ \delta_{\gamma} | \ldots \sim \text{Inverse-Gamma}\left( \frac{1}{2} + a_{\gamma}, \frac{\sum_{i}(\gamma_i - \nu_{\gamma})^2}{2} + b_{\gamma} \right) \]

For the given priors,

\[ \delta_{\gamma} | \ldots \sim \text{Inverse-Gamma}\left( \frac{1}{2} + 2, \frac{\sum_{i}(\gamma_i - \nu_{\gamma})^2}{2} + \frac{1}{4} \right) \]

**Parallel Model**

In this model, all \( \gamma \) parameters follow a Truncated-Normal distribution that is limited between \(-\infty\) and 0. The density for \( \gamma \) in this case is:

\[ f(\gamma) = (2\pi\delta_{\gamma})^{-\frac{1}{2}} \exp\left( -\frac{(\gamma - \nu_{\gamma})^2}{2\delta_{\gamma}} \right) \mathcal{I}(\gamma < 0) \Phi\left( \frac{0 - \nu_{\gamma}}{\sqrt{\delta_{\gamma}}} \right)^{-1} \]

\[ = (2\pi\delta_{\gamma})^{-\frac{1}{2}} \exp\left( -\frac{(\gamma - \nu_{\gamma})^2}{2\delta_{\gamma}} \right) \mathcal{I}(\gamma < 0) \Phi\left( \frac{-\nu_{\gamma}}{\sqrt{\delta_{\gamma}}} \right)^{-1} \]

where \( \mathcal{I}(\gamma < 0) \) is an indicator function that is equal to 1 when \( \gamma \) is less than 0 and is equal to 0 otherwise and \( \Phi \) is the cumulative density function of a standard normal distri-
\[ f(\gamma|\ldots) \propto \exp\left( -\frac{\sum_j \sum_k \sum_l \gamma_i^2 - 2\gamma_i \gamma_i (-1)^{j+k} + 2\eta_i \gamma_i (-1)^{j+k} + 2\alpha \gamma_i (-1)^k}{2\sigma^2} + 2\beta \gamma_i (-1)^j \right) \times \exp\left( -\frac{\gamma_i^2 - 2\nu \gamma_i + \nu_i^2}{2\delta} \right) \mathcal{I}(\gamma < 0) \\
= \exp\left( -\frac{1}{2} \left[ \gamma_i^2 \left( \frac{L_{ij} \eta_i + \gamma_i^2}{\sigma^2 + \frac{1}{\delta_i}} \right) - 2\gamma_i \left( \frac{\sum_j \sum_k \sum_l \gamma_i (-1)^{j+k} - L_{ij} \eta_i - L_{ij} \alpha + L_{ij} \beta_i}{\sigma^2} \right) \right] + \frac{\nu_i}{\delta} \right) \mathcal{I}(\gamma < 0) \\
\Rightarrow \gamma|\ldots \sim \text{Truncated-Normal} \left( \nu_i', \delta_i', \{-\infty, \ldots, 0\} \right) 
\]

for \( \delta_i' = \left( \frac{L_{ij}}{\sigma^2} + \frac{1}{\delta_i} \right)^{-1} \)
and \( \nu_i' = \left( \frac{\sum_j \sum_k \sum_l \gamma_i (-1)^{j+k} - L_{ij} \eta_i - L_{ij} \alpha + L_{ij} \beta_i}{\sigma^2} \right) + \frac{\nu_i}{\delta} \). 

\[ f(\nu_i|\ldots) \propto \exp\left( -\frac{\nu_i - \nu_i^2}{2\delta} \right) \Phi\left( \frac{-\nu_i - \nu_i^2}{2\theta} \right) \exp\left( -\frac{(\nu_i - \phi_i)^2}{2\theta_i} \right) \\
= \exp\left( -\frac{1}{2} \left[ \nu_i^2 - 2\nu_i \nu_i + \nu_i^2 + \frac{\nu_i^2 - 2\nu_i \phi_i + \phi_i^2}{\theta_i} \right] \right) \Phi\left( \frac{-\nu_i - \phi_i^2}{2\theta} \right) \\
\approx \exp\left( -\frac{1}{2} \left[ -2\nu_i \nu_i + \nu_i^2 + \frac{\nu_i^2 - 2\nu_i \phi_i + \phi_i^2}{\theta_i} \right] \right) \Phi\left( \frac{-\nu_i - \phi_i^2}{2\theta} \right) \\
= \exp\left( -\frac{1}{2} \left[ \nu_i^2 \left( \frac{I}{\delta} + \frac{1}{\theta_i} \right) - 2\nu_i \left( \frac{I}{\delta} + \phi_i \theta_i \right) \right] \right) \Phi\left( \frac{-\nu_i - \phi_i^2}{2\theta} \right) 
\]

This distribution is not a Normal distribution. Thus, the Normal distribution is not a con-
jugate prior for this parameter. This distribution is also not of a commonly known form, as no common distribution has a density composed of a product of a Standard Normal density function scales by a Standard Normal CDF. Consequently, Metropolis-Hastings sampling is needed to estimate this parameter. Also, keep in mind that, because of the sign in the normalizing constant from the truncated distribution, positive values of $\nu_\gamma$ are weighted more heavily than negative values, even though all values of $\gamma_i$ in this model are negative. This may prove to be problematic as this may result in the chain failing to converge on a final estimate, possibly leading to non-identifiability in the $\gamma_i$ and $\delta_\gamma$ posteriors.

$\delta_\gamma$:

$$
\begin{align*}
\delta_\gamma : \\
\quad f(\delta_\gamma | \ldots) &\propto (\delta_\gamma)^{-\frac{1}{2}} \exp \left( -\frac{\sum_i (\gamma_i - \nu_\gamma)^2}{2\delta_\gamma} \right) \Phi \left( \frac{-\nu_\gamma}{\sqrt{\delta_\gamma}} \right)^{-1} \delta_\gamma^{-\alpha_\gamma - 1} \exp \left( \frac{b_\gamma}{\delta_\gamma} \right) \\
&= \delta_\gamma^{-\frac{1}{2} - \alpha_\gamma - 1} \exp \left( - \frac{1}{\delta_\gamma} \left[ \frac{\sum_i (\gamma_i - \nu_\gamma)^2}{2} + b_\gamma \right] \right) \Phi \left( \frac{-\nu_\gamma}{\sqrt{\delta_\gamma}} \right)^{-1}
\end{align*}
$$

This distribution is not an Inverse-Gamma. This, the Inverse-Gamma distribution is not a conjugate prior for this parameter. This distribution is also not of a commonly known form, as no common distribution has a density composed of a product of an Inverse-Gamma density function scales by a Standard Normal CDF. Consequently, Metropolis-Hastings sampling is needed to estimate this parameter.

**Coactive Model**

Aside from the support set for the $\gamma$ parameters, this model is identical to the parallel model. As such, the derivations are, for the most part, identical. The only difference is in
the normalizing constants used in the truncated-normal distributions, as illustrated here:

\[
f(\gamma) = (2\pi\delta\gamma)^{-\frac{1}{2}} \exp\left(-\frac{(\gamma_i - \nu\gamma)^2}{2\delta\gamma}\right) \mathcal{I}(\gamma_i > 0) \left(1 - \Phi\left(\frac{0 - \nu\gamma}{\sqrt{\delta\gamma}}\right)\right)^{-1}
\]

\[
= (2\pi\delta\gamma)^{-\frac{1}{2}} \exp\left(-\frac{(\gamma_i - \nu\gamma)^2}{2\delta\gamma}\right) \mathcal{I}(\gamma_i > 0) \Phi\left(\frac{\nu\gamma}{\sqrt{\delta\gamma}}\right)^{-1}
\]

with the values of 1 and 0 returned by the indicator function corresponding to values of gamma greater than and not greater than 0, respectively, and with the sign of the term in the Standard-Normal CDF term switched.

Because of this similarity, only the posterior distributions are listed here. Note that this means the posterior distributions for \(\nu\gamma\) and \(\delta\gamma\) in this model are also neither conjugate nor of known form, necessitating Metropolis-Hastings steps for estimation and potentially encountering the same failures.

\(\gamma\):

\[
\gamma_i | \ldots \sim \text{Truncated-Normal} \left(\nu'_{\gamma}, \delta'_{\gamma}, \{0, \ldots, \infty\}\right)
\]

\(\nu\gamma\):

\[
f(\nu\gamma | \ldots) \propto \exp\left(-\frac{1}{2} \left[\nu^2 \left(\frac{I_{\gamma}}{\delta_{\gamma}} + \frac{1}{\theta_{\gamma}}\right) - 2\nu\gamma \left(\frac{I_{\gamma}}{\delta_{\gamma}} + \frac{\phi_{\gamma}}{\theta_{\gamma}}\right)\right]\right) \Phi\left(\frac{\nu\gamma}{\sqrt{\delta\gamma}}\right)^{-I}
\]
\[ f(\delta_\gamma, \ldots) \propto \delta_\gamma^{\frac{1}{2} - \alpha_\gamma - 1} \exp \left( -\frac{1}{\delta_\gamma} \left[ \sum_i (\gamma_i - \nu_\gamma)^2 + b_\gamma \right] \right) \Phi \left( \frac{-\nu_\gamma}{\sqrt{\delta_\gamma}} \right)^{-1} \]

### 2.7 Density Derivations

Because we are using CMDE, the conditional posterior density derivations for the interaction parameters described above can also be used for the conditional marginal density estimation needed in the Savage-Dickey ratio. Additionally, because the priors on the interaction parameters are conjugate priors, we know the full conditional posterior density and do not need to integrate to derive the normalizing constants. Consequently, the estimated posterior density on each iteration is equal to the product of the densities, at zero, of the distribution of the interaction parameters \( \gamma_i \) conditional on the current values of the hyperparameters \( \nu_\gamma \) and \( \delta_\gamma \). The prior density is estimated similarly with the exception that the hyperparameters are sampled from their prior distributions instead of their conditional posterior distributions.

It should be noted that, because of the typically small values densities take on in high-dimensional spaces, accurate calculation of the densities can be difficult. This precision issue can be remedied by sampling the log-densities instead of the densities directly. This also avoids potential boundary issues, as the set of all log-densities is unbounded whereas densities are always greater than zero.

