January 2014

Bayesian Models of Sequential Dependencies in Binary and Multi-Interval Response Tasks

Jeffrey Scott Annis
University of South Florida, jannis@mail.usf.edu

Follow this and additional works at: http://scholarcommons.usf.edu/etd

Part of the Psychology Commons

Bayesian Models of Sequential Dependencies in Binary and Multi-Interval Response Tasks

by

Jeffrey Annis

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Kenneth Malmberg, Ph.D.
Amy Criss, Ph.D.
Emanuel Donchin, Ph.D.
Chad Dubé, Ph.D.
Sudeep Sarkar, Ph.D.
Toru Shimizu, Ph.D.

Date of Approval:
July 9, 2014

Keywords: measurement models, memory models, recognition memory, assimilation

Table of Contents

List of Tables

List of Figures

Abstract

A sequential dependency occurs when the response on the current trial is correlated with responses made on prior trials. Sequential dependencies have been observed in a variety of both perception and memory tasks. Thus, sequential dependencies provide a platform for relating these two cognitive processes. However, there are many issues associated with measuring sequential dependencies and therefore it is necessary to develop measurement models that directly address them. Here, several measurement models of sequential dependencies for both binary and multi-interval response tasks are described. The efficacy of the models is verified by applying them to simulated data sets with known properties. Lastly, the models are then applied to real-world data sets which test the critical assumption that the underlying processes of sequential dependencies are modulated by attention. The models reveal increased vigilance during testing decreases the degree of sequential dependencies.

Chapter 1

An Overview of Sequential Dependencies in Perception and Recognition

Suppose you are the witness to a crime and the police subsequently ask you to identify the perpetrator in a photographic lineup. When you arrive at the police station, you are presented with pictures of non-suspects, called "fillers", along with the actual suspect. Your task is to identify the suspect as the perpetrator and reject the fillers. There is an extensive body of research on how eyewitness lineups should be designed and performed. Steblay, Dysart, and Wells (2011) suggest that photographic lineups should be presented sequentially in contrast to being presented all at once. Imagine that you are given one of these sequential lineups and on the first trial you are presented with a filler person. You correctly reject the person as not being the perpetrator. On the next trial, the perpetrator's mug-shot is presented, however, you mistakenly reject the person depicted in the photograph as not being the one who committed the crime. Why did you make this mistake? One plausible reason is that you simply did not recognize the person in the photograph as being the perpetrator. A less obvious reason is that your previous response influenced your current response; because you said "No" on the first trial, you were more likely to say "No" on the subsequent trial. This is an example of a *sequential dependency*.

Sequential dependencies have been studied in a wide variety of perception including absolute identification (Stewart, Brown, & Chater 2005) categorization (Jones, Love, & Maddox, 2006), and perceptual detection (Howarth & Bulmer, 1956). Several models of sequential dependencies have been proposed in the perception literature (e.g. Treisman & Williams, 1984; Stewart, Brown, & Chater, 2005; Petrov & Anderson, 2005; Brown, Marley, Donkin, &

1

Heathcote, 2008). There has been considerable debate regarding the locus of sequential

dependencies. For example, Treisman and Williams (1984) proposed sequential dependencies

were the result of non-random changes in response bias across test trials. On the other hand,

Annis and Malmberg (2013) proposed a model of sequential dependencies in recognition

memory that assumes stimulus information from the previous trial carries over to next trial.

Although these models may propose different mechanisms to produce sequential dependencies, a

commonality of these models is that they all contain a memory component in which some type of

information about the previous trials is stored. This is not too surprising if one considers the long

historical relationship between models of memory and models of perception that can be traced

back as far as Miller's (1956) famous paper on capacity limitations. For example, Signal

Detection Theory (Green & Swets, 1966) was originally designed to distinguish between

changes in bias vs. changes in discriminability in perception tasks, but was later applied to

recognition memory tasks. This relationship is not limited to decisional models like Signal

Detection Theory (SDT). Indeed, at the process level, the Generalized Context Model (Nosofsky,

1986) was developed in order to simultaneously account for perceptual categorization and

recognition memory tasks.

1.1.   Sequential Dependencies in Perceptual Detection

One of the first perception tasks in which sequential dependencies were studied was in a

perceptual detection task (Howarth & Bulmer, 1956). In this task, the participant is presented

with tones or light that vary in intensity. The task of the participant is to respond "Yes" when

they detect the signal and "No" otherwise. Howarth and Bulmer found the probability of a "No"

response increases when the previous response was a "No" response vs. if the previous response

was a "Yes" response (Figure 1.1). This positive correlation between the current response and

previous response is known as *assimilation* or a *positive sequential dependency*. Assimilation has

also been observed in other perception tasks including absolute identification (Stewart, Brown, &

Chater 2005) and categorization (Jones, Love, & Maddox, 2006).
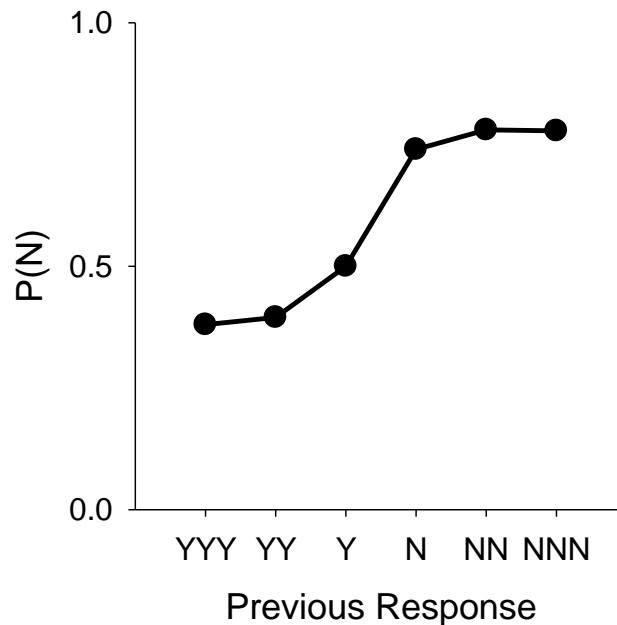
Figure 1.1. Results of Howarth and Bulmer (1956). There is a positive correlation between the
previous response and the current response.

    A task that is very similar to perceptual detection, but is a memory task, is a *yes/no*

recognition memory task. In this task, the participant is usually presented with a list of to-be-

remembered words. Following the study list, a test list is presented that is composed of words

that were studied and words that were unstudied. The task of the participant is to respond "Yes"

to studied words, and "No" to unstudied words. These stimulus-response combinations are

known as *hits* and *misses*, respectively. Studied words are sometimes referred to as *targets*, while

unstudied words are referred to as *foils*.

    Thus, both perceptual detection and yes/no recognition tasks have the same decisional

structure in that they are both *binary choice tasks*. That is, in both tasks, participants make one of

two responses on each trial, "Yes" or "No." If we assume that tasks that share the same

decisional structure also share the same decisional process, then any differences in the patterns of

sequential dependencies that are observed between tasks of perception and tasks of memory are

not due to differences in decisional process, but are due to differences in the cognitive processes

of perception and memory. Figure 1.2 shows a schematic of this concept. The output of

mnemonic and perceptual processes is mapped by a decisional process onto some response

structure. If the response structures of the memory and perception tasks are the same, then this

decisional process is assumed to also be the same for both tasks. Thus, any differences in the

patterns observed between tasks cannot be due to differences in decisional process, but must be

due to differences in the cognitive processes of memory and perception.
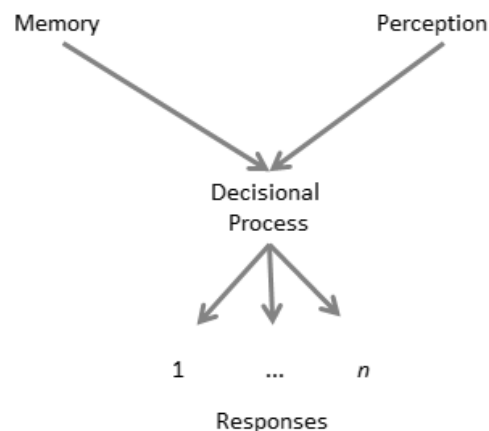
Figure 1.2. Schematic depicting the mapping of the output of the cognitive processes of memory
and perception onto a response structure.

1.2. Sequential Dependencies in Yes/No Recognition

Although both perceptual detection and yes/no recognition are similar to the extent that

they share the same response structure, there are key differences that must be considered. For

example, in a perceptual detection task, every trial presented is a test trial. That is, on every trial,

the participant must make a response to some event. On the other hand, in yes/no recognition, a

study list is presented in which participants do not make a response. The importance of this consideration becomes readily apparent when one considers the classic recognition priming literature of Ratcliff and McKoon (1978) in which they found that when the study order was equal to the test order, a recognition priming effect occurred, and responses were faster than when study and test orders were different. Schwartz et al. (2005) proposed that testing an item brings to mind items that occurred at nearby serial positions on the study list. Both of these accounts are types of the *enhanced memory hypothesis* of sequential dependencies. For example, consider Figure 1.3. Suppose items "Candle" and "Lamp" are studied at nearby serial positions and are then tested consecutively. If "Candle" is recognized and the participant responds "Yes" (this is also known as *hit*), according to Shwartz et al., this brings to mind items that were studied at nearby serial positions, namely "Ocean" and "Lamp". Thus, on the next trial, the probability of a "Yes" response to "Lamp" increases given the previous item was recognized as being studied. In other words the probability of a hit increases when there was a hit on the previous trial. On the other hand, if "Candle" is not recognized, then those nearby items on the study list are not brought to mind and the subsequent item enjoys no benefit from the previous trial.
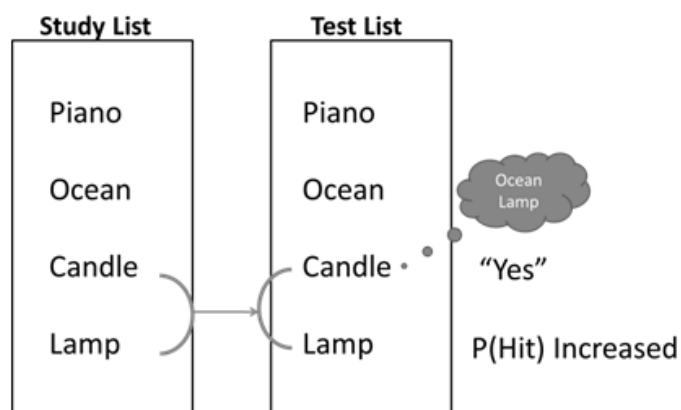
Figure 1.3. According the Schwartz et al. (2005) recognizing an item brings to mind items that were studied at nearby serial positions.

However, we know that in perceptual detection the probability of a "Yes" response is more likely following a "Yes" response than a "No" response. Thus, according to an *assimilation hypothesis*, study order should not matter. The probability of a hit should be greater when following a hit than when following a miss regardless of the study and test order sequence. Thus, there are two competing hypotheses of sequential dependencies in yes/no recognition: *assimilation* and *enhanced memory access.*

In order to test these hypotheses, we presented participants with a study list of 80 pairs of words (Malmberg & Annis, 2012). Following the study list, each item from the pair was consecutively tested along with foils. We refer to this condition as the *near pair condition.* An abbreviated example of the study and test list is shown in Figure 1.4 below. According the enhanced memory hypothesis, sequential dependencies should be observed in this condition. Specifically, the probability of a hit following a hit should be greater than the probability of a hit following miss. The assimilation hypothesis makes the exact same prediction.
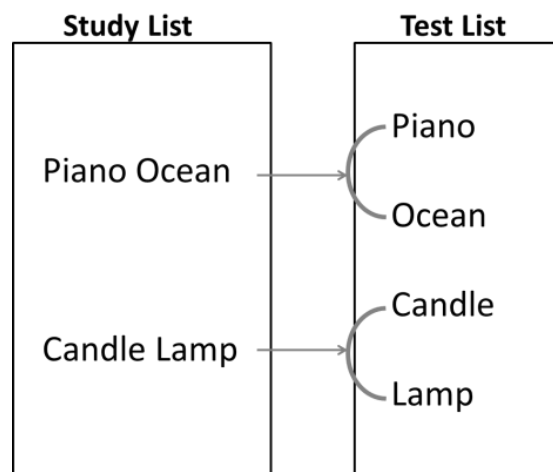
Figure 1.4. The near pair condition in which each item from the studied pair is tested consecutively.

In another condition, participants were again presented with pairs of words, but unlike the *near pair condition*, there were at least six intervening items between each item from the pair during test (see Figure 1.5). This condition we refered to as the *distant pair condition.* For this condition, the enhanced memory hypothesis predicts no sequential dependencies should occur. Specifically, the probability of a hit should be roughly the same regardless of whether it followed a hit or a miss. However, the *assimilation* hypothesis makes the same prediction as in the *near pair condition* – the probability of a hit is greater when following a hit than when following a miss.
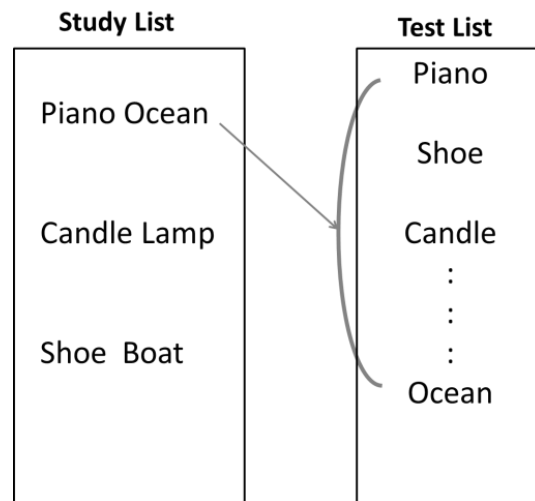
Figure 1.5. The distant pair condition in which there were at least six intervening items between each item from a given studied pair.

The results of the experiment are shown below in the left panel of Figure 1.6. In both the near pair and distant pair conditions, the probability of a hit following a hit was greater than the probability of a hit following a miss. Thus, this pattern of results is inconsistent with the enhanced memory hypothesis, but is consistent with the assimilation hypothesis. An even stronger test of the enhanced memory hypothesis is to look for sequential dependencies that exist between words that were not studied. If the word was not studied then a recollective process could not be involved in producing a sequential dependency. The right panel of Figure 1.6 shows

7

that a "yes" response to an unstudied item is more likely following a "yes" response than a "no" response. This pattern was found in the near and distant pair condition. These results were replicated with pictures, nonwords, and in single-item study. Thus, these data are inconsistent with the enhanced memory hypothesis.
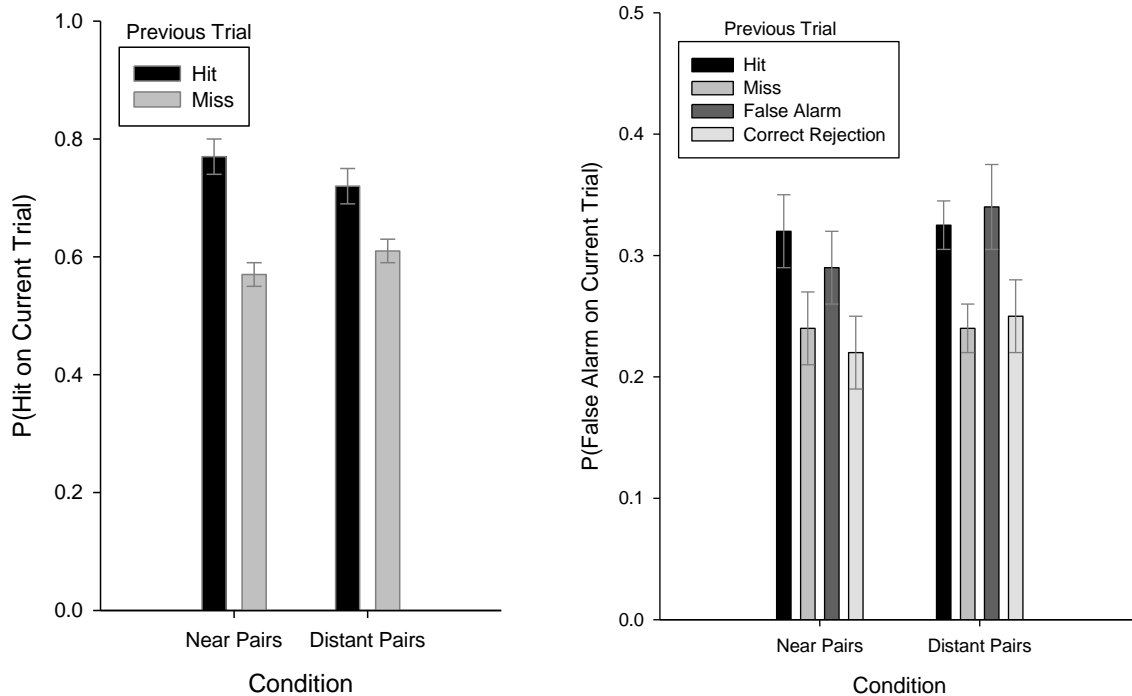
Figure 1.6. Results of near pair and distant pair conditions. The results are inconsistent with the enhanced memory access explanation of sequential dependencies.

The results are consistent with an assimilation hypothesis of sequential dependencies and inconsistent with the enhanced memory hypothesis; the probability of a "Yes" response was found to be greater when followed by "Yes" response than when followed by a "No" response regardless of the study-test-order mapping. Thus, the previous finding of Schwartz et al. may have been part of a broader pattern of assimilation. These positive sequential dependencies are similar to those found in perceptual detection.

1.3. Sequential Dependencies in Absolute Identification

As previously noted, there are a wide range of perception tasks besides perceptual detection in which sequential dependencies have been observed. In this section, I will focus on a task known as *absolute identification*. The reason for this is three-fold. Firstly, the sequential dependencies found in absolute identification are robust. Secondly, sequential dependencies have been extensively studied in absolute identification. Finally, there is an historical relationship between absolute identification and memory that dates back to Miller's (1956) paper on capacity limitations.

In an absolute identification task, the participant is presented with stimuli that vary along a single dimension. For example, the stimuli might be lines that vary in length. Each stimulus has an associated corresponding response. For example, Figure 1.7 shows that the smallest line length corresponds to a response of "1." The largest line length corresponds to a response of "10." The participant is usually given a series of practice trials with feedback until a given level of accuracy is achieved.
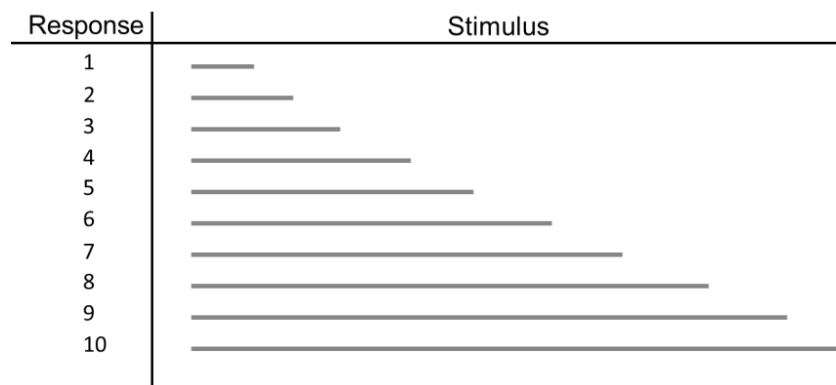
Figure 1.7. Example stimuli and their corresponding responses in an absolute identification task.

Assimilation is a robust finding in absolute identification. For example, suppose that a large line corresponding to a response of "10" is shown on trial 1 and the participant correctly identifies the stimulus as such. On the subsequent trial, suppose that a much shorter line is shown

9

that corresponds to a response of "1." When the previous response/stimulus is greater than the current stimulus, participants tend to overestimate the current stimulus value. That is, participants *assimilate* toward the previous stimulus/response. This is usually represented graphically by plotting the error on the current trial as a function of the previous stimulus and the current stimulus. The left panel of Figure 1.8 shows the results of Ward and Lockhead (1970) in which participants were presented with an absolute identification task. The stimuli were tones that varied in decibel level. Small stimuli are generally overestimated, indicated by positive error, and large stimuli are underestimated, indicated by negative error. More importantly, the left panel of Figure 1.8 also shows that as the previous stimulus increases participants increasingly overestimate the current stimulus value. Thus, participants *assimilate* toward the previous stimulus.

Another type of sequential dependency that is found in absolute identification is known as *contrast*. Unlike assimilation, which is a positive correlation between the previous stimulus/response and the current response, contrast is a negative correlation. Contrast is usually found between trial *n* and *n-j*, where *j* is referred to as *lag* and is greater than 1. The right panel of Figure 1.8 again shows the results of Ward and Lockhead (1970). At lags of 1, assimilation is present, however, at lags of 2 and greater, this pattern reverses. This is contrast. Thus, participants respond *towards* the previous stimulus and *away* from the stimuli that occur more than 1 trial back.

A task that is very similar to absolute identification, but is found in the recognition memory literature is known as a *judgment of frequency task*. In this task, participants are presented with a study list in which items are repeated a various number of times.
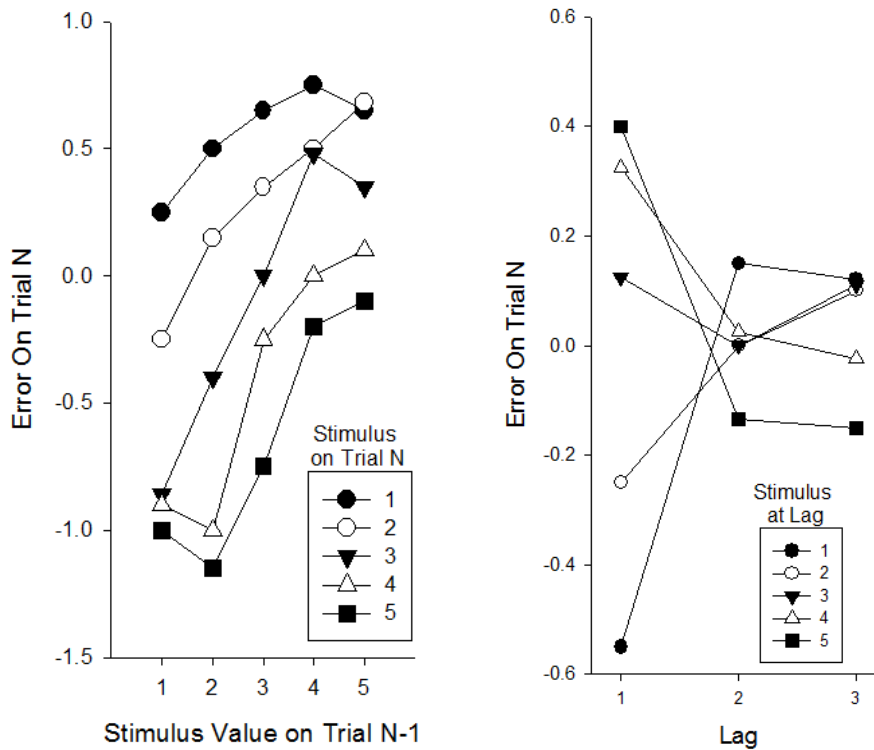
Figure 1.8. The left panel shows assimilation in an absolute identification task. The right panel shows contrast at lags greater than 1 in absolute identification. Points are estimated from the original figure found in Ward and Lockhead (1970).

For example, 10 stimuli may be presented 1 time, 10 other stimuli may be repeated 2 times, and so on. At test, the task of the participant is to determine the number of times each stimulus was presented. Thus, the response structure of absolute identification and judgments of frequency are the same. However, a different pattern of sequential dependencies arises from each task. Figure 1.9 shows the results of Malmberg and Annis (2012). In absolute identification (left panel), there is assimilation towards the previous stimulus and contrast at lags greater than 1, however, in judgments of frequency only assimilation is observed (right panel). Thus, if we assume that tasks that share the same decisional structure share the same decisional process, then the differences in the patterns of sequential dependencies that are observed between absolute

identification and judgments of frequency are due to differences in the perceptual and mnemonic

processes.
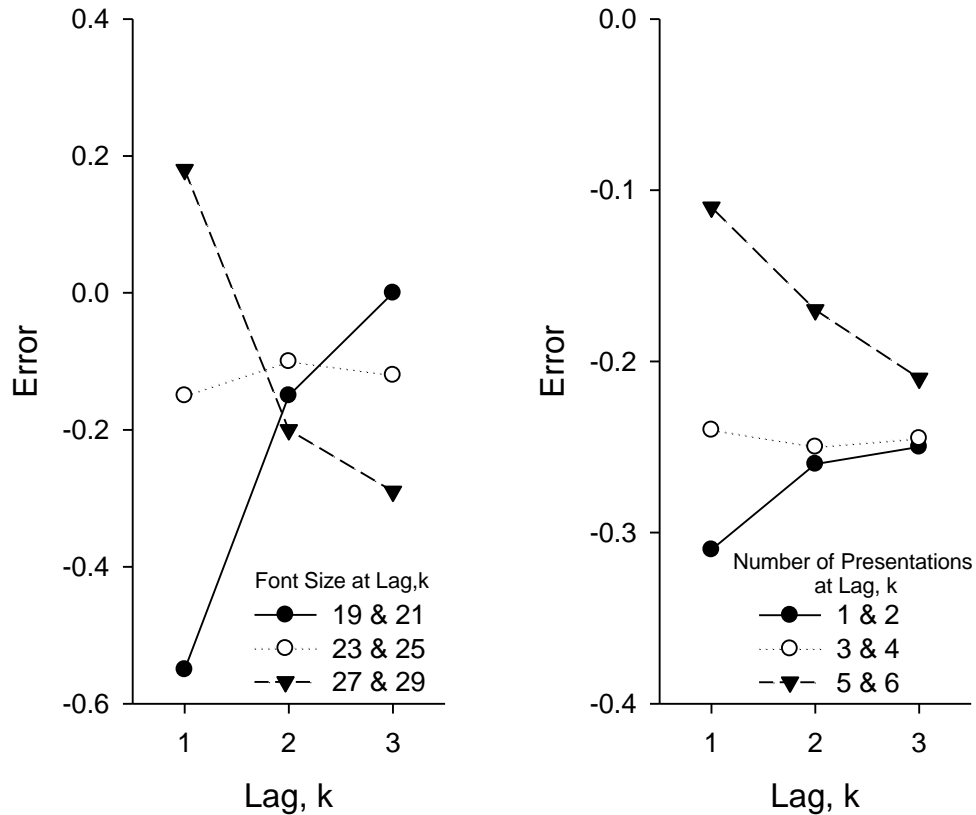
Figure 1.9. The left panel shows the results of the absolute identification task. The right panel shows the results of judgment of frequency task. Assimilation is present in both data sets at lags of 1. Contrast is present in the absolute identification task, but not in the judgment of frequency task.

1.4.   A Process Model of Sequential Dependencies in Judgments of Frequency

Annis & Malmberg (2013) developed a model that captures the patterns of sequential

dependencies seen in recognition memory. We chose to do this within the Retrieving Effectively

from Memory framework (REM; Shiffrin & Steyvers, 1997). REM assumes that memory traces

are represented as vectors of w geometrically distributed, features values. The environmental

base rate of feature values is determined by the geometric distribution parameter g. When an

item is studied, its lexical/semantic trace is activated, and t attempts to store a feature to an

12

episodic trace are made. Figure 1.10 gives an example in which the word "DOG" is presented at study. This activates the semantic memory trace associated with "DOG" and is then copied to episodic memory.