Additionally, by using log-densities instead of densities, we can calculate the log-Bayes Factor by taking the difference between the posterior and prior log-densities. Log-Bayes Factors can be considered easier to interpret than straight Bayes Factors because the sign of the log-Bayes Factor directly identifies the preferred model. Additionally, the equivalence
point, the point where evidence is equivocal with regard to both models, for log-Bayes Factors is 0 instead of the 1 observed with Bayes Factors. Last, the changes are put on an easier to interpret “additive” scale, as opposed to the “multiplicative” scale odds ratios typically exhibit.

It is also important to note that, during development, an error in the estimation of Bayes Factors occurred in which it was possible to find evidence in favor of both the Parallel and Coactive models relative to the Serial model but which favored the Serial model over the General mode. Considering that the Parallel and Coactive models are both restrictions of the General model that make opposing predictions, this doesn’t make sense. At the recommendation of Dr. Richard Morey and Dr. Jeff Rouder, I had added in a correction factor to account for this, under the assumption that this failure was due to the Serial portion of the Bayes Factor ratio differing with the other model being considered. After subsequent analyses, this assumption was found to be mistaken as the ordering of the Bayes Factors was actually due to misspecification in the sampling methods (specifically, an excessively large acceptance probability in the Metropolis-Hastings samplers used for the hierarchical means in the Parallel and Coactive models). This error has been corrected and no “correction factor” is actually needed.

2.8 Issues

As stated earlier, the Normal framework has two major issues: it fails to accurately specify both the positive skew and the direct relation between mean and variance in reaction-time data. Reaction-time data are known to be positively skewed and the Normal framework cannot account for this skew because the Normal distribution has no skew by definition. More importantly, reaction-time data also have variances that are directly related to their means. That is, subsets of reaction-time data with larger means are also more variable than subsets with smaller means. The Normal framework cannot account for this change
because the mean and variance are independent of each other. Although this second issue
could be accounted for with a reparameterization, later frameworks were instead developed
with completely different distributional assumptions, primarily because these frameworks
take advantage of distributions known to have both characteristics. These frameworks, the
Ex-Gaussian and Gamma frameworks, will be described in more detail in the next chapter.
Chapter 3

Additional Frameworks

In this chapter I will describe the two additional frameworks developed in this project and how they relate to the analytic goals described in the previous chapters.

3.1 Ex-Gaussian Framework

The third framework I attempted to use for this analysis was the ’Ex-Gaussian’ framework, which specified the following:

\[ y_{ijkl} \sim \text{Ex-Gaussian} \left( \mu_{ijk}, \sigma^2, \mu_{ijk} \lambda \right) \]  

for participant \( i = 1, \ldots, I \), change levels \( j = 1, 2 \) and \( k = 1, 2 \) for the first and second features, respectively, and repetition \( l = 1, \ldots, L_{ijk} \) of the \( jk^{th} \) condition for participant \( i \).

This is equivalent to saying that reaction time \( y_{ijkl} \) is equal to the sum of two independent random variables \( S_{ijk} \) and \( T_{ijk} \) where:

\[ S_{ijk} \sim \text{Normal} \left( \mu_{ijk}, \sigma^2 \right) \]  

(3.2)
for mean parameter $\mu_{ijk}$, variance parameter $\sigma^2$, and scale parameter $\lambda$. If $\mu_{ijk}$ is defined as in the Normal framework, most of the structure follows similarly from the Normal framework, with the constraint on $\lambda$

$$\lambda \sim \text{Gamma}(\kappa, \tau)$$

(3.4)

being the primary addition.

This similarity to the Normal framework is one of the primary motivations for this extension. Aside from the additional parameters needed to fully specify the Ex-Gaussian distribution, most of the framework follows directly from the Normal framework, so only the additional parameterization needs to be figured. Additionally, the extra parameters included help to resolve both the issue of skewed data and the issue of non-independence of the mean and variance, both of which were critical faults with the Normal framework.

Because the mean of an Ex-Gaussian distribution is equal to $\mu_{ijk} + \mu_{ijk}\lambda$, the MIC in this framework is calculated as follows:

$$\text{MIC}_i = (\mu_{i11} + \mu_{i21}\lambda) + (\mu_{i22} + \mu_{i22}\lambda) - (\mu_{i12} + \mu_{i12}\lambda) - (\mu_{i21} + \mu_{i21}\lambda)$$

(3.5)

$$= \mu_{i11}(1 + \lambda) + \mu_{i22}(1 + \lambda) - \mu_{i12}(1 + \lambda) - \mu_{i21}(1 + \lambda)$$

(3.6)

$$= (1 + \lambda)(\mu_{i11} + \mu_{i22} - \mu_{i12} - \mu_{i21})$$

(3.7)

$$= (1 + \lambda)\gamma_i$$

(3.8)

Using the ANOVA-like approach described in Chapter 2, this simplifies to

$$\text{MIC}_i = 4(1 + \lambda)\gamma_i$$

(3.9)
And, because $\lambda$ follows a Gamma distribution, it is strictly positive, meaning that placing sign constraint on the interaction term would suffice. As such, the same priors that were applied to the interaction term in the Normal framework can be used in this framework.

However, due to computational complexity issues that arose during development, this framework was eventually abandoned in favor of a similar one with fewer parameters: the Gamma framework.

### 3.2 Gamma Framework

The fourth framework I attempted to use for this analysis was the 'Gamma' framework, which specified the following:

$$y_{ijkl} \sim \text{Gamma}(\theta, \mu_{ijk})$$  \hspace{1cm} (3.10)

for participant $i = 1,...I$, change levels $j = 1,2$ and $k = 1,2$ for features 1 and 2, respectively, and repetition $l = 1,...,L_{ijk}$ of the $jk^{th}$ condition for participant $i$.

In the Gamma framework, the within-condition mean is equal to $\theta \mu_{ijk}$. Consequently,

$$MIC_i = (\theta \mu_{i11}) + (\theta \mu_{i22}) - (\theta \mu_{i12}) - (\theta \mu_{i21})$$  \hspace{1cm} (3.11)

$$= \theta (\mu_{i11} + \mu_{i22} - \mu_{i12} - \mu_{i21})$$  \hspace{1cm} (3.12)

$$= 4 \theta \gamma_i$$  \hspace{1cm} (3.13)

Using the ANOVA-like approach described in Chapter 2, this simplifies to

$$MIC_i = 4 \theta \gamma_i$$  \hspace{1cm} (3.14)

Because $\theta$ is strictly positive, placing sign constraint on the interaction is sufficient. As
such, the same priors that were applied to the interaction term in the Normal framework can be used in this framework. However, because of certain consequences of this framework, some of the non-interaction parameters do not necessarily need to follow a normal distribution.

Some of the derivation details for the Gamma framework are described below, with the rest described in the supplementary materials:

For $i = 1,...I$, $j = 1,2$, $k = 1,2$, $l = 1,...,L_{ijk}$, $m = 1,...,4$, the common assumptions are as follows:

\[
\begin{align*}
 y_{ijkl} & \sim \text{Gamma}\left(\theta, \mu_{ijk}\right) \\
 \mu_{ijk} &= \eta_i + \alpha_i(-1)^j + \beta_i(-1)^k + \gamma_i(-1)^{j+k} \\
 \theta & \sim \text{Gamma}\left(\kappa_{\theta}, \tau_{\theta}\right) \\
 \eta_i & \sim \text{Gamma}\left(\kappa_{\eta}, \tau_{\eta}\right) \\
 \alpha_i & \sim \text{Normal}\left(\nu_{\alpha}, \varepsilon_{\alpha}\right) \\
 \beta_i & \sim \text{Normal}\left(\nu_{\beta}, \varepsilon_{\beta}\right) \\
 \nu_{\alpha} & \sim \text{Normal}\left(\zeta_{\alpha}, \lambda_{\alpha}\right) \\
 \nu_{\beta} & \sim \text{Normal}\left(\zeta_{\beta}, \lambda_{\beta}\right) \\
 \varepsilon_{\alpha} & \sim \text{Inverse-Gamma}\left(\kappa_{\alpha}, \tau_{\alpha}\right) \\
 \varepsilon_{\beta} & \sim \text{Inverse-Gamma}\left(\kappa_{\beta}, \tau_{\beta}\right)
\end{align*}
\]

In this framework, $\eta_i$ follows a Gamma distribution instead of a Normal distribution as an attempt to keep the (typically largest portion of the) scale parameter positive. However, aside from making a slight change to how $\eta_i$ is sampled, this has no substantial effect on
the other parameters.

As in the Normal framework, the specific model assumptions are as follows:

### 3.2.1 Serial Model

Because the MIC predicted by serial processing is zero, this model assumes that all interaction parameters, $\gamma_i$, are exactly zero. Because these parameters are all constant, there is no need for any further hierarchical specification in this model.

### 3.2.2 General Model

The general model places minimal constraint on the possible range of the interaction parameters. For the purposes of this current research, we will assume that the interaction parameters each follow a normal distribution with a hierarchical structure described as follows:

\[
\begin{align*}
\gamma_i & \sim \text{Normal} (\nu_\gamma, \epsilon_\gamma) \\
\nu_\gamma & \sim \text{Normal} (\zeta_\gamma, \lambda_\gamma) \\
\epsilon_\gamma & \sim \text{Inverse-Gamma} (\kappa_\gamma, \tau_\gamma)
\end{align*}
\]

where $\zeta_\gamma$, $\lambda_\gamma$, $\kappa_\gamma$, and $\tau_\gamma$ are hyperparameters that are specified before analysis. It should be noted that the assumptions in this model are very similar to those in the next two models. The only major change is in the distribution of the interaction parameter itself, with the hyperpriors on $\nu_\gamma$ and $\epsilon_\gamma$ remaining the same. As such, only the change to the distribution to $\gamma_i$ will be noted.
3.2.3 Parallel Model

Because parallel processing predicts negative MICs, this model assumes that all interaction parameters must be negative. This is accomplished by the following restriction:

\[ \gamma_i \sim \text{Truncated-Normal} \left( \nu, \epsilon, (-\infty, 0] \right) \]

with \((-\infty, 0]\) being the support set for these truncated distributions. This limitation to the support set constitutes the sign constraint needed for this model and only allows interaction parameters to equal zero in order to allow for Savage-Dickey estimation.

3.2.4 Coactive Model

Because coactive processing predicts positive MICs, this model assumes that all interaction parameters must be positive. This is accomplished by the following restriction:

\[ \gamma_i \sim \text{Truncated-Normal} \left( \nu, \epsilon, [0, \infty) \right) \]

with \([0, \infty)\) being the support set for these truncated distributions. This limitation to the support set constitutes the sign constraint needed for this model and only allows interaction parameters to equal zero in order to allow for Savage-Dickey estimation.