Figure 1.10. Studying an item activates the semantic memory trace which is then copied to episodic memory.

However, this copying process is not perfect. Sometimes a failure to store a feature in the episodic trace will occur. This assumption is formalized in the model by assuming the probability that a feature will be stored is u^*. In addition to storage failures, the model also assumes copy errors take place as well. This is formalized by assuming features are copied correctly from the lexical/semantic trace with probability c. If the feature is not copied correctly, then the stored value is drawn randomly from the geometric distribution. Thus, the episodic trace is usually a noisy and incomplete copy of the semantic trace.

During single item recognition, the test item's associated lexical/semantic trace serves as the retrieval cue. The retrieval cue is matched in parallel against episodic traces stored during study. This global matching process is graphically represented in Figure 1.11. For each episodic trace, the odds that the item was studied is computed. The odds represent the familiarity strength of the tested item. The odds are then compared to a criterion. If the odds are greater than the criterion value, an "Old" response is given, otherwise a "New" response is made. Multiple criteria values can be assumed for tasks involving multi-interval response sets.
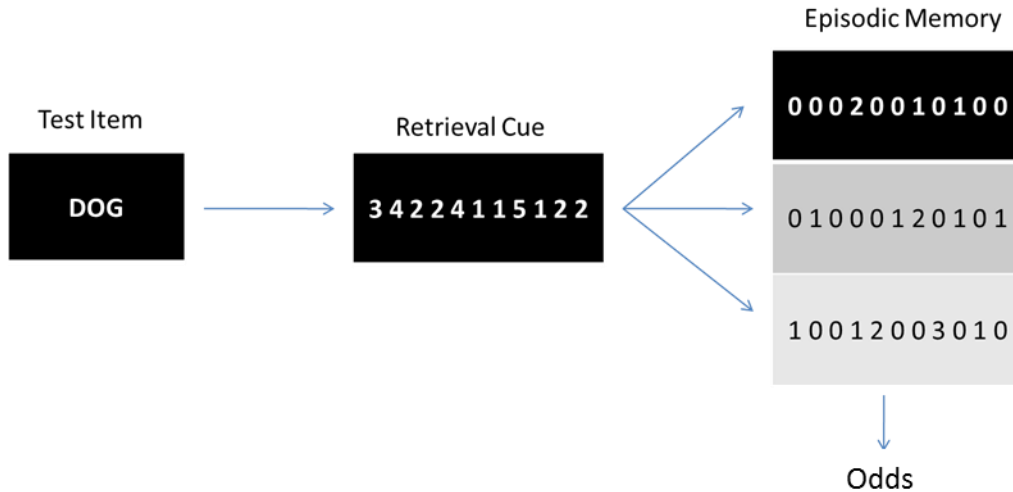
Figure 1.11. The retrieval cue is globally matched to an activated episodic memory set. The result is the log odds that the item was studied. This is the familiarity strength of the test item.

In order to model assimilation, Annis and Malmberg assumed that on each trial, there is a probability, 1-*a*, that a carryover process occurs in which features from the retrieval cue on the previous trial carry over to the retrieval cue on the current trial (cf. Huber, Shiffrin, Lyle, & Ruys, 2001). Therefore, as the *a* parameter increases, the number of trials in which carryover occurs decreases. Figure 1.12 shows an example of how the model works. Suppose on trial 1, the word DOG is presented. On trial 2 the word SHOE is presented, but features from the previous trial carry over. This is illustrated by the red arrows in Figure 1.12. Thus, the odds on trial 2 will be correlated with the odds on trial 1 resulting in response assimilation. On trial 3, with probability *a*, no carry over occurs. Thus, the response on trial 3 will be independent of previous responses. For readers interested in comprehensive fits of the model to judgment of frequency data, I refer them to our original manuscript (Annis & Malmberg, 2013).
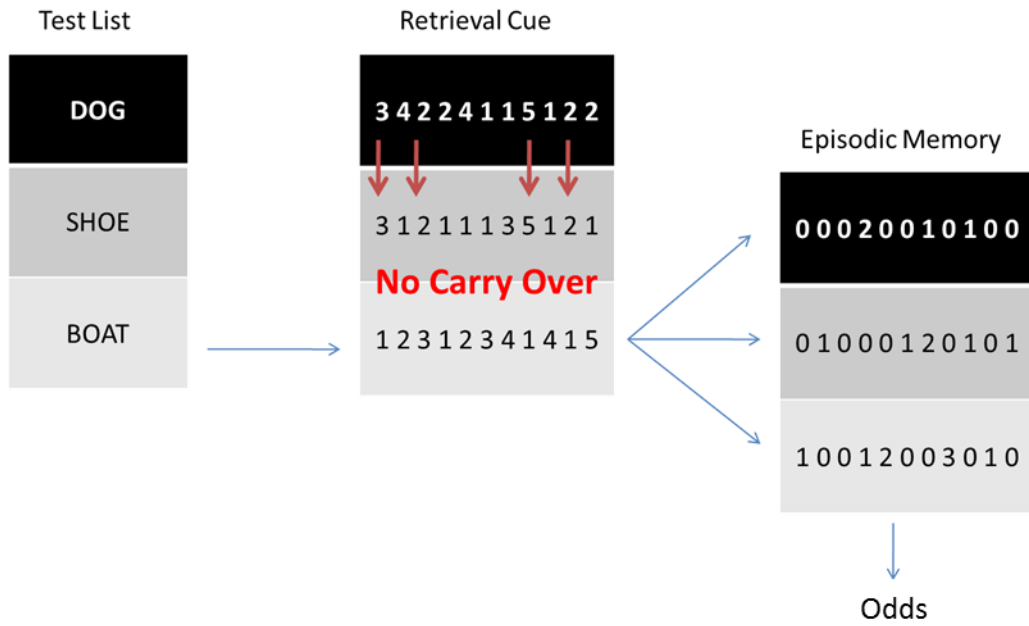
Figure 1.12. A model of assimilation in the REM framework. On each trial, with probability 1-*a*, features from the previous retrieval cue carry over to the next trial. No carryover occurs with probability, *a*.

Chapter 2

A Bayesian Model of Sequential Dependencies in Yes/No Recognition

Although sequential dependencies are robust and ubiquitous, and the model accounts for those found in recognition memory testing, there are inherent difficulties with measuring them within the frequentist architecture traditional to psychological research which I describe in section 2.3. The key problem that I identified is the ability of the models to separately measure sequential dependencies and shifts in response bias that occur between lists or individuals. The measurement issue is classically motivated insofar as researchers are often concerned with independently measuring bias and sensitivity. Here I will first describe this problem concretely and then I will provide several solutions within the frameworks of Bayesian Hierarchical models. I will show that these solutions provide independent measures of bias, sequential dependencies, and sensitivity.

2.1.   The Relationship between Bias and Sequential Dependencies

List-wide bias may artificially inflate the amount sequential dependencies present in the data. To pick a trivial example, if the participant responds "Yes" on nearly every trial, then the probability of a "Yes" response following a "Yes" response will be close to 1.0. However, this does not reflect a sequential dependency, rather a bias to respond "Yes." The key question is the extent to which the data reflect response bias or sequential dependencies.

In a binary choice task, there are four different types of response dependencies. Let the probability of responding "Yes" following a "Yes" response be q_{YY}, the probability of a "Yes"

response following a "No" response be q_{NY}, the probability of a "No" response following a "No"

response be q_{NN}, and the probability of a "No" response following a "Yes" response be q_{YN}.

These probabilities can be represented in the following matrix which shows all the different

types of current response and previous response combinations.

Current Response

		Yes	No
Previous Response	Yes	q_{YY}	q_{YN}
	No	q_{NY}	q_{NN}

Figure 2.1 Shows the probabilities associated with each current and previous response combination.

An important property of this matrix is that each row sums to 1. Therefore, $q_{YY} + q_{YN} = 1$ and

$q_{NY} + q_{NN} = 1$. The relationship between sequential dependencies and bias can be described

within this framework. The overall proportion of "Yes" responses is simply the average of the

first column.

$$P("Yes") = \frac{q_{YY} + q_{NY}}{2}. \tag{2.1}$$

Notice that as the probability of "Yes" responses increases, q_{YY} and q_{NY} also increase. Thus, if

one were to compare the probability of repeating a "Yes" response across conditions of an

experiment, there would be no way of knowing whether those differences were due to response

bias, sequential dependencies, or both. However, if one were to develop a formal model one may

interpret raw data within its framework. Therefore, the critical issue of this thesis is to develop

and test a model that allows one to separately measure sequential dependencies and response bias

under a wide range of conditions and for several tasks in which bias and sequential dependencies may affect performance.

One solution to the problem is similar to that of the classic problem of differentiating between response bias and sensitivity. In order to elucidate this similarity and perhaps glean an analogous solution to the problem, I will briefly outline the problem of differentiating between response bias and sensitivity. In a yes/no recognition test (see Chapter 1), there are two stimulus classes – studied items and unstudied items – to which there are two possible responses, "Yes" and "No". A "Yes" response to a studied item is known as a *hit*. A "Yes" response to an unstudied item is a *false alarm*. A "No" response to a studied item is a *miss*, and a "No" response to an unstudied item is a *correct rejection*. The probabilities of the stimulus and response combinations can be represented with the following matrix.

Figure 2.2. Stimulus-response matrix.

The overall probability of a "Yes" response can be calculated by averaging the probability of a hit and false alarms.

$$P(\text{"Yes"}) = \frac{P(Hit) + P(False\ Alarm)}{2}.$$ (2.2)

Notice that increases in the probability of "Yes" responses are confounded with increases in the hit rate. If the hit rate changed across conditions of an experiment it would not be possible to know whether the change was due to a change in the participant's ability to distinguish studied from unstudied items or whether the change was due to changes in response bias. One solution to this problem involves "correcting" the hit rate by subtracting the false alarm rate (Snodgrass & Corwin, 1988). When this difference is 0, this indicates that hit and false alarm rates are equal, and thus the subject was unable to differentiate between studied and unstudied items. When this measure is 1, the subject had perfect discriminability between studied and unstudied items. This logic might be similarly applied to distinguish bias from sequential dependencies. The probability of a "Yes" response following a "No" response will be subtracted from the probability of a "Yes" response following a "No" response. In terms of Figure 2.1, this quantity is equal to $q_{YY} - q_{NY}$. While this may be a simple measure, this solution is known to be confounded with bias and I do not consider it further (Snodgrass & Corwin, 1988).

Another solution to differentiating between bias and sensitivity comes from the application of signal detection theory to recognition memory (Green & Swets, 1966). Signal detection theory assumes that when an item is tested, a continuous, random variable representing the familiarity strength of the item is generated. This random variable is assumed to be drawn from a normal distribution that corresponds to one of the two stimulus classes. The unstudied distribution has mean, μ_N and the studied distribution has mean μ_S. The standardized difference between these two distributions is known as d'. If the variances of the target and foil distributions are equal, d' can be estimated by calculating the difference between standardized hit and false-alarm rates.

$$d' = \frac{\mu_S - \mu_N}{\sigma} = z(HIT) - z(FA). \qquad (2.3)$$

This measure is known to be independent of bias, under some conditions (Snodgrass & Corwin, 1988; Macmillan and Creelman, 1991). A measure of bias, referred to as c, can be estimated as the midpoint between standardized hit and false-alarm rates. This measure of bias is also independent of sensitivity, under some conditions (Snodgrass & Corwind, 1988).

$$c = \frac{z(HIT) + z(FA)}{2}. \tag{2.4}$$

However, the conditions under which these measures are independent are not satisfied by the recognition memory paradigm, and therefore various work-arounds were developed (Green & Swets, 1966; Egan, 1958). In any case, the important point is that under certain assumptions, it is possible to obtain independent measures of theoretically important constructs that are not revealed by raw observations.

An analogous remedy can be applied to the problem of distinguishing between bias and sequential dependencies. To generalize the results above, it is not necessary to assume the underlying distributions represent familiarity strength. On the other hand, the important properties of these measures being independent of one another will still apply. I define a new measure, given below, that represents a measure of sequential dependencies independent of bias.

$$\alpha = z(q_{YY}) - z(q_{NY}). \tag{2.5}$$

Alpha is the difference in the tendency to respond "Yes" following a "Yes" response on the prior trial and following a "No" response on the prior trial. It measures the effect on the tendency to response "Yes" of the prior response.

The same logic can be applied to a measure of bias that is independent of sequential dependencies. Equation 2.6 shows a measure of bias can be obtained by averaging over the z-transformed proportion of "Yes" responses following "Yes" response and the z-transformed proportion of "No" responses following "Yes" responses.

$$b = \frac{z(q_{YY}) + z(q_{NY})}{2}. \qquad (2.6)$$

This measures the average tendency to respond "Yes", without regard to the prior response. This is analogous to the measure of bias, c, in Equation 2.4.

The first model I develop will estimate α and determine the effect size of the sequential dependency. The second model I describe, will simultaneously estimate both bias and sequential dependencies, and provide effect sizes for both. Before describing these models I will first argue why I am going to use a Bayesian approach. Although these measures can be calculated in the frequentist framework as simple point estimates, there are a plethora of advantages to the Bayesian approach which I outline in the next section.

2.2.   The Bayesian Approach

There are a vast number of works that already describe Bayesian inference in detail and its many advantages (e.g. Gelman, Carlin, Stern, & Rubin, 2004; Jaynes, 2003; Kruschke, 2011; Lee & Wagenmakers, 2013; Rouder & Lu, 2005; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). Therefore, I will provide an outline of some these arguments and refer the reader to those works listed above for a more rigorous and thorough discussion.

In recent years, cognitive psychology has seen something akin to a "Bayesian revolution" in data analysis (e.g. Dennis, Lee, & Kinnell, 2008; Krushke, 2011; Lee, 2008; Lee & Webb, 2005; Morey, Pratte, & Rouder, 2008; Pitt, Myung, & Zhang, 2002; Rouder, Lu, Speckman, Sun, & Jiang, 2005) and cognitive modeling (e.g. Anderson, 1991; Griffiths, Sanborn, Canini, Navarro, & Tenenbaum, 2011; Lee, 2006; Navarro, & Griffiths, 2008). Given the number of advantages the Bayesian approach has, this is not surprising. A root cause for these advantages may come from representing parameter estimates as probability distributions. By representing parameters as probability distributions it is easy to embed knowledge about parameters, via

21

priors, in said probability distributions. It also provides a principled way, via Bayes Theorem, to update our knowledge about parameters in light of new evidence. Because parameter estimates are treated as probability distributions and not point estimates, the Bayesian approach provides us with a measure of uncertainty about the estimated parameter values.

There are also advantages that are specific to the problem of measuring sequential dependencies. The number of observations per cell is not predefined by the experimenter when measuring sequential dependencies because the number of observations per cell is dependent on the responses of the subject during testing and not on the experimental design. For example, the number of target trials following hits is inversely related to the number of target trials following misses. Therefore, if we ensure that the number of target trials following hits is high, then the number of target trials following misses will be low.

The frequentist approach to estimating the parameters of probabilistic models is to use maximum likelihood estimation (MLE; see Rice, 1995) which gives the parameter value that maximizes the likelihood of the data. A low number of observations leads to less reliable maximum likelihood estimates. For example, after observing a single coin flip produce heads, the maximum likelihood estimate for the rate at which the coin produces heads is equal to 1. This is a rather rash decision to make considering only a single coin flip was made. As more coin flips are made we can become more and more certain of the estimate, but this uncertainty is not explicitly conveyed by the maximum likelihood estimate. However, the Bayesian approach takes the number of observations into account when determining the uncertainty of the parameter estimate. For example, if the number of observations is 1 and we have uninformative prior knowledge about the parameter estimate, then a single observation is usually not very likely to drastically change our belief about the parameter estimate. As the number of coin flips increases,

the influence of our prior knowledge about the parameter decreases and the updated parameter estimate (the posterior) increasingly reflects the proportion in the data. For a detailed, mathematical discussion of this point see Kruschke (2011).

Another problematic issue for the frequentist arises when applying the z-transform to rates that are 1 or 0. If the rate is 1 or 0, then the z-transformed value is positive or negative infinity, respectively. In order to circumvent this issue, the researcher must use some type of edge correction such as adding .5 to the hit and false alarm rate count (Snodgrass & Corwin, 1988). This is not a problem for the Bayesian. This is because Bayesian inference can be implemented via Gibbs sampling which allows for the direct sampling of the posterior probability distribution, and although theoretically possible, given the nature of recognition data, it is unlikely that a rate of 1 or 0 will be sampled from the posterior distribution. It is also easy to implement constraints on the interval from which samples are drawn in order to ensure that infinite values are not a problem.

Finally, the Bayesian approach is hierarchical and allows the researcher to represent knowledge about parameter estimates at both the individual and group level simultaneously (Rouder & Lu, 2005). Group level estimates inform and constrain the individual estimates and the individual estimates inform the group. Thus, the parameter estimates for individuals inform the estimates for other individuals. This makes Bayesian inference resistant to outliers and is known as *shrinkage* (Kruschke, 2011). Given these advantages, the Bayesian approach will be used to address the issue of independently measuring bias and sequential dependencies.

2.3.   A Model of Binary Decisions

Figure 2.4 depicts the graphical model of a Bayesian *t*-test with binomially distributed data (Wagenmakers et al., 2010). Each node represents a variable, while the arrows represent the

conditional relationships between each variable. Shaded nodes represent observed variables (i.e. the data), while unshaded nodes represent unobserved random variables. Nodes containing concentric circles represent variables defined by a deterministic function. The square surrounding the nodes is called a "plate" and represents independent iterations of the model for $i...N$ subjects. Nodes outside the plate are not iterated for each subject, but are constrained by the entire group of subjects. The model can be implemented using the JAGS software (Plummer, 2003). The JAGS code for the model can be found in Appendix A.
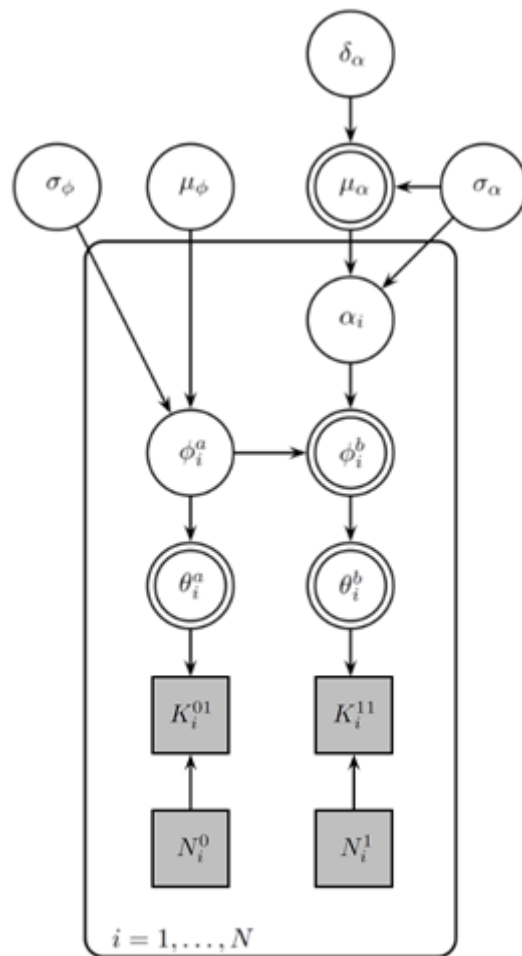
Figure 2.3. Shows a Bayesian graphical model of sequential dependencies in binary choice tasks.

I will describe the model in terms of "Old" vs. "New" responses, but the model could just as easily be described in terms of any binary response-set such as "Yes" vs. "No", "High" vs.

"Low", etc. Starting at the bottom left of the graphical model, the node, N_i^0, represents the number of "New" responses for subject i. Further up the graph, the node, θ_i^a, is the rate at which a "New" response is followed by an "Old" response. Thus, the number of "Old" responses following "New" responses, K_i^{01}, can be modeled via a binomial distribution with rate parameter, θ_i^a, and number of trials, N_i^0. Formally,

$$K_i^{01} \sim Binomial(\theta_i^a, N_i^0). \tag{2.7}$$

Since we are interested in comparing the rate of "Old" responses following "Old" responses and the rate of "Old" responses following "New" responses, the number of "Old" responses, N_i^1, is entered into the model (see the bottom right corner of the graphical model). The number of "Old" responses following "Old" responses, K_i^{11}, is binomially distributed with rate parameter θ_i^b and number of trials, N_i^1.

$$K_i^{11} \sim Binomial\left(\theta_i^b, N_i^1\right). \tag{2.8}$$

The goal of the model is to develop a measure of sequential dependencies that is independent of response bias. As detailed in section 2.1, this can be done by determining whether the z-transformed rate of responding "Old" differs as a function of the previous response. In terms of the current model, this is the difference, α_i, between the z-transform of the rates, θ_i^b and θ_i^a. Instead of modeling the rates and applying the z-transformation as in Eq. 2.5, I directly model the z-transformed rates, $z(\theta_i^a)$ and $z\left(\theta_i^b\right)$, as normally distributed random variables, ϕ_i^a and ϕ_i^b, respectively. Thus,

$$\alpha_i = \phi_i^b - \phi_i^a \tag{2.9}$$

In order to get back to the rate scale, the probit transformation is applied. The probit transformation can map any real number to the rate scale. Figure 2.4 shows the probit transformation maps, for example, a value of 0 on the probit scale to a value of .5 on the rate

scale. Rearranging Eq. 2.9 and applying the probit transform, the following reparameterization

for θ_i^b is obtained:

$$\theta_i^b = \Phi(\phi_i^b) = \Phi(\phi_i^a + \alpha_i). \tag{2.10}$$

This will later allow estimates of effect size to be made. I assume that the differences between

the transformed rates for each participant are drawn from a group-level normal distribution with

mean μ_α and standard deviation, σ_α. ϕ_i^a is modeled as a normally distributed variable with

group-level mean μ_ϕ and standard deviation, σ_ϕ.
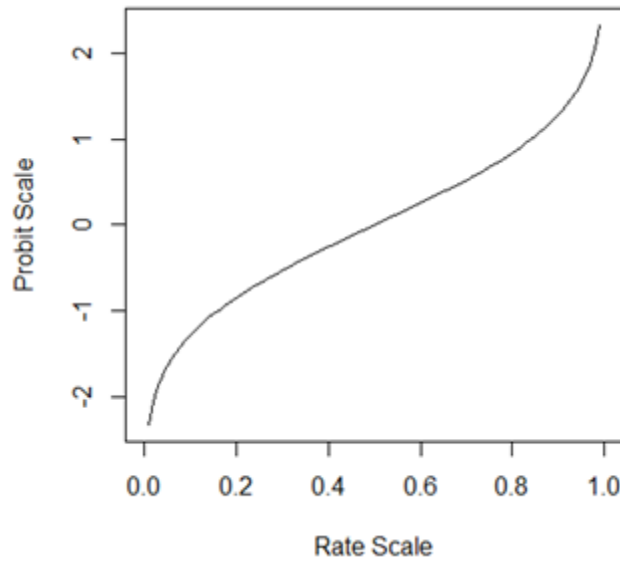
Figure 2.4. The probit function relates the rate scale and the probit scale.

The next step in defining the model is to define the group-level priors. The group-level

mean for the effect of the previous response on the current response is μ_α. Instead of modeling

μ_α directly, it is determined by the product of the effect size δ_α and the standard deviation σ_α.

$$\mu_\alpha = \delta_\alpha \times \sigma_\alpha, \tag{2.11}$$

This allows priors to be placed on the effect size, δ_α. The prior I choose to place on δ_α is a

weakly informative prior known as the "unit information prior" (Wagenmakers et al. , 2010;

Kass and Wasserman, 1995) and contains as much information as a single observation. The

motivation for this is to conduct a type of Bayesian hypothesis test which is described in the next section. Following Wagenmakers et al. (2010) the priors placed on the variances σ_ϕ and σ_α were reasonably uninformative uniform distributions.

$$\sigma_\phi, \sigma_\alpha \sim Uniform(0, 10). \tag{2.12}$$

Finally, I assumed the group-level mean of the z-transformed rates of responding "Old" given a "New" response, μ_ϕ, followed a standard normal distribution.

$$\mu_\phi \sim Normal(0, 1). \tag{2.13}$$

2.3.2. Measuring Changes in List-wide Bias and SDs. Having described the model, I will now describe how the model can be used to estimate the rate of sequential dependencies and test whether, for example, the rate of observing a "Yes" response is greater following a "Yes" response than when following a "No" response. These two analyses can be classified as parameter estimation and Bayesian hypothesis testing, respectively. In Section 2.5, I will test whether this model is valid measure of bias and SDs.

Parameter estimation is carried out by estimating the 95% Highest Density Interval (95% HDI; Kruschke, 2011). The 95% HDI contains 95% of area under the posterior probability curve and which follows the constraint that any density estimate inside the interval is greater than any density estimate outside the interval. In addition to parameter estimation, Kruschke (2011) suggests the HDI can also be used in the context of hypothesis testing. Following the decision rule proposed by Kruschke (2011), parameter values falling within the 95% HDI are defined as "credible" while those falling outside the HDI are defined as "not credible."

Another method of testing a hypothesis within a Bayesian framework is to calculate the Bayes Factor.

$$BF_{01} = \frac{P(D|H_0)}{P(D|H_1)}, \tag{2.14}$$

where the numerator is the likelihood of the data under the null hypothesis and the denominator is the likelihood of the data given the alternative hypothesis. Thus, if $BF_{01} > 1$ then the null hypothesis is more likely than the alternative. If $BF_{01} < 1$ then the alternative hypothesis is more likely than the null. It can be shown that the the ratio of the height of the posterior probability and prior probabilities is the Bayes Factor. This is known as the Savage-Dickey Ratio Test (Dickey, 1971; Wagenmakers et al., 2010). In the context of the model described above, I test the hypotheses that the effect size is equal to zero (i.e. $\delta_\alpha = 0$) vs. the hypothesis that the effect size is not equal to zero (i.e. $\delta_\alpha \neq 0$).

2.4. Simulating Binary Data with a Markov Chain

It often useful to test a model's ability to measure what it intends to measure with data having known properties. In order to generate the data set to test the model, I will employ the use of a Markov chain to create simulated data with known properties. For a comprehensive treatment of Markov chains I refer the reader to Kemeny and Snell (1976). A Markov chain is a series of discrete time steps each associated with a random variable, X_t, whose outcome only depends on the previous time step, *t*-1. Thus,

$$P(X_t = i_0 | X_{t-1} = i_1)$$
$$= q_{i_0,i_1}, \tag{2.15}$$

where $i = \{1,...m\}$, and *m* is the total number of states. A state transition diagram is a graphical representation of a Markov chain. The diagram depicted below shows a Markov chain with two-states, Y and N, and the associated state transitions.
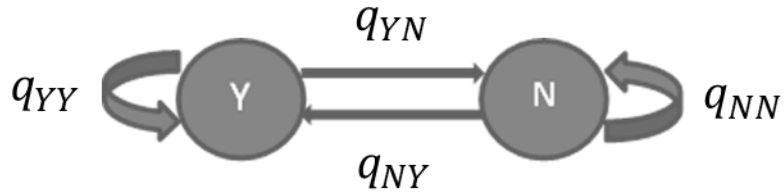
Figure 2.5. State transition diagram of a two-state Markov chain.