3.2.5 Joint Distribution and Posterior Derivation of Common Parameters

Because the common parameters are independent of the specific parameters and only depend on them when they are assumed to be known, the priors for the model-specific parameters are effectively normalizing constants that can be dropped during derivation up to proportionality. As such, those distributions are of no interest to the current joint analysis.
and will be omitted. Consequently, the joint distribution of all common parameters is as follows:

\[
f(Y, \theta, \vec{\eta}, \vec{\alpha}, \nu, \nu, \epsilon, \epsilon) = \prod_i \prod_j \prod_k \prod_l (\mu_{ijk})^{-\theta} \Gamma(\theta)^{-1} (y_{ijkl})^{-\theta - 1} \exp \left( -\frac{y_{ijkl}}{\mu_{ijkl}} \right) \times \]

\[
(\tau_\theta)^{-\kappa_\theta} \Gamma(\kappa_\theta)^{-1} (\theta)^{\kappa_\theta - 1} \exp \left( -\frac{\theta}{\tau_\theta} \right) \times \left( \prod_i (\tau_\eta)^{-\kappa_\eta} \Gamma(\kappa_\eta)^{-1} (\eta_i)^{\kappa_\eta - 1} \exp \left( -\frac{\eta_i}{\tau_\eta} \right) \right) \times \]

\[
\prod_i (2\pi \varepsilon_\alpha)^{-\frac{1}{2}} \exp \left( -\frac{(\alpha_i - \nu_\alpha)^2}{2\varepsilon_\alpha} \right) \times \prod_i (2\pi \varepsilon_\beta)^{-\frac{1}{2}} \exp \left( -\frac{(\beta_i - \nu_\beta)^2}{2\varepsilon_\beta} \right) \times \]

\[
(2\pi \lambda_\alpha)^{-\frac{1}{2}} \exp \left( -\frac{(\nu_\alpha - \zeta_\alpha)^2}{2\lambda_\alpha} \right) \times (2\pi \lambda_\beta)^{-\frac{1}{2}} \exp \left( -\frac{(\nu_\beta - \zeta_\beta)^2}{2\lambda_\beta} \right) \times \]

\[
(\tau_\alpha)^{\kappa_\alpha} \Gamma(\kappa_\alpha)^{-1} (\epsilon_\alpha)^{-\kappa_\alpha - 1} \exp \left( -\frac{\tau_\alpha}{\epsilon_\alpha} \right) \times (\tau_\beta)^{\kappa_\beta} \Gamma(\kappa_\beta)^{-1} (\epsilon_\beta)^{-\kappa_\beta - 1} \exp \left( -\frac{\tau_\beta}{\epsilon_\beta} \right) \times \]

Unfortunately, none of the prior distributions are conjugate priors, nor are most of the hyperprior distributions. More importantly, most of the conditional posterior distributions do not have a frequently-used functional form, as can be seen from the example below. Consequently, most of the chains for this framework will rely largely on Metropolis steps to generate adequate samples.

An example of the general lack of a clean functional form can be seen in the posterior distribution of \( \alpha_i \):

\[
f(\alpha_i | \cdot) \propto \prod_j \prod_k \prod_l (\mu_{ijk})^{-\theta} \exp \left( -\frac{y_{ijkl}}{\mu_{ijkl}} \right) \times \exp \left( -\frac{(\alpha_i - \nu_\alpha)^2}{2\varepsilon_\alpha} \right) \]

\[
= \prod_j \prod_k \prod_l (\eta_i + \alpha_i (-1)^j + \beta_i (-1)^k + \gamma_i (-1)^{j+k})^{-\theta} \]

50
\[
\exp\left(-\frac{y_{ijkl}}{\eta_i + \alpha_i (-1)^j + \beta_i (-1)^k + \gamma_i (-1)^{j+k}}\right) \times \exp\left(-\frac{(\alpha_i - v_\alpha)^2}{2\varepsilon_\alpha}\right)
\]

The main problem is the \((\mu_{ijk})^{-\theta}\) term, which becomes more complex as \(\theta\) increases and, for \(\theta \neq 1\), prevents separation of any component from the others. This might not be as much of an issue if \(\theta\) is held constant, which will be done in current work as earlier Gamma framework development has found that chain estimates of the \(\theta\) parameter seem to climb until the sample space for the parameters of interest becomes too small to analyze, but derivation of the posterior distribution was still done in the interest of future refinement. Because this is a common issue for all parameters (and because the other parameters have similarly non-standard posterior distributions), the posterior distributions for most parameters are omitted here for brevity and, along with the analysis code, will be included in supplementary materials for any interested parties. However, the posterior distributions for the model-specific interaction terms will be included subsequently for reference.

### 3.2.6 Specific Posterior Derivations

The posterior distributions for the hierarchical parameters, \(v_\gamma\) and \(\varepsilon_\gamma\) in the Gamma framework are effectively identical to their Normal framework counterparts, \(v_\gamma\) and \(\delta_\gamma\), respectively, as the specific distributions involved are the same. Consequently, the reader is referred to their derivation in the previous chapter, and only the posterior distribution for the interaction parameters themselves will be included here.
**Serial Model**

Just like in the Normal framework, the Serial model assumes that all interaction parameters, $\gamma$, are equal to 0. As such, there is no need to estimate either $\nu_\gamma$ or $\varepsilon_\gamma$, as they effectively do not exist for this model.

**General Model**

In this model, the prior distribution over each $\gamma_i$ parameter is similar to the priors for the $\alpha_i$ (described above) and $\beta_i$ parameters. As such, the posterior derivations for $\gamma_i$ are also similar to the posteriors for $\alpha_i$ and $\beta_i$:

$$
\begin{align*}
    f(\gamma_i|\cdot) &\propto \left( \prod_j \prod_k \prod_l (\mu_{ijkl})^{-\theta} \exp \left( -\frac{y_{ijkl}}{\mu_{ijkl}} \right) \right) \times \exp \left( -\frac{(\gamma_i - \nu_\gamma)^2}{2\varepsilon_\gamma} \right) \\
        &= \left( \prod_j \prod_k \prod_l (\eta_i + \alpha_i (-1)^j + \beta_i (-1)^k + \gamma_i (-1)^{j+k})^{-\theta} \times \exp \left( -\frac{y_{ijkl}}{\eta_i + \alpha_i (-1)^j + \beta_i (-1)^k + \gamma_i (-1)^{j+k}} \right) \right) \times \exp \left( -\frac{(\gamma_i - \nu_\gamma)^2}{2\varepsilon_\gamma} \right)
\end{align*}
$$

This derivation is not extended much further because, as described earlier, there is no simple expansion of the $(\mu_{ijkl})^{-\theta}$ term when $\theta$ is not constant. Additionally, because this term is multiplied across iterations within conditions, the actual scale of the exponent is substantially larger than $\theta$, further complicating any attempt at expansion. However, it must be noted that the un-normalized posterior density is still easy to calculate as it is effectively proportional to the product of a normal density and a series of gamma densities, with all
parameters known.

**Parallel Model**

In this model, all $\gamma_i$ parameters follow a Truncated-Normal distribution that is limited between $-\infty$ and 0. The density for $\gamma_i$ in this case is:

$$f(\gamma_i | \cdot) \propto \left( \prod_j \prod_k \prod_l (\mu_{ijk})^{-\theta} \exp\left( -\frac{y_{ijkl}}{\mu_{ijk}} \right) \times \exp\left( -\frac{(\gamma_i - \nu_{\gamma})^2}{2\varepsilon_{\gamma}} \right) \times \mathcal{I}(\gamma_i < 0) \right)$$

Aside from the indicator function used to enforce the restriction, the posterior distribution is effectively identical to that of the General Model. Because of this, this density can be calculated in roughly the same manner.

**Coactive Model**

Aside from the support set for the $\gamma_i$ parameters, this model is identical to the parallel model. As such, the derivations are, for the most part, identical. The only differences are in the indicator function and Normal-distribution CDF used in the truncated-normal distributions, with only the indicator being included in the posterior distribution as the CDF correction does not explicitly depend on $\gamma_i$ and, consequently, is constant with respect to it. Because of this similarity, only the posterior distribution is listed here.
Again, because this is proportional to a product of a few well-known distributions, this functional form is relatively easy to calculate, even though the normalizing constant is not known.

3.3 Density Derivations

Unlike in the Normal Framework, the posterior densities are not fully determined. Although the functional form is known, CMDE requires that the normalizing constant is also known. The normalizing constant can be determined with a numerical integration step. Indeed, numerical integration is how the normalizing constant is determined in this application. However, much like in the Normal framework, the density that is being sampled is determined by the density function from which the sample is generated, which is the posterior distribution described in the previous section. Additionally, the densities calculated from the Parallel and Coactive models suffer from the same proportionality issue that was observed in the Normal framework. Consequently, the correction factor described in the previous chapter is applied similarly to the Gamma framework analysis.

For the purposes of this paper, both the Normal and Gamma frameworks will be used to analyze the data from the experiments detailed in later chapters. For both computational ease and ease of reporting, both of which were commented on in Chapter 2, log Bayes
Factors, with a logarithmic base of 10, will be reported instead. This has the benefit of making the Bayes Factor even easier to interpret, with positive values corresponding to Bayes Factors greater than 1 (support for the first model), negative values corresponding to Bayes Factors less than 1 (support for the second model), and values near 0 corresponding to relatively no evidence towards either side. Because this value is an exponent, the difference between a log Bayes Factor of 1 and 2 and the difference between 2 and 3 are multiplicative instead of additive, with the multiplicative factor being 10 in this case. That is, a log Bayes Factor of 2 provides 10 times as much support for the first model, relative to the second, as a log Bayes Factor of 1, and a log Bayes Factor of 3 provides 10 times as much support again as a log Bayes Factor of 2. Similarly, a log Bayes Factor of -2 provides 10 times as much support for the second model, relative to the first model, as a log Bayes Factor of -1, and so forth.

Now that both frameworks of interest have been described, I will explain the specifics of the experiments and analyses originally mentioned in Chapter 1. The following chapter will describe the general experimental method, the details of analysis, and the methods and results of the four experiments in this project.
Chapter 4

Methods and Results

4.1 Method Summary

There are four experiments in total for this paper: two stimulus types, Screwhead stimuli and Digit Pair stimuli, with two experiments, one perceptual and one memory, for each stimulus type. By running both experiments, we can determine if the forced-funneling signature described in Chapter 1 is present for each stimulus type and, thus, determine if the stimulus is one that would typically be chunked.