The probability of transitioning from state Y, at time t -1, to state Y, at time t, is q_{YY}. The

probability of repeating state N is q_{NN}. The probability of transitioning from state Y, at time t-1,

to state N, at time t, is q_{YN}, and the probability of transitioning from state N to state Y is q_{NY}. All

of these state transitions can be compactly represented in a matrix. This matrix is known as the

transition matrix. The transition matrix, Q, for the state diagram above is

$$Q = \begin{bmatrix} q_{NN} & q_{NY} \\ q_{YN} & q_{YY} \end{bmatrix}, \tag{2.16}$$

where each row must sum to 1. The overall distribution of each state is given by a row vector, v,

known as the *state distribution vector*, whose elements must also sum to 1. The state distribution

vector represents the probabilities of the system being in each particular state.

$$v = [q_N \ q_Y]. \tag{2.17}$$

Multiplying the transition matrix by the state distribution vector results in the state

distribution at $t + 1$. For example, suppose the transition matrix is

$$\begin{bmatrix} .6 & .4 \\ .8 & .2 \end{bmatrix}, \tag{2.18}$$

and the state distribution vector at time t is

$$[.3 \ .7]. \tag{2.19}$$

If the transition matrix is multiplied by the state distribution vector, the state distribution at time,

$t + 1$ is obtained.

$$[.3 \ .7] \begin{bmatrix} .6 & .4 \\ .8 & .2 \end{bmatrix} = [.74 \ .26]. \tag{2.20}$$

29

We can repeat this process to find the state distribution at the next given time step by multiplying the new state distribution vector with the transition matrix.

$$[.74 \; .26] \begin{bmatrix} .6 & .4 \\ .8 & .2 \end{bmatrix} = [.652 \; .348]. \tag{2.21}$$

If one wanted to find the state distribution vector 100 steps into the future this process would be quite cumbersome. A much easier way to find the state distribution at some time step, n, is to raise the transition matrix by n and multiply it by the state transition vector. For example, if n were 20 then the state distribution vector at time step 20 is given by

$$[.3 \; .7] \begin{bmatrix} .6 & .4 \\ .8 & .2 \end{bmatrix}^{20} = [.67 \; .33]. \tag{2.22}$$

An important property of the Markov chains described here is that as the number of time steps increases, the state distribution stabilizes. In order to give a concrete example, suppose we wish to find the state distribution at $n = 100$ using the given state distribution and transition matrix below.

$$[.3 \; .7] \begin{bmatrix} .6 & .4 \\ .8 & .2 \end{bmatrix}^{100} = [.67 \; .33] \tag{2.23}$$

Notice the state distribution vector did not change (or changed insignificantly) between 20 and 100 time steps. This state distribution vector is known as the *steady state vector*. It should be noted that with large enough time steps, this method will return the same steady state vector regardless of what the initial state distribution vector is.

2.4.1. Simulation Details. In order to simulate a data set with *no* bias, but *with* sequential dependencies, I used the transition matrix Q described above and set $q_{NY} = .2$ and $q_{YY} = .8$.

$$Q = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix}, \tag{2.24}$$

30

The steady state of the transition matrix can be found by applying the method above (see

Kemeny and Snell for the analytic method). Again, the state distribution vector is arbitrary and

will return the same steady state no matter what it is.

$$[.1 \ .9] \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix}^{500} = [.5 \ .5] \tag{2.25}$$

Thus, the overall proportion of this Markov chain will have an equal proportion of Y states and

N states. The transition matrix used to generate a data set with both bias and sequential

dependencies had $q_{NY} = .6$ and $q_{YY} = .8$. Thus, the transition matrix for this data set was

$$Q = \begin{bmatrix} .4 & .6 \\ .2 & .8 \end{bmatrix}, \tag{2.26}$$

The overall proportion of Y states was therefore .75. The transition matrix for the data set

containing no bias and no sequential dependencies was

$$Q = \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix}, \tag{2.27}$$

Obviously, the proportion of Y states was also .5. Finally, the transition matrix for the data set

containing bias, but no sequential dependencies was

$$Q = \begin{bmatrix} .2 & .8 \\ .2 & .8 \end{bmatrix}, \tag{2.28}$$

Thus, the probability of a Y state was .8. In order to generate the chain I implemented the

Markov chain in R (see Appendix B for the R code) using a chain length of 500 for 20 subjects.

The algorithm is as follows:

1. Sample row i of the transition matrix according to a uniform distribution. The row
   number, i, corresponds to the state of the Markov chain at time t-1.

2. Given the row vector sampled on the previous step, sample the state corresponding to
   column j. The column number, j, corresponds to the state of the Markov chain at time t.

3. Sample row j of the transition matrix.

4. Set $t = t + 1$

5. Repeat steps 2 through 4.

2.5.   Tests of the Binary Model with Simulated Data

In this section, I will take data generated from the Markov model described in Section 2.4 to test the model described in Section 2.3. In all simulations, 20 chains, each 50 elements long were generated. The model was implement using the open-source JAGS software (Plummer, 2003). To run this model, three chains, each consisting of 53,000 samples, were generated. The chains were visually checked for convergence and the first 3000 burn-in samples were discarded. The panels of Figure 2.6 show the results of four tests of the model in which combinations of bias and sequential dependencies were either present or absent in the simulated data. In the Bayesian t-test, the height of the posterior distribution over zero of the effect size, δ, and the height of prior distribution over zero are computed. The prior distribution is normal with a mean of 0 and standard deviation of 1. This is known as the "unit information prior" and carries with it the amount of information provided by one observation (Kass and Wasserman, 1995). The ratio of the height of the posterior density and prior density over zero gives the Bayes Factor, of the effect size, δ, at zero. Bayes Factors greater than 1 indicates support for the hypothesis that the effect size, δ, is equal to 0. Bayes Factors less than 1 indicate support for the hypothesis that the effect size is greater than or less than 0.

The top left panel of Figure 2.6 is the result of a simulation of Markov chain which contained no bias, but did contain sequential dependencies. Specifically, the probability of an "Old" or a "New" response were both set to .5. The probability of an "Old" response given an "Old" response was set to .8, and the probability of an "Old" response given a "New" response was .2. The top left panel shows the posterior and prior densities for δ. The dots represent the

height of the posterior and prior densities at zero. The height of posterior density of δ at zero is lower than the height of the prior density at zero. This indicates that the data decreased the support for the hypothesis that the effect size, δ is equal to zero. Thus, the model correctly detected the sequential dependencies in the data.

The top right panel was the result of an analysis of a data set that contained both bias and sequential dependencies. Specifically, the probability of an "Old" response was set to .75. The probability of repeating an "Old" response was set to .8 and the probability of an "Old" response given a "New" response was set to .6. The bottom left panel shows the posterior density of the effect size and the prior effect size. The model was again able to detect sequential dependencies in the data.

The bottom left panel of Figure 2.6 shows the results of the analysis over data which contained neither bias nor sequential dependencies. The probability of an "Old" response following an "Old" response was equal to the probability of an "Old" response following a "New" response. The probability of an "Old" response was also equal to the probability of a "New" response. The posterior density of the effect size over zero increased from the prior density over zero. Thus, the data increased support for the effect size being equal to zero and the model correctly identified the absence of any sequential dependencies in the data.

The bottom right panel of Figure 2.6 shows the critical analysis of data in which there was bias, but no sequential dependencies. This was implemented by setting the probability of an "Old" response following an "Old" response was set to .8, and the probability of an "Old" response following a "New" response to .8. The posterior density of the effect size over zero increased from the prior density over zero indicating that the data increased support for an effect

size equal to zero. Thus, the model was able to distinguish between bias and sequential dependencies.
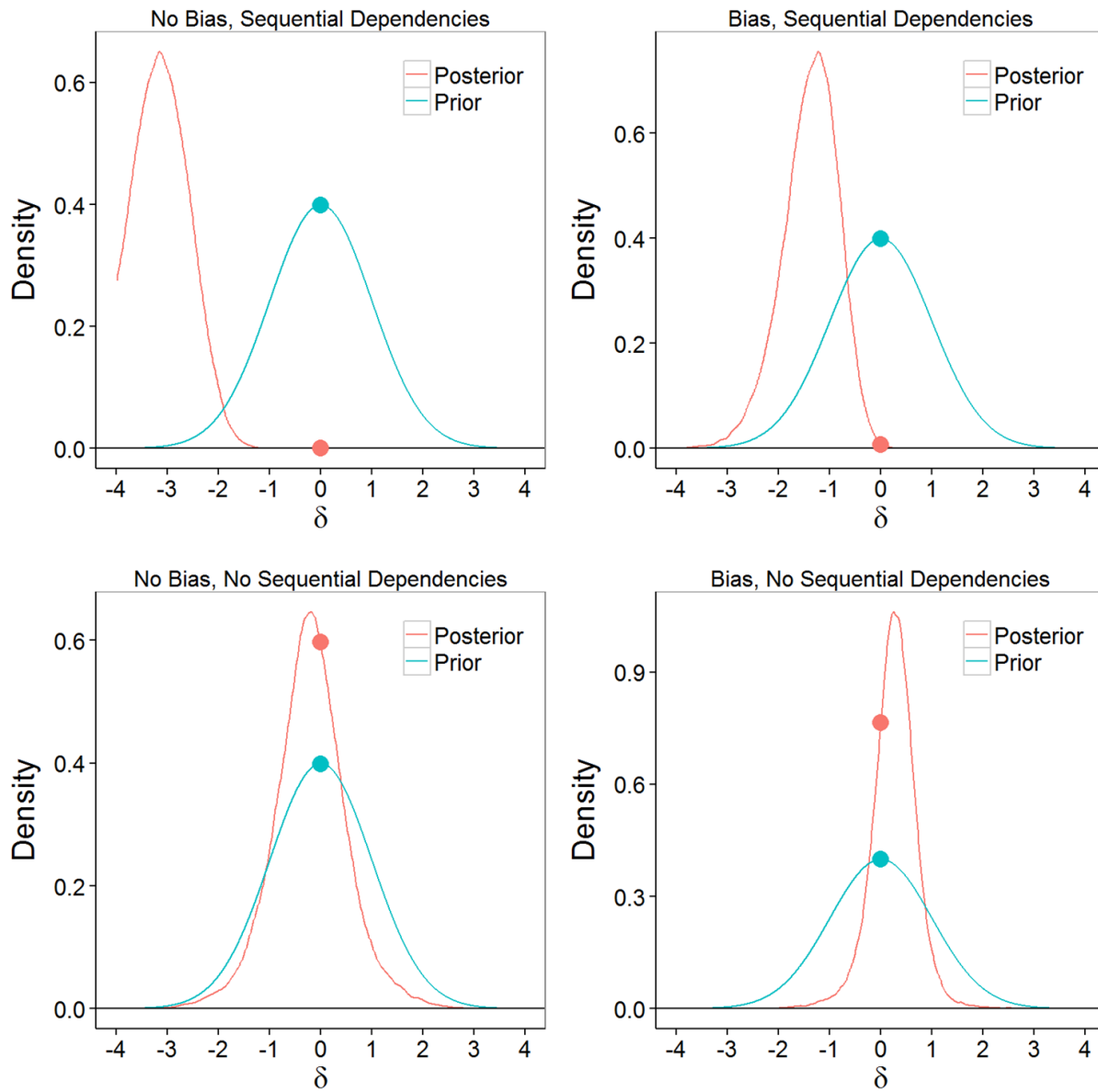
Figure 2.6. Analysis of simulated binary data containing combinations of bias and sequential dependencies.

2.6.   Modeling Bias and Sequential Dependencies Simultaneously

Although the previous model was able to distinguish between bias and sequential dependencies, it would be convenient to model both bias and sequential dependencies simultaneously. As we will see, by explicitly modeling both bias and sequential dependencies, the model becomes more efficient in detecting bias and sequential dependencies or a lack thereof. That is, the model requires less data to reach the same conclusion about bias and sequential dependencies as its counterpart described above. The model is depicted in Figure 2.7. The code can be found in Appendix C.

We again begin by modeling the frequency of "Old" following "Old" responses, K_i^{11} as a binomial distribution with rate parameter, θ_i^b. The total number of "Old" response trials is N_i^1. On the bottom right side of Figure 2.7, the number of "Old" responses following "New" responses, K_i^{11} is depicted and is assumed to be binomially distributed with rate parameter, θ_i^a. The number of "New" response trials is N_i^0. In order to explicitly model bias, I define the bias parameter, β_i, as the midpoint between the z-transformed rates, θ_i^a and θ_i^b.

$$\beta_i = \frac{z(\theta_i^a) + z(\theta_i^b)}{2}. \tag{2.29}$$

Therefore, if β_i is greater than 0, then this indicates that participant, i, is biased to respond "Old". An explicit sequential dependency parameter, α_i, can also be defined as the difference between the z-transformed rates, θ_i^b and θ_i^a.

$$\alpha_i = z(\theta_i^b) - z(\theta_i^a). \tag{2.30}$$

Thus, if α_i, is greater than 0, this indicates that the ith participant has exhibited a tendency to respond "Old" following "Old" more so than "Old" following "New." Notice that this model is isomorphic to Signal Detection Theory (Green & Swets, 1966; Lee, 2008), the only

difference being, I am modeling the rate at which participants respond "Old" following either

"Old" or "New", instead of modeling hits and false alarm rates.
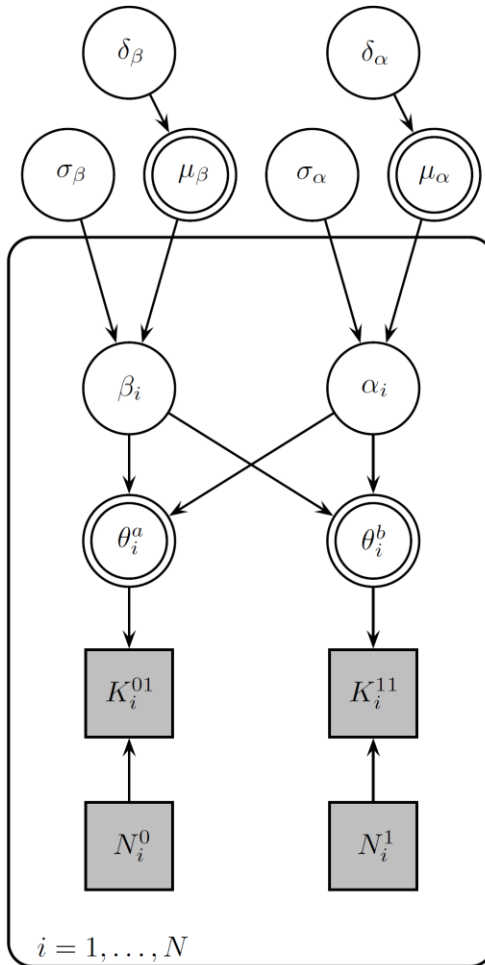
Figure 2.7. Simplified version of the binary choice sequential dependency model.

　　　　Having explicitly defined the bias and sequential dependency parameters, the next step is

to reparameterize the above equations in terms of θ_i^a and θ_i^b . This is a necessary step in order to

explicitly model the bias and sequential dependency parameters in a Bayesian hierarchical

framework (Lee, 2008). Rearranging the equation and applying the probit transform as discussed

in section 2.4, the reparameterization is given by Eq. 2.31 and Eq. 2.32.

$$\theta_i^a = \Phi\left(b_i - \frac{\alpha_i}{2}\right). \tag{2.31}$$

$$\theta_i^b = \Phi\left(b_i + \frac{\alpha_i}{2}\right). \tag{2.32}$$

Having rearranged the equations in terms of θ_i^a and θ_i^b it is now possible to model the bias and sequential dependency parameters as normal distributions. The sequential dependency parameter for each subject, α_i, is drawn from the group-level normal distribution with mean, μ_α, and standard deviation, σ_α. The bias parameter, β_i, is drawn from a group-level normal distribution with mean, μ_b and standard deviation, σ_α.

The group-level mean for the effect of the previous response, μ_α, is the product of the effect size δ_α and the standard deviation σ_α. Likewise, the group-level mean for bias, μ_b, is set to the product of the effect size δ_b and the variance σ_b. This allows for a Savage-Dickey ratio test to be performed and the Bayes Factor over the effect size to be calculated. The priors for the effect size of sequential dependencies, δ_α, and bias, δ_b, are unit information priors (Jeffreys, 1961). The priors for the group-level standard deviations of sequential dependencies and bias are reasonably uninformative (Wagenmakers et al., 2010).

$$\sigma_\alpha, \sigma_b \sim Uniform(0, 10). \tag{2.33}$$

2.7. Tests of the Bias Model with Simulated Data

All simulation details were exactly the same as those detailed in the previous section. The top panels of Figure 2.8 are the result of a simulation which contained no bias. Specifically, the probability of an "Old" or a "New" response were both set to .5. In addition, the data set contained sequential dependencies. The probability of an "Old" response given a "New" response was set to .8, and the probability of an "Old" response given a "New" response was .2. The top left panel shows the posterior and prior densities for the bias effect size, δ_β. The black dots represent the height of the posterior and prior densities at zero. Note the prior density

(dotted line) is not visible because the density of δ_β over zero is very large. This indicates that there was a large amount of support for the hypothesis that the effect size for bias, δ_β, is equal to zero. This is correct, in that, the probability of an "old" or "new" response was equal in this data set. The top right panel shows the posterior and prior densities for δ_α. The dots show the height of the densities at zero. The height of posterior density of δ_α at zero is lower than the height of the prior density at zero. This indicates that the data decreased the support for the hypothesis that the effect size, δ_α, is equal to zero. Thus, the model correctly detected the sequential dependencies in the data. It is important to note the estimates of the effect sizes show much less variance than those obtained from the previous model. By explicitly taking into account bias, the model is able to estimate the effects more efficiently. This pattern of improved efficiency is true for all of the analyses that I conducted for this model.

The bottom two panels were the result of an analysis of a data set that contained both bias and sequential dependencies. Specifically, the probability of repeating an "Old" response was set to .75. The probability of repeating an "Old" response was set to .8 and the probability of an "Old" response given a "New" response was set to .6. The bottom left panel shows the posterior density of the effect size for bias, δ_β, and the prior effect size. It is clear the posterior density of δ_β is sharply peaked about zero indicating the data decreased support for the hypothesis that the effect size, δ_β, is equal to zero. Thus, the model successfully detected bias in the data. The bottom right panel shows the height of the posterior density of δ_α over zero is less than that of the height of the prior density over zero indicating the data decreased support for the hypothesis that the effect size for sequential dependencies, δ_α, is equal to zero.
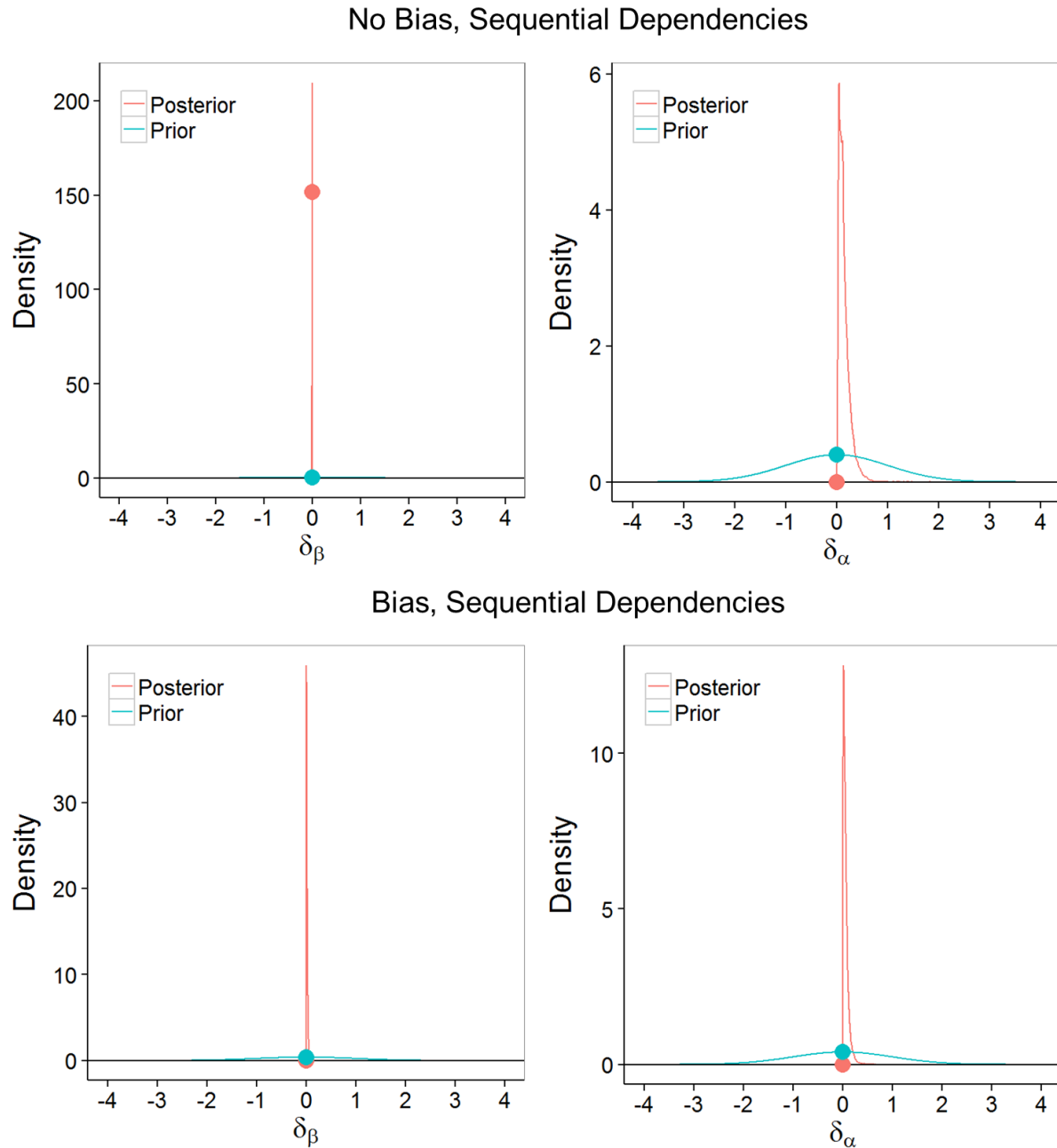
Figure 2.8. Analysis of simulated binary data using the bias model. The top two panels are the result of data which contained no bias and sequential dependencies. The bottom two panels are the result of data which contained both bias and sequential dependencies. The left panels depicts the Savage-Dickey ratio test over the bias parameter, δ_β, while the right panels shows the results of the sequential dependency parameter δ_α.

Figure 2.9 shows further tests of the model. The data, whose analysis is shown in the top two panels, were generated via a Markov chain (details described above) in which the probability of an "Old" response was set to .8. Thus, the data set contained a substantial bias to respond

"Old". The probability of an "Old" response following an "Old" response was set to .8 and the probability of an "Old" response following a "New" response was also set to .8. Thus, the data set did not contain sequential dependencies.

The top left panel shows the posterior and prior densities of the effect size of bias, δ_β. The posterior density is sharply peaked over zero indicating decreased support for the hypothesis that δ_β is equal to zero. Thus, the model successfully detected the large amount of bias in the data. The top right panel shows the posterior and prior densities of the effect size for sequential dependencies, δ_α. The height of the posterior density over zero is substantially larger than the prior density over zero indicating increased support for the null hypothesis, $\delta_\alpha = 0$. Thus, the model correctly identified the absence of sequential dependencies in the data.

The bottom two panels of Figure 2.9 show the final analysis in which neither sequential dependencies nor bias were present in the data. In order to simulate this data set, the probability of an "Old" and "New" response were both set to .5. The probability of an "Old" response following an "Old" response was equal to the probability of an "Old" response following a "New" response. The bottom left panel shows the prior and posterior densities of the effect size for bias, δ_β. The bottom right panel shows the densities of the effect size for sequential dependencies, δ_β. In both cases, the height of the posterior density over zero is substantially larger than the height of the prior density over zero, correctly indicating that the data increased support for the null hypotheses that the effect size for bias, δ_β, and the effect size for sequential dependencies, δ_α, is equal to zero. All of the posterior probability estimates for δ_α in the fours simulations reported above appeared to converge on an estimate more quickly than the model that did not contain a bias parameter. In order to verify this, the 95% HDIs were calculated for both models.

40

Figure 2.9. Analysis of simulated binary data using the bias model. The top two panels are the result of data which contained bias and no sequential dependencies. The bottom two panels are the result of data which contained no bias and no sequential dependencies. The left panels depicts the Savage-Dickey ratio test over the bias parameter, δ_β, while the right panels shows the results of the sequential dependency parameter δ_α. Note that the dotted line indicating the prior distribution is not visible in some of the panels because of the large scale of the density.

Figure 2.9 shows the upper and lower bounds of the 95% HDI for δ_α under the model

with no bias parameter and under the model with the bias parameter. In all simulations, the

distance between the upper and lower bounds of the 95% HDI were smaller when the bias

parameter was included. On average, this distance decreased by 92% when the bias parameter

was considered. This provides good evidence that including the bias parameter increases the

efficiency of the model.

Table 2.1. Shows the upper and lower bounds of the 95% HDI of the sequential dependency
parameter δ_α under models with and without a bias parameter. The interval distance is the
absolute value of the difference between the upper and lower bounds. The distances are
compared by computing the percentage decrease.