4.1.1 Stimuli Used

The stimuli in question are Screwhead stimuli (top left in Figure 4.1), which consist of a circle with a single diameter line and which vary in size and the orientation of the diameter line, and Digit Pairs (bottom right in Figure 4.1), which are fairly self-explanatory and vary in the tens and ones digits. The Screwhead stimuli are borrowed from prior research in feature integrality (Garner & Morton, 1969; Garner & Felfoldy, 1970; Ashby & Maddox, 1990; Ashby & Maddox, 1993), whereas the Digit Pairs have been adapted from prior research on the "Distance from Five” effect (Moyer & Landauer, 1967; Link, 1990; Rouder
Figure 4.1: Examples of the two stimuli used in this research: Screwhead stimuli (left), which vary in size and the orientation of the dividing line, and Digit Pairs (right), which vary from 1 through 9 for each digit.

4.1.2 General Method

The general trial schematics for the two experiment types, the perceptual experiment and the memory experiment, are depicted in Figure 4.2 and Figure 4.3, respectively. In perceptual experiment trials, participants are provided with two stimuli that do or do not vary from each other and are asked to compare them and state whether or not both features differ. Memory experiment trials follow a similar scheme with the primary exception being that participants are only shown the items one at a time and must hold the first item in memory in order to compare it to the second item. In both cases, the stimuli will either match on a feature or differ by either a “small” or a “large” amount, with the exact amounts defined for each experiment.
Figure 4.2: General Trial Schematic of the Perceptual experiment trials (with Screwhead stimuli for an example). Participants see a fixation cross at the center before seeing both stimuli, which remain presented until the participant provides a response.
Figure 4.3: General Trial Schematic of the Memory experiment trials (with Screwhead stimuli for an example). Participants see a fixation cross at the center before seeing one stimulus for a short interval (200 ms listed here, but might differ in specific experiments). After this interval, the screen is left blank for another waiting period (1000 ms listed here, but might differ in specific experiments) before the second stimulus is presented, with the second stimulus displayed until a response is provided.
4.1.3 Predictions

Perceptual processing of the Screwhead and Digit Pairs are expected to be a mix of both serial and parallel and parallel, respectively. The processing architecture used in memory was not specified beforehand for either stimuli. Consequently, the perceptual portions of the tasks are considered confirmatory and the memory portions exploratory.

4.2 Analysis Summary

Both the Normal and Gamma frameworks were used in these analyses. Inaccurate participant responses to the four change conditions were removed before analysis, as were responses faster than 300 ms and slower than 5000 ms. This cleaning resulted in the removal of about 15% of the overall analyzable data. Most of this cleaning occurred in the first two experiments, as participant accuracies were substantially lower with the Screwhead stimuli than with the Digit Paris stimuli. Although responses to non-change stimuli were also removed, these data are not of primary interest to the analysis and, as such, are not included in the above cleaning proportions.

4.2.1 Prior Settings

For the purposes of the analyses in this paper, all of the hyperpriors that need to be specified beforehand are listed here and are common across all uses of this framework. Additionally, although the prior distributions of the interaction terms differ, the hyperprior structure for all but the serial model are effectively identical, with the only main difference resulting from the truncation of the distribution of the interaction terms themselves. Consequently, the relevant prior settings are common across all of those models. Any differences between this approach and what was actually done for any given experiment will be included in the details of the relevant experiment.
Normal Framework Priors

\[
\phi_{\eta} = 2 \\
\theta_{\eta} = 1 \\
a_{\eta} = 3 \\
b_{\eta} = .2 \\
\phi_{\alpha} = 0 \\
\theta_{\alpha} = .001 \\
a_{\alpha} = 3 \\
b_{\alpha} = .2 \\
\phi_{\beta} = 0 \\
\theta_{\beta} = .001 \\
a_{\beta} = 3 \\
b_{\beta} = .2 \\
\phi_{\gamma} = 0 \\
\theta_{\gamma} = .0001 \\
a_{\gamma} = 3 \\
b_{\gamma} = .02 \\
a_{\sigma} = 2 \\
b_{\sigma} = .25
\]

Gamma Framework Priors

\[
\kappa_{\eta} = 5
\]
\[ \tau_\eta = .2 \]
\[ \xi_\alpha = 0 \]
\[ \lambda_\alpha = .001 \]
\[ \kappa_\alpha = 3 \]
\[ \tau_\alpha = .2 \]
\[ \xi_\beta = 0 \]
\[ \lambda_\beta = .001 \]
\[ \kappa_\beta = 3 \]
\[ \tau_\beta = .2 \]
\[ \xi_\gamma = 0 \]
\[ \lambda_\gamma = .0001 \]
\[ \kappa_\gamma = 3 \]
\[ \tau_\gamma = .02 \]

You will note that the prior parameters for the shape parameter \( \theta \) are not listed. This is due to \( \theta \) being treated as a constant, at \( \theta = 2 \), for the present analyses. This is being considered due to initial work in jointly estimating shape and other parameters resulting in non-convergence of the chains. Specifically, shape estimates continued to increase until the range of possible values for other parameters contracted to a null set, causing the chain to fail to run. The exact cause for this climb are unknown and will be addressed in the future, possibly with changes in the prior structure used, but will currently be ignored for practical purposes.

Example hyperpriors for the gamma parameter’s hierarchical location and scale parameters, as well as the marginal priors for both the gamma parameters and for the actual MICs, are depicted in Figure 4.4. Note that the priors for the gamma parameters between
the Normal and Gamma frameworks are the same. However, because of the differences in parameterization, the effective prior on the interaction effect itself is different. Specifically, the actual MIC is equal to $4\gamma_i$ in the Normal framework and to $4\theta_i\gamma_i$ in the Gamma framework and, consequently, the prior on the actual MIC in the Gamma framework is more diffuse than it is in the Normal framework with variation being $\theta^2$ times as large in the Gamma framework. This can be seen in the bottom-right plot in Figure 4.4, with the red line depicting the Normal framework’s prior on the actual MIC and the black line depicting the Gamma framework’s prior. Also note that, because the range of plausible values for the hierarchical scale parameter is so small (with most of the mass being below $0.04s^2$), there appears to be little difference between the marginal prior for the gamma parameter and the hyperprior for the corresponding hierarchical location parameter. This is highlighted in the bottom-left plot in Figure 4.4, with the hyperprior depicted in red and the marginal prior depicted in black.

4.2.2 Chain Runtimes and Output

The chains in the Normal framework analysis were run for 100000 iterations, with the first 55000 iterations used for burn-in and the remaining portions of the chains thinned by a factor of 50, producing sets of 900 iterations from which the results are computed.

The chains in the Gamma framework analysis were run using adaptive Metropolis-Hastings steps on the interaction parameter with adaptation (i.e. chain restarting and adjustment of the Metropolis bandwidth specific to the parameters being adapted) being run after acceptance probability checks over the prior 100 iterations upon completing 105 iterations. These checks were done to determine if the acceptance probabilities for each interaction parameter was between 0.45 and 0.6, with the chains restarting if at least one parameter had an acceptance probability outside of this range and continuing to the end otherwise. The full chains were run out to 10000 iterations with burn-ins of 5000. The
Figure 4.4: Hyperpriors for the hierarchical mean (top-left) and variance (top-right) of the interaction parameters in units of seconds and squared seconds, respectively. The bottom plots describe the conditional and marginal densities of the hierarchical interaction mean parameter and the interaction parameters, respectively, (left) and the marginal prior for the interaction parameters (in seconds) when scaled (by 4 for the Normal framework or by $4\theta$ for the Gamma framework, both of which are assumed to be dimensionless constants) to match the scale of the surface MIC estimates.
chains were not run as thoroughly as the Normal framework chains due to a substantial increase in computation time, with 100 iterations taking roughly 6 or 7 minutes (for a total runtime of 10 to 12 hours for each experiment), not including the hour or two needed for the initial adaptive phase, compared to the 15 minutes needed for the Normal framework chains to run in entirety for one experiment.

Although initial attempts to estimate the Bayes Factors with an even shorter chain (of length 2000) disagreed substantially with Bayes Factors calculated from the Normal framework, autocorrelations were low enough to suggest that there is little to no need for substantial burn-in or thinning (details will be included in the results section).

4.2.3 Simulation Analyses

Analyses with simulated data suggest that the Normal analysis is fairly accurate in predicting the model used in generating the data. The reaction time data were generated from a Truncated-Normal distribution with a lower limit of .3 seconds, an upper limit of 5 seconds, a variance of .400 seconds$^2$, and a set of four other parameters that determined the mean parameter of this generating distribution: mean reaction time ($\eta_i$), reaction time effect for the first feature ($\alpha_i$), reaction time effect for the second feature ($\beta_i$), and reaction time interaction effect ($\gamma_i$), all measured in seconds. The first three parameters were generated once for the set of 40 fictional participants and were otherwise identical through all used simulated data sets ($\eta_i \sim \text{Normal}(1.2,.01), \alpha_i \sim \text{Normal}(.1,.01), \beta_i \sim \text{Normal}(.1,.01)$). The interaction parameters, however, were derived differently for each simulated data set, with the generating distributions used for each set being listed in Table 4.1 along with the model simulated by these interaction parameters and the resulting Serial-Other log Bayes Factors when the Normal framework analysis was applied to the relevant simulated data. As can be seen, although this analysis may have some difficulties with distinguishing small amounts of true variation in MIC from a total lack of variation from 0, it otherwise accu-
Simulated Model | Interaction Generating Distribution | General | Parallel | Coactive
---|---|---|---|---
Serial | Degenerate(0) | 4.14 | 13.01 | 8.16
General 1 | Normal(0, .0025) | 2.92 | 7.40 | 10.94
General 2 | Normal(0, .01) | -16.15 | 0.61 | -1.62
General 3 | Normal(0, .04) | -81.88 | -21.83 | -36.65
Parallel 1 | Truncated − Normal(0, .0025, {−∞, 0}) | 0.08 | 0.04 | 17.24
Parallel 2 | Truncated − Normal(0, .01, {−∞, 0}) | -25.09 | -33.16 | 26.73
Coactive 1 | Truncated − Normal(0, .0025, {0, ∞}) | -7.03 | 26.54 | -15.54
Coactive 2 | Truncated − Normal(0, .01, {0, ∞}) | -19.76 | 26.41 | -29.38

Table 4.1: Eight Simulated data sets, with the distributions used to generate the interaction parameters, and the Serial-Other Log (base 10) Bayes Factors, rounded to two decimal places, comparing the Normal-framework Serial model to the other three models: General, Parallel, and Coactive. Positive values constitute evidence in favor of the Serial model and negative values constitute evidence in favor of the opposing “other” model. Log Bayes Factors comparing any other pair of models (within a given framework) can be derived by taking the difference between two log Bayes Factors, with positive values of this difference suggesting evidence in favor of the “other” model corresponding to the second log Bayes Factor and negative values in favor of the “other” model corresponding to the first log Bayes Factor.

rately identifies the generating model.

In the interest of saving both time and space, only Normal framework simulations are presented here.

### 4.2.4 Other Analytic Approaches

In addition to the above modelling approaches, two quality control analyses will be included: a Bernoulli test and a participant-wise interaction-only ANOVA.

The Bernoulli test will compare the total number of negative (or positive) MICs in each experiment to the amount that would occur by chance if each sign was equally likely will be run as a quality control step. Under this test, the null hypothesis is that either the serial or general models are true and, consequently, the division of MICs is roughly equal between positive and negative. The alternative hypothesis is that either the parallel model or the coactive model is true, with the number of negative (or positive) MICs corresponding to
the exact conclusion. For the purposes of this step, the error rate for the test will be set at .05, as is typical for psychological research.

In the ANOVA, each participant’s reaction time data will be analyzed under an ANOVA model with only an interaction effect and classified into serial, parallel, or coactive according to significance and the sign of the surface MIC (which is unlikely to differ substantially from the relevant ANOVA contrast). This approach is based on a similar categorization approach as seen in Fific et al. (2008), and only assumes the interaction effect as earlier attempts to include main effects as well as interactions almost always favored the serial model. Two error rates will be used in each experiment: a base .05 error rate for each participant and a Bonferroni-corrected error rate applied to each experiment as a whole. Although a paper-wide Bonferroni correction could also be used, this would likely bias the results in favor of the serial model and, consequently, is not used.

4.3 Changes from the Original Proposal

The stimuli mentioned above are not the stimuli that were originally planned for this research. The originally planned stimuli, nested squares and plaids, are not being studied because the former violates an essential assumption of the current methods and the latter appeared, in pilot testing, to be too complex to expect participants to perform at adequate accuracy.

The nested square stimuli used by Luck and Vogel (1997) and Wheeler and Treisman (2002) will not be analyzed in these experiments for two primary reasons. First, developmental work on these stimuli made it clear that the selection of changes is not a trivial task, due to the complexities of various color spaces. Second, the Systems Factorial methodology requires that changes in one stimulus does not affect the amount of time it takes to process the other stimulus, an assumption called selective influence, which I believe is likely to be violated by nested square stimuli because of contrast effects. An example of
such a contrast effect is depicted in Figure 4.5. In this figure, both inner squares are the same shade of blue, but the blue square on the right looks brighter because it is surrounded by purple whereas the blue square on the left looks darker because it is surrounded by yellow. This difference in perception is likely to influence reaction times and, perhaps more importantly, will reduce response accuracy enough to impact estimation of mean reaction times and, consequently, the MIC or otherwise make accuracy measures relatively meaningless. As such, I will avoid using the nested square stimuli for this research.

Plaids, as shown in the top right of Figure 4.1 and in Figure 4.6 were also considered. However, pilot research with research assistants suggests that they might be too difficult for most participants to be of any use. As such, they are no longer being considered for the current project, although they will be considered if this line of research continues. Other stimuli that were considered but not used included right triangles (seen in the bottom left of Figure 4.1) that would vary in the lengths of both legs. These stimuli were not included
Figure 4.6: Examples of the plaids stimuli and the magnitudes of change considered.

due to time constraints.

In the next sections, the experiments and their results will be described in more detail.
4.4 Experiment 1: Screwhead Perceptual Task

4.4.1 Participants

32 participants from the University of Missouri, 1 of which was an undergraduate research assistant and 31 of which were students from an introductory psychology class, were tested in this experiment.

4.4.2 Data Cleaning

An additional 16 participants were run but not included in the final analysis because of insufficient accuracy. Data was further cleaned by removing trials in which participant responses were faster than 300 ms, were slower than 5000 ms, or were incorrect. This cleaning accounted for the removal of 18% of the data in this experiment. The by-condition accuracies can be found in Figure A.1A in Appendix A.

4.4.3 Stimuli

Screwhead stimuli that varied in both the size of the circle (and, consequently, the length of the bisecting line) and in the orientation of the bisecting line were used in this experiment. The stimuli were presented concurrently at a quarter screen-width (about 360 pixels) to the left and right of center. The stimuli were allowed to vary in size between 54 pixels and 180 pixels in radius and in angle between -90 degrees (90 degrees counterclockwise from vertical) and 90 degrees (90 degrees clockwise from vertical). When the stimuli were different in a feature they differed by 15 or 30 percent of the original size and by 20 or 60 degrees in orientation for small and large changes, respectively.
4.4.4 Procedure

Participants were presented with a fixation cross for 500 ms, followed by two screwheads, one to either side of center, which remained on screen until the participant provided a response, as seen in Figure 4.2. This task was repeated a total of 378 times, with the first 18 trials used as practice trials that were not entered into the analysis.

4.4.5 Results

Participant-wise estimates for the interaction parameters estimated from both the normal and gamma frameworks can be seen plotted against ”surface” MIC estimates (that is, MIC estimates derived directly from the mean reaction times in each condition) in Figures 4.7A and 4.7B, respectively. These estimates are plotted against each other in Figure 4.7C, and the Maximum Absolute Autocorrelation factors for each model are plotted in Figure 4.7D. In all cases, black corresponds to the Serial model (and should always be 0 by definition), red to the General model, green to the Parallel model, and blue to the Coactive model. In the autocorrelation plots, the solid lines correspond to the Normal framework values and the dashed lines correspond to the Gamma framework values.

As can be seen, the General model parameter estimates from both frameworks roughly correspond to the surface estimates of MIC, albeit with fairly substantial shrinkage towards zero (that is, pull of parameter estimates towards the mean of the prior distribution due to over-informative or misspecified priors or due to an insufficient amount of data). Additionally, in the Normal framework, parameter estimates for the Parallel and Coactive models tended to approach the surface estimates for extreme estimates concordant with the respective model and approached zero for extreme estimated contradictory to the respective model. However, the Gamma framework estimates exhibit substantial shrinkage above and beyond what the Normal framework exhibits, suggesting that the priors may yet be too constraining or that the model may otherwise be misspecified. In either case, the maximum
Table 4.2: Number of participants classified as using Serial, Parallel, or Coactive processing according to the participant-wise ANOVA analysis, using both a raw .05 error rate and an experiment-wise .05 error rate determined by Bonferroni correction.

<table>
<thead>
<tr>
<th>Correction</th>
<th>Serial</th>
<th>Parallel</th>
<th>Coactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>11</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>20</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

absolute autocorrelations suggest that little, if any, thinning or adjustment beyond what was already applied to chain estimates needs to be made. Additional information regarding the MIC estimates and by-condition reaction times can be seen in Figures A.1B through D in Appendix A.

In addition to the above characteristics, the distribution of these statistics, whether the surface estimates or either of the model estimates, is even enough to suggest that neither the Parallel nor the Coactive model would be definitively supported over the Serial or General models. This is supported by a binomial test on the surface estimates, 19 of which were negative, (n=32, p=.38). The ANOVA classification also provides partial support for this, with most participants classified as using serial processing (see Table 4.2).

The Serial-Other log Bayes Factors from both the Normal and Gamma frameworks are printed in Table 4.3. Both frameworks suggest that the Serial model is the most appropriate explanation for the current data, relative to the other three models. Additionally, both weigh the remaining models similarly, with the General model being the second most likely outcome.

### 4.5 Experiment 2: Screwhead Memory Task

#### 4.5.1 Participants

30 participants from introductory psychology classes at the University of Missouri were tested in this experiment.
Figure 4.7: A. Surface-Normal plot for all four models. Serial model estimates (black) are included for completeness. General model estimates (red), Parallel model estimates (green) and Coactive model estimates (blue) all increase monotonically with the surface MIC estimates, albeit with substantial shrinkage. B. Surface-Gamma plot for all four models. C. Normal-Gamma plot for all four models. D. Maximum Absolute Autocorrelation plot across all individual interaction parameters within each model.
### Table 4.3: Serial-Other Log (base 10) Bayes Factors, rounded to two decimal places, comparing the Serial model to the other three models: General, Parallel, and Coactive. Positive values constitute evidence in favor of the Serial model and negative values constitute evidence in favor of the opposing model. Log Bayes Factors comparing any other pair of models (within a given framework) can be derived by taking the difference between two log Bayes Factors, with positive values of this difference suggesting evidence in favor of the model corresponding to the second log Bayes Factor and negative values in favor of model corresponding to the first log Bayes Factor.

<table>
<thead>
<tr>
<th>Framework</th>
<th>General</th>
<th>Parallel</th>
<th>Coactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>2.40</td>
<td>13.95</td>
<td>14.05</td>
</tr>
<tr>
<td>Gamma</td>
<td>6.37</td>
<td>16.20</td>
<td>18.84</td>
</tr>
</tbody>
</table>

#### 4.5.2 Data Cleaning

An additional 21 participants were not included in this analysis because of insufficiently accurate performance, much of which occurred due to the difficulty of the original version of the task. Data was further cleaned by removing trials in which participant responses were faster than 300 ms, were slower than 5000 ms, or were incorrect. This cleaning accounted for the removal of 24% of the data in this experiment. The by-condition accuracies can be found in Figure A.2A in Appendix A.

#### 4.5.3 Stimuli

The stimuli in experiment 2 are the same as those in experiment 1 except the background was grey instead of black. This change in background color was done to reduce the occurrence of after-images that would impact the validity of the memory task. Additionally, the study and test stimuli were located in the same spatial location instead of being in distinct spatial locations. This was done to reduce the need for saccades and to give participants the best chance to perform accurately. The magnitudes of the changes were also changed to 30% and 50% of the original size and 35 or 70 degrees for small and large changes, respectively, due to insufficient performance for the original 21 participants that received...
the stimuli with the original change magnitudes.

4.5.4 Procedure

Participants were presented with a fixation cross for 500 ms, followed by the presentation of one screwhead at center for 200 ms. After the study period, the screen went blank for 2000 ms before the second screwhead was presented at center until the participant provided a response. This trial structure is largely similar to that shown in Figure 4.3.