Simulation Type	δ_α 95% HDI No Bias Parameter			δ_α 95% HDI With Bias Parameter			Comparison
	Lower Bound	Upper Bound	Interval Distance	Lower Bound	Upper Bound	Interval Distance	Interval Distance % Decrease
No Bias, Seq. Dep.	-4.41	-2.02	2.39	0.02	0.36	0.34	85.80
Bias, Seq. Dep.	-2.51	-0.37	2.14	0.00	0.17	0.17	92.24
No Bias, No Seq. Dep.	-1.73	1.28	3.01	-0.04	0.04	0.09	97.08
Bias, No Seq. Dep.	-0.65	1.14	1.80	-0.10	0.04	0.14	92.08

2.8.   Analyzing Multiple Conditions

The models described above are useful for determining whether sequential dependencies

exist in a single condition of an experiment. In this section, I will extend the model to account for

changes in sequential dependencies across multiple conditions of an experiment. This is a

desirable property insofar as the researcher is interested in how changes in certain factors (e.g.

vigilance) might cause changes in the degree of sequential dependencies. Figure 2.10 shows the

Bayesian graphical model of sequential dependencies for two conditions. The code can be found

in Appendix D.
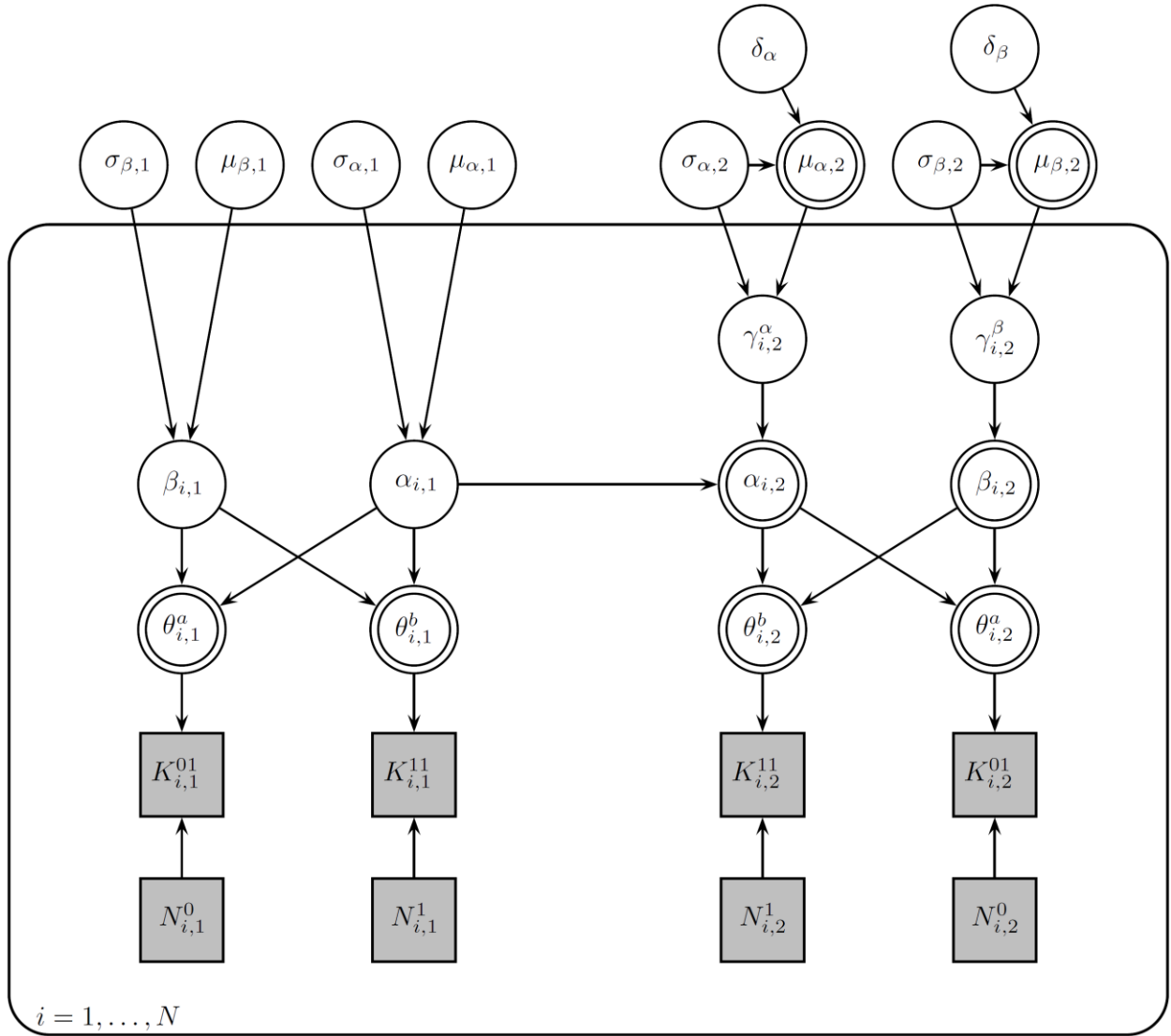
Figure 2.10. Bayesian graphical model of sequential dependencies for comparing two conditions.

The graphical model is very similar to the previous model described above. The number

of "Old" responses following "New" responses in condition 1, $K_{i,1}^{01}$, is binomially distributed

with rate parameter, $\theta_{i,1}^{a}$. The total number of "New" responses in condition 1 is $N_{i,1}^{0}$. Thus,

$$K_{i,1}^{01} \sim Binomial\left(\theta_{i,1}^{a}, N_{i,1}^{0}\right). \tag{2.34}$$

The number of "Old" responses following "New" responses in condition 2 mirrors the structure

for condition 1.

$$K_{i,2}^{01} \sim Binomial(\theta_{i,2}^a, N_{i,2}^0). \tag{2.35}$$

The number of "Old" responses following "Old" responses for conditions 1, and 2 are also modeled as a binomial distributions.

$$K_{i,1}^{11} \sim Binomial(\theta_{i,1}^a, N_{i,1}^1). \tag{2.36}$$
$$K_{i,2}^{11} \sim Binomial(\theta_{i,2}^a, N_{i,2}^1). \tag{2.37}$$

The rate at which an "Old" response follows an "Old" response for conditions 1 and 2 are the probit transform of the bias parameter, β, and the sequential dependency parameter α (see previous model for derivation).

$$\theta_{i,1}^a = \Phi\left(\beta_{i,1} - \frac{\alpha_{i,1}}{2}\right). \tag{2.38}$$
$$\theta_{i,2}^a = \Phi\left(\beta_{i,2} - \frac{\alpha_{i,2}}{2}\right). \tag{2.39}$$

The rate at which an "Old" response follows a "New" response for conditions 1 and 2 follow the same logic.

$$\theta_{i,1}^b = \Phi\left(\beta_{i,1} - \frac{\alpha_{i,1}}{2}\right). \tag{2.40}$$
$$\theta_{i,2}^b = \Phi\left(\beta_{i,2} - \frac{\alpha_{i,2}}{2}\right). \tag{2.41}$$

The bias parameter for condition 1, $\beta_{i,1}$, is distributed normally with mean, $\mu_{\beta,1}$ and standard deviation, $\sigma_{\beta,1}$.

$$\beta_{i,1} \sim Normal(\mu_{\beta,1}, \sigma_{\beta,1}). \tag{2.42}$$

The change in bias from condition 1 to condition 2, $\gamma_{i,2}^\beta$, is assumed to be normally distributed mean, $\mu_{\beta,2}$ and standard deviation, $\sigma_{\beta,2}$.

$$\gamma_{i,2}^\beta \sim Normal(\mu_{\beta,2}, \sigma_{\beta,2}). \tag{2.43}$$

Therefore, the bias parameter for condition 2 is deterministic.

$$\beta_{i,2} = \beta_{i,1} + \gamma_{i,2}^\beta \tag{2.44}$$

The sequential dependency parameter for condition 1, $\alpha_{i,1}$, is distributed normally with mean, $\mu_{\alpha,1}$, and standard deviation, $\sigma_{\alpha,1}$.

$$\alpha_{i,1} \sim Normal(\mu_{\alpha,1}, \sigma_{\alpha,1}). \tag{2.45}$$

The difference between $\alpha_{i,1}$ and $\alpha_{i,2}$ is given by $\gamma_{i,2}^{\alpha}$ which follows a normal distribution with mean, $\mu_{\alpha,2}$ and standard deviation, $\sigma_{\alpha,2}$.

$$\gamma_{i,2}^{\alpha} \sim Normal(\mu_{\alpha,2}, \sigma_{\alpha,2}). \tag{2.46}$$

The sequential dependency parameter for condition 2 is given by

$$\alpha_{i,2} = \alpha_{i,1} + \gamma_{i,2}^{\alpha}. \tag{2.47}$$

The effect size for bias and sequential dependencies are modeled by δ_{β} and δ_{α}, respectively.

$$\delta_{\beta} \sim Normal(0,1) \tag{2.48}$$
$$\delta_{\alpha} \sim Normal(0,1) \tag{2.49}$$

2.9. Tests of the Multiple Conditions Model

Data were simulated according to the Markov chain technique described in section 2.4. Figure 2.11 shows the results of the analysis of a data set which contained neither differences in bias nor differences in sequential dependencies across the two conditions. In both conditions, the probability of an "Old" response following either an "Old" or "New" response was set to .8. In both conditions, the overall probability of an "Old" response was set to .8. The posterior density estimates over 0 for both δ_{α} and δ_{β} are greater than the prior indicating increased support for the null hypothesis that the effect size is equal to zero. Thus, the model successfully identified there were no differences in either bias or sequential dependencies between conditions.
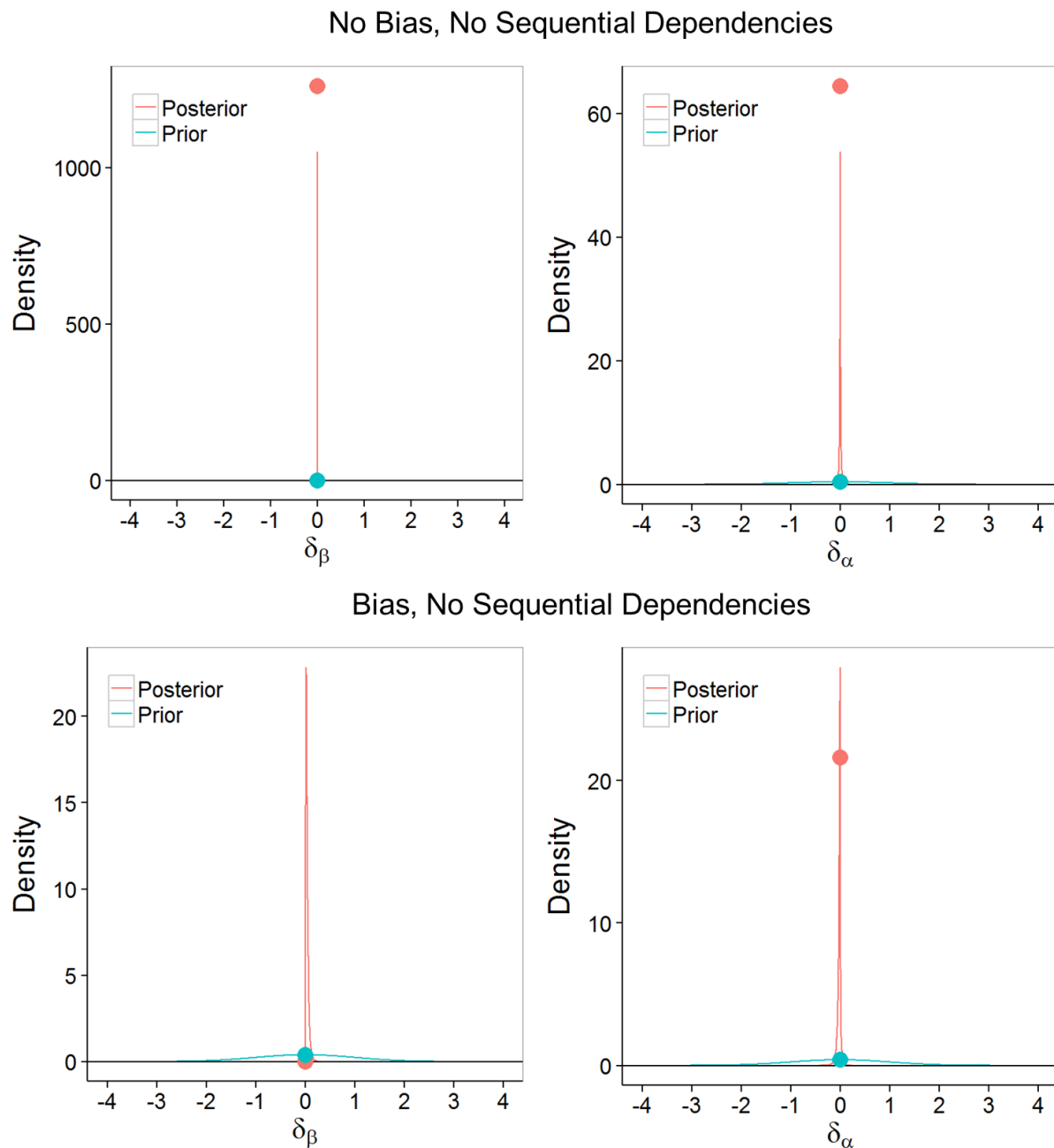
Figure 2.11. Analysis of simulated binary data using the bias model. The top two panels are the result of data which contained neither differences in bias nor differences sequential dependencies across conditions.

Figure 2.12 shows the results of an analysis in which the data did not contain differences in bias across conditions, but did contain differences in sequential dependencies. In the first condition, the probability of repeating an "Old" response was .8, while the probability of responding "Old" following a "New" response was .2. In the second condition, the probability of responding "Old" following either an "Old" or "New" response was .5. In both conditions, the probability of responding "Old" was .5. The left panel shows the posterior density for δ_β over zero increased from the prior density indicating the data increased support for the null hypothesis. The right panel shows the opposite outcome for δ_a in which the data decreased the support for the null. The model, therefore, was able to successfully detect differences in sequential dependencies between conditions, while, at the same time, indicating no differences in bias between conditions.