4.5.5 Results

Participant-wise estimates for the interaction parameters estimated from both the normal and gamma frameworks can be seen in Figure 4.8. These plots represent the same types of information as seen in Figure 4.7.

Like in Experiment 1, despite some substantial shrinkage, the estimates of the interaction parameter from the Normal framework closely approximate the surface MICs. Additionally, the parameter estimates from the Gamma framework still exhibit some substantial shrinkage, but also suggest that neither the serial and general models are the two most likely models. Autocorrelation estimates are also tolerable. Additional information regarding the MIC estimates and by-condition reaction times can be seen in Figures A.2B through D in Appendix A.

The General model estimates in both the Normal and Gamma frameworks are fairly distributed across the range with neither positive nor negative values occurring more frequently, favoring serial and general interpretations. This evidence is further supported by both the bernoulli test, with 15 of the MICs being negative (n=30, p=1), and the ANOVA classification (as seen in Table 4.4).

The Serial-Other log Bayes Factors from both the Normal and Gamma frameworks are printed in Table 4.5. Like in Experiment 1, both the Normal and Gamma frameworks
Figure 4.8: A. Surface-Normal plot for all four models. Serial model estimates (black) are included for completeness. General model estimates (red), Parallel model estimates (green) and Coactive model estimates (blue) all increase monotonically with the surface MIC estimates, albeit with substantial shrinkage. B. Surface-Gamma plot for all four models. C. Normal-Gamma plot for all four models. D. Maximum Absolute Autocorrelation plot across all individual interaction parameters within each model.
<table>
<thead>
<tr>
<th>Correction</th>
<th>Serial</th>
<th>Parallel</th>
<th>Coactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>22</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>28</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.4: Number of participants classified as using Serial, Parallel, or Coactive processing according to the participant-wise ANOVA analysis, using both a raw .05 error rate and an experiment-wise .05 error rate determined by Bonferroni correction.

<table>
<thead>
<tr>
<th>Framework</th>
<th>General</th>
<th>Parallel</th>
<th>Coactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>2.11</td>
<td>13.34</td>
<td>16.38</td>
</tr>
<tr>
<td>Gamma</td>
<td>7.42</td>
<td>15.95</td>
<td>16.86</td>
</tr>
</tbody>
</table>

Table 4.5: Serial-Other Log (base 10) Bayes Factors, rounded to two decimal places, comparing the Serial model to the other three models: General, Parallel, and Coactive. Positive values constitute evidence in favor of the Serial model and negative values constitute evidence in favor of the opposing model. Log Bayes Factors comparing any other pair of models (within a given framework) can be derived by taking the difference between two log Bayes Factors, with positive values of this difference suggesting evidence in favor of the model corresponding to the second log Bayes Factor and negative values in favor of model corresponding to the first log Bayes Factor.

support the Serial model over the other three models, with the General model being the next most likely. The models generate more evidence in favor of the Parallel model relative to the Coactive model, despite there being exactly 15 negative MICs and 15 positive MICs. This is likely a consequence of the few participants with extremely negative $(-.3)$ MICs and the lack of participants with similarly extreme positive MICs.

### 4.6 Experiment 3: 2-Digit Memory Task

#### 4.6.1 Participants

28 participants from introductory psychology classes at the University of Missouri were tested in this experiment.
4.6.2 Data Cleaning

Data was cleaned by removing trials in which participant responses were faster than 300 ms, were slower than 5000 ms, or were incorrect. This cleaning accounted for the removal of 9% of the data in this experiment. The by-condition accuracies can be found in Figure A.3A in Appendix A.

4.6.3 Stimuli

The stimuli used were pairs of digits written in white and displayed against a grey background. The pairs were displayed separately at a quarter-screen width to the left and right of center. The digits in the study pair were randomly selected from the set 4, 5, 6 with the digits in the test pair differing by either 1 or 3, in either direction, for small and large changes, respectively.

4.6.4 Procedure

The procedure for Experiment 3 is largely similar to the procedure for Experiment 2. The only substantial difference, outside of the different stimulus, is that study and test stimuli were moved to the left and right sides of center, respectively, as in Experiment 1 and in Figure 4.3, instead of being presented at center.

4.6.5 Results

Participant-wise estimates for the interaction parameters estimated from both the normal and gamma frameworks can be seen in Figure 4.9. These plots represent the same types of information as seen in Figure 4.7.

As in the previous experiments, the Gamma framework experiences substantially more shrinkage than the Normal framework but both frameworks exhibit tolerable autocorrela-
Table 4.6: Number of participants classified as using Serial, Parallel, or Coactive processing according to the participant-wise ANOVA analysis, using both a raw .05 error rate and an experiment-wise .05 error rate determined by Bonferroni correction.

<table>
<thead>
<tr>
<th>Correction</th>
<th>Serial</th>
<th>Parallel</th>
<th>Coactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>11</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>16</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.7: Serial-Other Log (base 10) Bayes Factors, rounded to two decimal places, comparing the Serial model to the other three models: General, Parallel, and Coactive. Positive values constitute evidence in favor of the Serial model and negative values constitute evidence in favor of the opposing model. Log Bayes Factors comparing any other pair of models (within a given framework) can be derived by taking the difference between two log Bayes Factors, with positive values of this difference suggesting evidence in favor of the model corresponding to the second log Bayes Factor and negative values in favor of model corresponding to the first log Bayes Factor.

<table>
<thead>
<tr>
<th>Framework</th>
<th>General</th>
<th>Parallel</th>
<th>Coactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1.68</td>
<td>7.35</td>
<td>19.02</td>
</tr>
<tr>
<td>Gamma</td>
<td>7.67</td>
<td>12.36</td>
<td>18.44</td>
</tr>
</tbody>
</table>

There is a larger proportion of negative MICs (21) than positive MICs (7), which suggests that these stimuli may be processed in parallel by most participants ($n=28$, $p=.01$). Several of the observed MICs are fairly close to 0, which is a known shortcoming of this particular test, but the ANOVA classification, seen in Table 4.6, suggests that the processing for these stimuli is fairly evenly split between serial and parallel processing.

The Serial-Other log Bayes Factors from both the Normal and Gamma frameworks are printed in Table 4.7. The Bayes Factors here again support the Serial model over all other models, although it is important to note that the evidence in favor of the serial model over the general model is rather slight, as suggested by both the bernoulli test and the ANOVA classification. Additionally, the above analyses and the current analysis all suggest that these stimuli were generally not processed coactively.
Figure 4.9: A. Surface-Normal plot for all four models. Serial model estimates (black) are included for completeness. General model estimates (red), Parallel model estimates (green) and Coactive model estimates (blue) all increase monotonically with the surface MIC estimates, albeit with substantial shrinkage. B. Surface-Gamma plot for all four models. C. Normal-Gamma plot for all four models. D. Maximum Absolute Autocorrelation plot across all individual interaction parameters within each model.
4.7 Experiment 4: 2-Digit Perception Task

4.7.1 Participants

28 participants from introductory psychology classes at the University of Missouri were tested in this experiment.

4.7.2 Data Cleaning

Data was cleaned by removing trials in which participant responses were faster than 300 ms, were slower than 5000 ms, or were incorrect. This cleaning accounted for the removal of 9% of the data in this experiment. The by-condition accuracies can be found in Figure A.4A in Appendix A.

4.7.3 Stimuli

The stimuli used in this experiment are identical to those used in Experiment 3.

4.7.4 Procedure

The trial structure for this experiment is identical to that for Experiment 1, with the stimuli being the primary difference.

4.7.5 Results

Participant-wise estimates for the interaction parameters estimated from both the normal and gamma frameworks can be seen in Figure 4.10, which is similar in structure to Figure 4.7. The Gamma framework estimates are still over-shrunk but, otherwise, useable, the Normal framework estimates track the surface estimates fairly well, and autocorrelation in
Table 4.8: Number of participants classified as using Serial, Parallel, or Coactive processing according to the participant-wise ANOVA analysis, using both a raw .05 error rate and an experiment-wise .05 error rate determined by Bonferroni correction.

<table>
<thead>
<tr>
<th>Correction</th>
<th>Serial</th>
<th>Parallel</th>
<th>Coactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>8</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Bonferroni</td>
<td>19</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4.9: Serial-Other Log (base 10) Bayes Factors, rounded to two decimal places, comparing the Serial model to the other three models: General, Parallel, and Coactive. Positive values constitute evidence in favor of the Serial model and negative values constitute evidence in favor of the opposing model. Log Bayes Factors comparing any other pair of models (within a given framework) can be derived by taking the difference between two log Bayes Factors, with positive values of this difference suggesting evidence in favor of the model corresponding to the second log Bayes Factor and negative values in favor of model corresponding to the first log Bayes Factor.

<table>
<thead>
<tr>
<th>Framework</th>
<th>General</th>
<th>Parallel</th>
<th>Coactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1.33</td>
<td>9.08</td>
<td>18.40</td>
</tr>
<tr>
<td>Gamma</td>
<td>7.66</td>
<td>11.94</td>
<td>17.63</td>
</tr>
</tbody>
</table>

the interaction parameter estimates is generally acceptable. Additional information regarding the MIC estimates and by-condition reaction times can be seen in Figures A.4B through D in Appendix A.

Like in experiments 1 and 2, the General model estimates are fairly evenly distributed above and below 0, with 18 estimates below 0 (n=28, p=.18) despite the two extreme scores, suggesting that the Serial and General models are more likely to be correct. This is somewhat supported by the ANOVA classification (see Table 4.8), although the classification with no correction factor identifies nearly half of all participants as using parallel processing.

The Serial-Other log Bayes Factors from both the Normal and Gamma frameworks as applied to this experiment are printed in Table 4.9. In both frameworks, evidence supports the Serial model over all other models, with the evidence against coactive processing being somewhat stronger than evidence against parallel processing.
Figure 4.10: A. Surface-Normal plot for all four models. Serial model estimates (black) are included for completeness. General model estimates (red), Parallel model estimates (green) and Coactive model estimates (blue) all increase monotonically with the surface MIC estimates, albeit with substantial shrinkage. B. Surface-Gamma plot for all four models. C. Normal-Gamma plot for all four models. D. Maximum Absolute Autocorrelation plot across all individual interaction parameters within each model.
Chapter 5

Discussion and Conclusion

The results suggest that both the Screwhead stimuli and the Digit Pair stimuli are processed serially in both the perception experiment and the memory experiment, with the forced-funneling signature being absent. Consequently, there is no evidence that either type of stimuli is chunked in Working Memory, because the lack of coactive processing in the memory tasks suggests that the features are maintained separately and, thus, no chunking occurs. Consequently, I did not succeed in my original goal.

Additionally, although the serial processing of Screwhead stimuli is in line with past research, the same cannot be said with the Digit Pair stimuli, which I expected would have been processed in parallel in the perception experiment. Although no hypothesis was presented for the memory experiments, the fact that the processing architecture detected for the memory experiments did not differ from either the perceptual experiments or from each other makes me wonder if the experimental method was not well-specified.

The conclusions derived from the quality control steps largely agree with conclusions derived from the above tests, with some qualifications. Although the Bernoulli test on Experiment 3 suggests that most participants processed the Digit Pair stimuli in parallel in that task, the test itself does not take the magnitude of each effect into account and, thus, can be fooled by a preponderance of statistics that are close to, but not equal to, zero. This
is likely the case for this experiment, as Figure 4.9A shows a fairly substantial clump of MICs that are just below 0, and the ANOVA classification (Table 4.6) also showed a fairly even division between serial and parallel processing, even though this was the most extreme distribution of classifications out of the four experiments.

5.1 Shortcomings

Although I am fairly confident in the initial results of this research, there are a few shortcomings that I do need to acknowledge. These shortcomings come from three main sources: analytic assumptions, experimental assumptions, and inefficient data usage.

5.1.1 Analytic Assumptions

The analyses that I have used make several assumptions that might not hold, which could either account for misspecification and error in the model or provide alternative interpretations that I was unable to provide with the given methods. There are three main assumptions to consider: selective influence in stimuli, a lack of cross-talk in processing, and appropriateness of the chosen analytic method.

First, I assume that Selective Influence, that a change in one stimulus feature does not change the amount of processing time needed for the other feature, holds for the stimuli used in these experiments. This was done primarily because these experiments do not collect enough data to allow for a test of this assumption, but this means that any evidence against selective influence in these feature pairs would invalidate these conclusions. This is especially worrisome with the digit pair stimuli, as feature confusion, and subsequent cross-feature influence, is likely to occur.

Second, I also assume that there is no cross-talk, or some other influence of one feature in the processing of the other feature, in these stimuli as that is a central assumption to SFT.
Unfortunately, this assumption might not be warranted, especially considering the amount of neuronal interconnection and signal looping that is almost certainly involved in memory. This is especially important to consider, given that the processing architectures considered by Systems Factorial Technology are not exhaustive and can be mimicked by other architectures. Specifically, simulations by Tremblay et al. (2014) illustrate that exhaustive serial processing, the type of processing found in all four experiments in this project, is also the conclusion SFT reaches when analyzing reaction times produced by iterative recall neural networks, such as the Bidirectional Associative Memory models used in the simulation. This may be due to the fact that SFT was not designed with these alternative processing structures in mind, but this overlap should be considered for future research.

Last, although I have some confidence in this methods ability to determine if a single processing architecture is used for a given stimulus, there is both room for improvement and room for additional modelling approaches. The biggest issue to work with is the misspecification of the Gamma framework. The current analysis forced the shape parameter to be constant due to an issue in development in which the estimates for the shape parameter continued to increase until the range of possible estimates for the other (scale) parameters shrunk to the null set. One possible reason for this issue is that the Gamma distribution lacks a built-in shift parameter, leaving the shape and scale parameters to determine the mean and variance of the distribution. Unfortunately, the priors needed for accurate model assessment restricted the scale parameters to such a small value that the analysis consistently favored larger shape parameters to compensate. To account for this, adjusting the Gamma framework to include a shift parameter or refining and reparameterizing the failed ExGaussian or Log-Normal frameworks will be pursued.

Additionally, the favoring of the serial model over the other models may have been due to the inherent simplicity of the serial model as each other model had about 30 additional free parameters that were not present in the serial model. This additional dimensionality
may have provided a penalty substantial enough to eliminate any evidence in favor of the other models. However, this may be remedied by re-running the previous analyses with the priors on the interaction parameter’s hierarchical variance set such that said variance was as close to zero as possible, forcing all participants to have roughly the same interaction parameter. This adjustment, along with further attempts to check for misspecification, such as calculating Divergence Information Criterion or other badness of fit measures, or attempts at alternate model descriptions, such as a Multinomial Classification model, are being considered.

5.1.2 Experimental Assumptions

In addition to the above analytic assumptions, there are also a few methodological assumptions that I have made that need to be considered. These assumptions suggested that the task did not favor specific types of responses, either because of response type or because of allowed processing time. In this regard, there are three assumptions I will cover here: that the experiments do not have task demands, that there is no need for additional limits to response time, and that participant performed perfectly according to exhaustive processing.

First, I assumed that there would be no Task Demands in this experiment. That is, I did not expect that the method of the memory task might inherently bias participants by forcing them to potentially “break apart” any chunks that might have been formed to make the comparison. Specifically, by calling attention to the individual features, we may have forced them to use serial processing deliberately and removed any consideration of the gestalt stimuli. Consequently, an alternative trial format for the memory tasks, specifically one that forces the comparison to be remembered, must be considered alongside the current form of the experiment to attempt to preserve any gestalt perception and, thus, ensure that this method is still a viable means of studying chunking.

Second, I assumed that allowing participants to take as much time as they need to re-
spond would not affect the results. Although few trials were removed due to the amount of time needed to make the decision, the relative lack of urgency in responses might have encouraged less efficient processing of the stimuli and, consequently, biased the participants towards serial processing from the beginning. Future research should consider that processing architectures might depend on whether the trials were speeded or unspeeded.

Last, I assumed that participants always performed the task perfectly when they responded correctly. Specifically, I have not taken into account the possibility that some participants might have been guessing on some trials. Although the data cleaning performed prior to analysis was intended to account for trials on which response due to memory were unlikely, it might not have captured all trials in which a participant was guessing instead of responding faithfully to the trial. As such, future research should consider modelling or otherwise studying participant guessing behaviors.

5.1.3 Inefficient Data Usage

One other limitation in the current research is the relatively inefficient use of data, despite applying Bayesian methodology in the analysis. Specifically, data cleaning performed before analysis eliminated two types of information that may have been useable in other, related applications: incorrect responses and "no-change" responses to stimuli where at least one feature remained unchanged.

First, a substantial proportion of the data had been removed due to incorrect responses, especially in Experiments 1 and 2. Although this was considered a necessary step for initial research, doing so results in unbalanced designs, which complicates the analysis, and artificially decreases the precision of estimates of the interaction parameter. There are two ways to resolve this issue: artificially increase the experiment duration by dynamically assigning trials until the participant has produced a set number of correct responses to each of the four relevant conditions or use missing data imputation to determine what the reaction time
would have been if they did provide a correct response. Additionally, tradeoffs between accuracy and the observed interaction parameter should be considered, if only because certain processing architectures might motivate participants to make quicker, less accurate, responses regardless of any speeding in the experiments.

Second, even if the above methods were considered at least half of the data in the experiments, that from responses to stimuli in which there was at least one stimulus that did not change, is inherently discarded and left unused. Although not as throughly researched as the main interaction statistics, Systems Factorial Technology does admit the use of reaction time data from these no-change trials in an additional diagnostic (Fific, Little, & Nosofsky, 2010), and future research should consider its usage to allow for more thorough, and more efficient, analyses.

5.2 Future Research and Extensions

Although the above issues need to take priority, there are several possible extensions of this methodology that I would like to consider in future research outside of the obvious application to other arbitrary visual feature pairs. I cannot hope to list all possible extensions, as there are undoubtedly applications of this method which I have not thought of as of this writing, but there are four general extensions that I would consider: higher-dimensional stimuli, changes in architecture over time, out-of-the-lab variations, and generalization to non-visual features.

First, I would like to see if there is a way to generalize this method to study the combination of three or more features, if only for theoretical interest. At first, it seems clear that such attempts would likely never extend beyond this methodology as both the complexity of the relevant interaction parameters and the substantial increase in trial counts and, thus, experiment length, resulting from combinatorial explosion would prevent this. Indeed, I believe that generalization beyond combinations of 3 features would likely be very difficult
to attain. However, this could potentially be mitigated by nesting the combinations; that is, by determining the processing architecture of an arbitrarily-selected pair of features and, if that pair is found to be processed coactively, running a second test using the first feature pair as one “feature” and seeing if it and the third feature are then processed coactively as well.

Second, I would like to see if the processing architecture used is necessarily constant with respect to time and, in the event that it is not, I would like to determine what types of transitions in the architecture used tend to occur. This is especially important because of the implications this could have for skill-learning and automaticity research. It is also important to note that this extension is not necessarily new, as similar analyses have been run in prior SFT research (Fific et al., 2008) and, even though the current research does not produce as substantial a data set as prior SFT research, I do have to wonder if a simple by-block analysis could be run with my data. Additionally, I have to wonder if calculating a running interaction statistic would also provide insight into this, or other, questions. This, however, would likely require a data set substantially larger than the ones I currently have to be of any use.

Third, I would like to see this method applied to experimental protocols beyond looking at a two-stimulus feature and pressing one of two buttons, as the current implementation seems limited in its generalizability and, thus, its empirical validity, because it is almost always applied in a relatively artificial setting. As such, I propose two design options to consider for future research: using responses other than button presses, to see if the response modality impacts the processing architecture employed, and applying the method to similar decisions being made outside of laboratory conditions. The first type of change seems fairly straightforward, such as using verbal or saccade responses, and will not be elaborated on further for this reason. As for the second idea, one potential idea would be to use a shooting-gallery-type task, in which a participant is given a light-gun, like those
used in various arcade games, and is shown a series of two-feature stimuli and told that a specific combination is a "bad" stimulus and must be shot. The exact details regarding what identifies the "bad" stimulus will be determined by any researchers who pursue this method, suffice to say that this would provide one instance of a less artificial task to which this analytic and experimental method can be applied.

Last, Although this method has largely been considered with respect to visual stimuli, I do have to wonder if it is possible to extend it to stimuli with non-visual features. Stimuli that consist of one visual feature and one auditory feature immediately come to mind, as both are amenable to experimental manipulation and stimuli of both types are identified and resolved fairly quickly, but extending this method to tactile, olfactory, or gustatory stimuli or, indeed, any other type of stimulus that humans can experience, seems possible at first glance and, hypothetically, could answer many questions about how certain combinations of stimuli are perceived or, otherwise, extend our understanding of their perception.

Overall, although I did not succeed in identifying chunking, I still believe that this method could be employed in many avenues of future research to determine how stimuli are, or are not, combined by human perception. And, in doing so, we could develop a better understanding of what people tend to see and, consequently, what they were likely thinking leading up to a given action.
Appendix A

Supplemental Figures
Figure A.1: A. Marginal mean accuracies (pre-cleaning) by condition in Experiment 1. The change in angle is indicated by the x-axis while the change in size is indicated by the color of the points, with blue corresponding to no change, green corresponding to a small change, and red corresponding to a large change. The four points corresponding to the double-change trials are further highlighted with black borders. B. Surface MIC estimates from Experiment 1, sorted by magnitude, with error bars that correspond to the 80% confidence intervals of the means. The points and error bars are colored with regard to sign: green for negative, red for positive, and blue for intervals containing zero. C. Marginal mean reaction times (post-cleaning) by condition in Experiment 1. The arrangement, coloration, and highlighting are as in part A. D. Posterior means of MIC estimates from the General model in the Normal framework for Experiment 1, sorted by the magnitude of the surface MIC estimates in part B. Error bars correspond to the 90% posterior credible interval.
Figure A.2: A. Marginal mean accuracies (pre-cleaning) by condition in Experiment 2. The change in angle is indicated by the x-axis while the change in size is indicated by the color of the points, with blue corresponding to no change, green corresponding to a small change, and red corresponding to a large change. The four points corresponding to the double-change trials are further highlighted with black borders. B. Surface MIC estimates from Experiment 2, sorted by magnitude, with error bars that correspond to the 80% confidence intervals of the means. The points and error bars are colored with regard to sign: green for negative, red for positive, and blue for intervals containing zero. C. Marginal mean reaction times (post-cleaning) by condition in Experiment 2. The arrangement, coloration, and highlighting are as in part A. D. Posterior means of MIC estimates from the General model in the Normal framework for Experiment 2, sorted by the magnitude of the surface MIC estimates in part B. Error bars correspond to the 90% posterior credible interval.
Figure A.3:  
A. Marginal mean accuracies (pre-cleaning) by condition in Experiment 3. The change in the second digit is indicated by the x-axis while the change in the first digit is indicated by the color of the points, with blue corresponding to no change, green corresponding to a small change, and red corresponding to a large change. The four points corresponding to the double-change trials are further highlighted with black borders.  
B. Surface MIC estimates from Experiment 3, sorted by magnitude, with error bars that correspond to the 80% confidence intervals of the means. The points and error bars are colored with regard to sign: green for negative, red for positive, and blue for intervals containing zero.  
C. Marginal mean reaction times (post-cleaning) by condition in Experiment 3. The arrangement, coloration, and highlightings are as in part A.  
D. Posterior means of MIC estimates from the General model in the Normal framework for Experiment 3, sorted by the magnitude of the surface MIC estimates in part B. Error bars correspond to the 90% posterior credible interval.
Figure A.4: A. Marginal mean accuracies (pre-cleaning) by condition in Experiment 4. The change in the second digit is indicated by the x-axis while the change in the first digit is indicated by the color of the points, with blue corresponding to no change, green corresponding to a small change, and red corresponding to a large change. The four points corresponding to the double-change trials are further highlighted with black borders. B. Surface MIC estimates from Experiment 4, sorted by magnitude, with error bars that correspond to the 80% confidence intervals of the means. The points and error bars are colored with regard to sign: green for negative, red for positive, and blue for intervals containing zero. C. Marginal mean reaction times (post-cleaning) by condition in Experiment 4. The arrangement, coloration, and highlighting are as in part A. D. Posterior means of MIC estimates from the General model in the Normal framework for Experiment 4, sorted by the magnitude of the surface MIC estimates in part B. Error bars correspond to the 90% posterior credible interval.
References


Cheng-Ta Yang. Relative saliency in change signals affects perceptual comparison and

I was born on October 4th, 1986 at Heartland West Hospital in Saint Joseph, Missouri. I was born to Ivan Thiele, an engineer for the Northwest Missouri section of the Natural Resource Conservation Services, and Debra Thiele, a Book-Keeper and eventual Human Resources employee for H&R Block. I was the youngest of three children with two older sisters: Sheila, who is 8 years older, and Genevieve (or Jenny), who is 4 years older. I was raised in a home just east of the Saint Joseph border and lived there until I finished my Bachelor’s degree. I was always a rather odd child, and didn’t speak until I was three years old, although I had learned to read before then. I also enjoyed mathematics and science to the point that I looked to “Donald Duck in Mathemagic Land” and “Bill Nye the Science Guy” almost religiously, and loved to play with simple constructions of shape blocks. I generally viewed mathematics and science as toys and things to play with and understand, even if my understanding of understanding was rather unrefined.

For school, I went to Bessie Ellison Elementary, with a once-per-week visit to Skaith Elementary for the “Rainbows” program for gifted students from third through sixth grades. This was followed by two years in Bode Middle School, with participation in the “Odyssey” program that followed from “Rainbows”, and the end of my compulsory education at Central High School, where I spent much of my time in honors-level courses and, consequently, rarely saw most of the students in my grade level, focusing mostly on performing well in my classwork. And I performed very well, participating in several mathematics competitions from fourth grade on, being part of other academic clubs and activities, and even participating in other special academic programs with sessions at the local college. I also participated in academic activities every summer, from special summer-school classes in elementary school to marching band practice in high school. I especially excelled at exams, scoring a respectable 20 on the math ACT section in eighth grade and eventually earing a 30 overall during high school.
Although academic interests were of primary concern to me, I did develop interests in other topics during middle school and high school. It was around this time that I began to develop interests in video games, Japanese animation (or Anime), Rampant Daydreaming, and Internet “citizenship” at this time because of the relative lack of direct social interactions (and abundance of simulated social interactions) involved in them, and all of which continue to be big parts of my life to this day. I also enjoyed comedy, especially that from Dave Barry, George Carlin, Douglas Adams, and Monty Python, which all undoubtedly shaped my perspective on human behaviors.

It was also during this time that I started to wonder what it was about me that separated me from everyone else, aside from what seemed to be a general overperformance in math and science and similar underperformance in history and social studies. Although I did have a few friends here and there, I was largely isolated and generally avoided initiating anything beyond simple, fleeting conversations about largely trivial points, because I mostly didn’t relate to anyone else. I was also somewhat routinely teased and felt as though the people in charge of the school weren’t interested in fixing the issue, which I took to mean that I was the problem, even though I was still confused by this and most other behaviors people exhibited.

Towards the end of my time in high school, I had some extra class times to fill, as I had focused enough on my studies to meet all of the requirements for graduation. Because I happened to have the appropriate time vacant and saw little interest in the other offerings, I ended up taking a course in Psychology. This course completely changed the way I thought about people, taking them from some nebulous cloud of behaviors that could never be understood properly and which changed the way they responded to the same input in ways that never made sense and describing them *mechanically*, as a system that could be studied and understood by means of standard scientific principles. Because of my confusion and my performance in science coursework, this perspective greatly appealed to me. As such, I
decided to major in Psychology in college.

I went to college at Missouri Western State University in Saint Joseph. This decision was made primarily because I had enough scholarship money between my ACT scores and what the college itself offered that I could effectively study there for free, minus the costs of food and living, both of which were remedied by living at home with my father and riding in with him every day (and riding back home every night). My mother had started living in Kansas City at this time to avoid commuting an hour in each direction every day due to work, and both of my sisters had since started living on their own, so it was just my father and I. And, although we talked here and there about a few topics, I mostly didn’t relate because our expertises had started to diverge, much like how our interests had diverged once I entered middle school and continued to do so throughout high school and college, as I made a few friends who were also interested in video games and anime, and even helped the Anime Club become an official club that was recognized by the university.

It was also during this time that I started to become a little disillusioned with my initial views of Psychology, the amount of scientific rigor and mathematical aptitude involved in it, and the worth of a Bachelor’s degree in it, but I felt that I had crossed the Rubicon in deciding my major and finished the degree anyway. But this change in view, and the perceived expectation that I would keep going, led me to apply to a graduate program in Psychology. However, much like my initial reckless decision to major in a field based on a class taken on a whim, my decisions on which program to apply to were also rather sudden. Each program I considered (of which there were five) was considered mainly because a couple of the papers authored by the lab leader was interesting and because a guide on which schools were accepting new students listed (erroneously in one case) those professors as seeking new students. I was rejected in all but one application: my application to Jeff Rouder’s lab at the University of Missouri. It being the only program that accepted me, and it being in state, I decided to study there.
Which brings me to my current state. I have been here at the University of Missouri for almost six years now and, although I do not know how useful I have been or how useful I will be in the future, I can say that I have learned a lot that I did not know before now. Even though I had no real plan of study during my time here, I think I’ve gotten a better answer for what humans are than I did back in high school, and I think I have a better understanding of myself. However, I do still have a few regrets. Namely, I regret that I did not do more than I did regarding research, especially reading and writing.

Reading and writing have always been issues for me. They always felt like wastes of time compared to other pursuits because I generally did not understand them, either because the material was inherently uninteresting to me, because I understood the material but kept being told that I was “missing” something that the other party could never satisfactorily describe for me, or because I tried to communicate something that was either insufficient or lost on the other party. Even though the most enjoyable classes I ever took in high school were tenth and twelfth grade English (which focused on writing and reading, respectively), I never really enjoyed either activity, and this has continued to impact me ever since. As such, I never felt like I understood the essential material as well as the people around me did, nor did I feel I had the ability to talk to, or ask, them about it. And, as such, I felt that I shouldn’t try, and I didn’t produce as much as I could have because of this.

But, I’ve recently learned that I have done more than I think I have and, even with this perceived lack of ability, I can still contribute something. I do still have some time and ability to understand something and to relate it to other people, and I have a duty to do so. There is a good chance that this document will be the last academic document I ever produce, but that is okay. As long as I can put my skills to good use and help the people around me, especially those who will come after me, in some other way, then all will be well. I still regret that I have not communicated this well until now, especially since my mother will never see me at this level of performance due to her passing three years ago,
but I must not let that regret hold me back.

I am, essentially, a person who has learned a few things about math, science, and human behavior, who enjoys silly things on the internet, computer games, and Japanese animation, and who will likely continue to keep to themselves for the most part. And so long as I help society continue to move forward, that is all irrelevant and, thus, that is all okay.

“A lifetime spent solving the puzzles in this place would be one which missed everything of value in it; but we must remember that others will tread this path, and we must leave it better signposted than when we found it.”

—@ v. 17.1.0005 (excerpt from *The Talos Principle*)