Same Bias, Different Sequential Dependencies
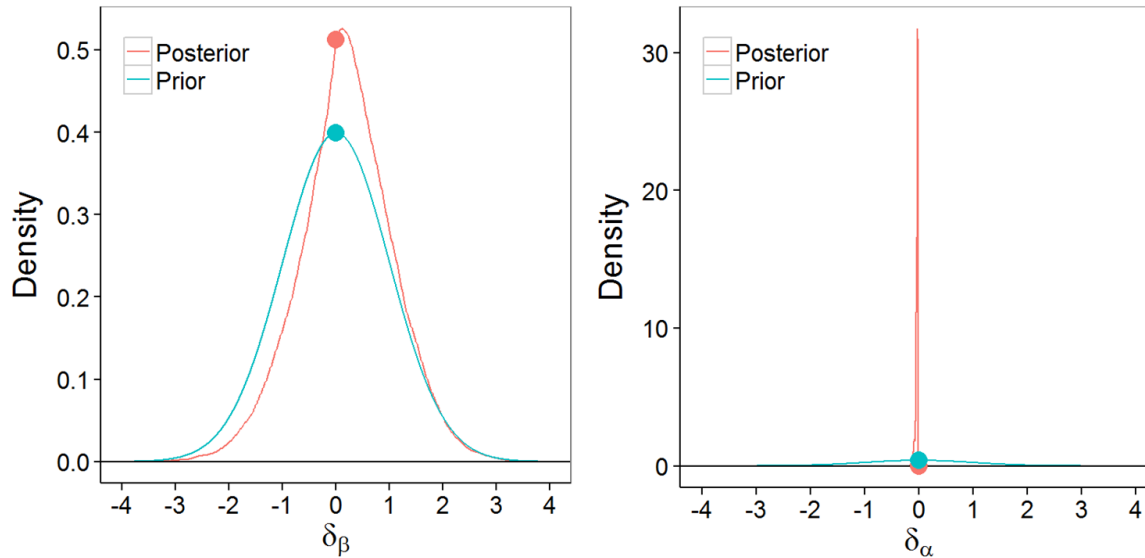
Figure 2.12. Shows the result of the model analysis with data which contained no differences in bias across conditions, but did contain differences in sequential dependencies.

Figure 2.13 shows the results of a simulation in which there were changes in bias between two conditions, but there were no changes in the amount of sequential dependencies across conditions. The probability of an "Old" response was set to .75 in the first condition, and set to .5 in the second condition. The probability of an "Old" response following a "New" response was set to .6 in the first condition and .4 in the second condition. In both conditions, "Old" responses were 20% more likely following "Old" responses than "New" responses. The posterior density over zero decreased for δ_β indicating the data increased support for the alternative hypothesis that there was a difference in bias between conditions. The right panel shows the data increased support for the null hypothesis that there were no differences in sequential dependencies between conditions. Thus, the model was able to successfully detect the differences in bias across conditions. In addition, it was able to find evidence for the null hypothesis that there no differences in sequential dependencies across the two conditions.

Different Bias, Same Sequential Dependencies
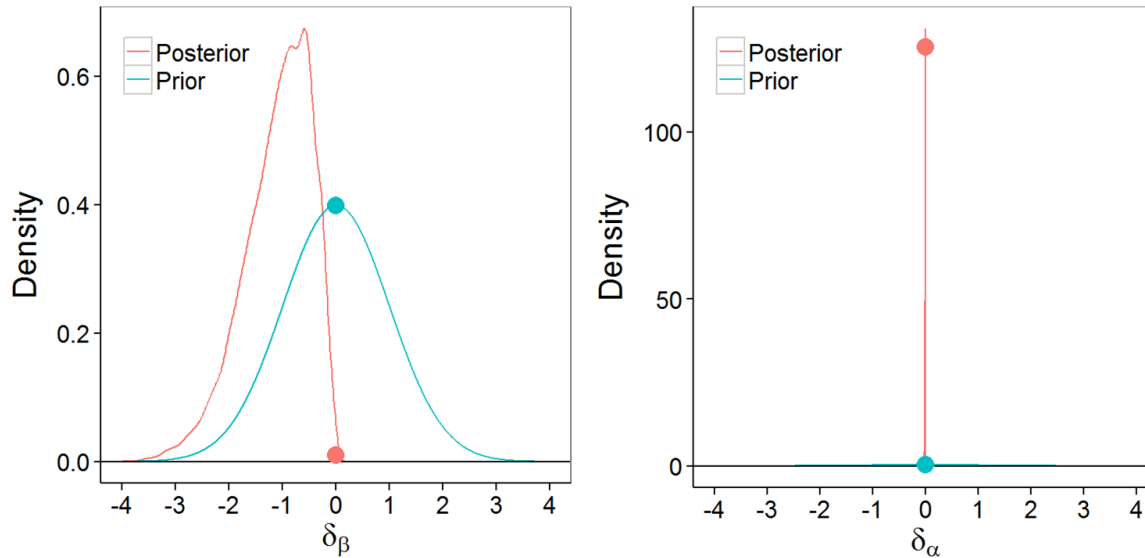
Figure 2.13. Analysis of simulated binary data using the bias model. Shows the result of data which contained differences in bias, but no differences sequential dependencies across conditions.

Figure 2.14 shows the results of an analysis in which there were differences in both bias and sequential dependencies across conditions. The probability of responding "Old" in the first condition was .75, while the probability of responding "Old" in the second condition was .5. The probability an "Old" response following a "New" response was .6 in the first condition and .8 in the second condition. The probability of an "Old" response was 20% higher in the first condition than in the second condition. As depicted in Figure 2.14, the model was successful at detecting both differences in bias and sequential dependencies across conditions.

Different Bias, Different Sequential Dependencies
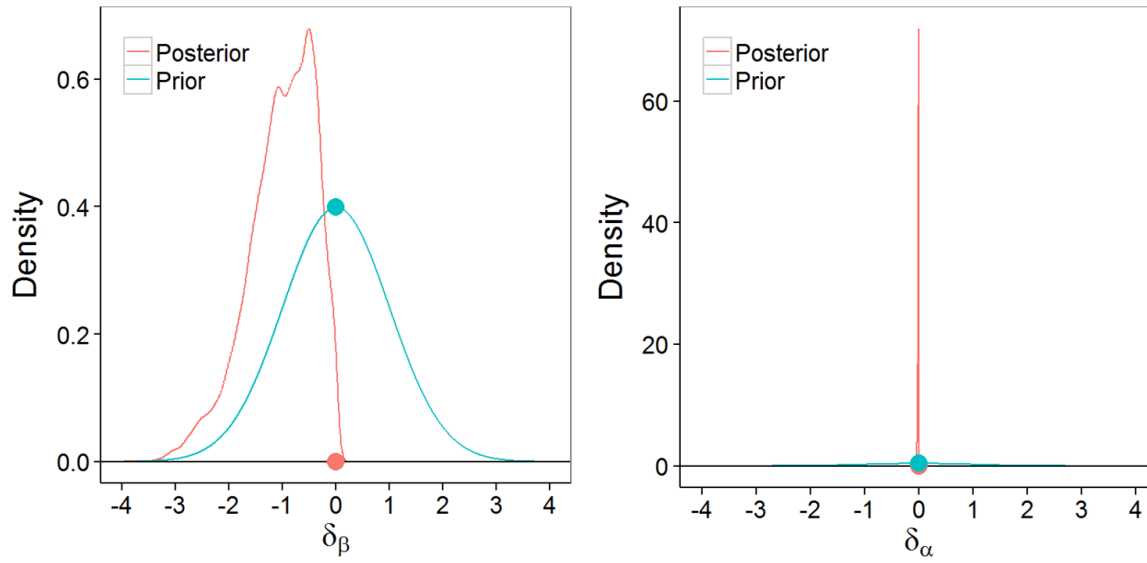
Figure 2.14. Analysis of simulated binary data using the bias model. Shows the result of data which contained both differences in bias and sequential dependencies across conditions.

Chapter 3

Models of Sequential Dependencies in Multi-Interval Tasks

The models outlined in the previous chapter were able to distinguish between list-wide response bias vs. sequential dependencies. Given the preliminary success of the Bayesian approach to modeling sequential dependencies in yes/no recognition, I seek to extend this model to multi-interval response tasks such as absolute identification and judgments of frequency. There are several issues associated with the standard frequentist approach that can be resolved by utilizing a Bayesian framework.

Prior attempts to model assimilation and contrast have often involved the use of frequentist regression models (e.g. Jones, Love, and Maddox, 2006; Jestedt, Luce, and Green, 1988). All of these models assume that the effect of the previous stimulus/response must be factored into the current response, but this holds only within the frequentist framework. This limitation does not exist in a Bayesian approach. For example, the proportion of trials in which sequential dependencies occur could be modeled by a binomial distribution whose rate parameter could be estimated.

Another general advantage of the hierarchical Bayesian approach is that it allows the simultaneous fitting of individual and group parameters. As far as the author knows, individual differences have not been studied with regard to sequential dependencies. A hierarchical Bayesian approach to measuring sequential dependencies would provide such information.

3.1. Measuring Sequential Dependencies in Multi-Interval Designs

Models of cognition can be divided into two broad classes: measurement models and process models (Malmberg, 2008). Although both models map stimuli onto responses via some latent mechanism, process models make explicit assumptions about how stimuli are represented and processed while measurement models skip this step and model decisional processes *given* the processed stimulus. The model that I describe will attempt *measure* sequential dependencies and the related decisional components involved in a multi-interval response task. Therefore, the model that I am proposing will *not* attempt to explain the underlying processes that are responsible for generating sequential dependencies (cf. Annis & Malmberg, 2013; Petrov & Anderson, 2005). Rather, the proposed model will contain a minimal number of theoretical assumptions regarding the cognitive mechanisms involved in the generation of sequential dependencies.

The utility of such a model should be readily apparent. For example, the model will be able to be generalized to a variety of different multi-interval tasks, and will increase our ability to accurately measure sequential dependencies. This is not to say that the model will bear no relationship to underlying process models. Indeed, the present model is tailored such that it can be related to specific assumptions found in process models such as the one described by Annis & Malmberg (2013). In this fashion, we are in a better position to relate measurement with underlying processes (Estes, 1975; Gillund & Shiffrin, 1984; Malmberg, 2008), an approach that is often overlooked. The current model will attempt to measure various decisional components in a multi-interval response task within a Bayesian hierarchical framework for the first time. The components that the model will attempt to describe are listed below.

List-wide bias may artificially inflate the amount sequential dependencies present in the data. Thus, the model will be able to distinguish between response bias and sequential dependencies.

The model will contain parameters that will describe the strength of the stimulus-to-response mapping. For example, participants often "bin" their responses on extreme ends of the response scale which results in increased accuracy for stimuli lying on extreme ends of the scale (Luce, Nosofsky, Green, & Smith, 1982). Thus, it becomes necessary to differentiate between this type of response bias and the actual sensitivity between adjacent stimulus values.

The model will contain parameters that will identify the *magnitude* and the *direction* of the sequential dependency between the current response and stimuli occurring on trials $n - j$, where $j > 0$. For example, the sign of the parameter could indicate whether the sequential dependency is negative or positive. Thus, depending on the sign of the parameter, the researcher could readily identify whether assimilation or contrast is present in the data. The value of the parameter could indicate the magnitude of the sequential dependency.

The model will contain one or more parameters that will indicate the *magnitude* and *direction* of the sequential dependency between the current response and the previous responses for lags greater than 1.

The model will attempt to differentiate the effect of the prior responses from prior stimuli. For example, as accuracy increases, the correlation between the previous stimulus and the previous response also increases. One way to decorrelate the effect of the previous stimulus from the previous response is to hold the previous stimulus constant while varying the previous response. Aside from the analysis being rather awkward, data sparsity also becomes an issue when conducting this analysis. For example, there are very few "6" responses to stimuli that

correspond to a "1" response. This is a very large overestimation that participants are not inclined to make very often. Therefore, in order to carry this analysis, it is usually necessary to bin the responses.

Unlike previously developed regression models (e.g. Jones, Love, & Maddox, 2005) the model that I will develop will not assume that the effect of prior stimuli/responses should always be factored into the current response. Rather, the model will attempt to identify the proportion of trials in which sequential dependencies occurred.

The Bayesian hierarchical framework naturally lends itself to describing individual differences while simultaneously estimating group level parameters. Thus, all of the above phenomena will be measured at both the group and individual level.

3.2.   Multi-Interval Model R: A Model of Response Sequential Dependencies

In some of the most theoretically useful experiments, decisions require the use of a multi-interval scale (e.g., Likert Scale), and sequential dependencies are quite commonly reported. In this section, I will describe a model of sequential dependencies in multi-interval tasks that attempts to address the bulleted points discussed above. It is related to several generalized measurement models by Kruschke (2011) and Morey, Pratte, and Rouder (2008). The JAGS code can be found in Appendix E. The probability of participant i giving response k on the jth trial, is the standardized area of the normal distribution, θ_{ij}, between the criterion, C, associated with response k and k-1.

$$p\big(y_{ij} = k \mid \theta_{ij}\big) = \Phi\left(\frac{C_k - \theta_{ij}}{\sigma_S}\right) - \Phi\left(\frac{C_{k-1} - \theta_{ij}}{\sigma_S}\right). \qquad (3.1)$$

The probability of making the highest response, K, is the area under the standardized normal above the highest criterion, C_{K-1}.

$$p(y_{ij} = K \mid \theta_{ij}) = 1 - \Phi\left(\frac{C_{K-1} - \theta_{ij}}{\sigma_S}\right). \tag{3.2}$$

The probability of making the lowest response is the area under the curve below the lowest criterion value.

$$p(y_{ij} = 1 \mid \theta_{ij}) = \Phi\left(\frac{C_1 - \theta_{ij}}{\sigma_S}\right). \tag{3.3}$$

The standard deviation associated with each stimulus value, S, follows an inverse gamma distribution with near uninformative priors.

$$\sigma_S \sim InverseGamma(.001, .001). \tag{3.4}$$

The criteria are assumed to be drawn from a normal distribution. The lowest criterion has a prior mean of 0. Each additional criterion's prior mean is incremented by .5 as suggested by Kruschke (2011). All criteria share the same prior standard deviation of 1. When there are only two response criteria, Morey et al. (2008) suggests fixing them at 0 and 1 for simplicity.

An instance of this model, in which there are two criteria and thus three response types, is represented graphically in Figure 3.1. The solid curve represents the distribution of θ_{ij} and the vertical dotted lines represent the criteria, C_1 and C_2. The area under the curve falling below C_1 representing the probability of making a "1" response is quite low. Therefore, participant i has a very low probability of responding "1" on trial j. The area falling above the highest criterion, C_2, is also small which reflects a low probability of responding "3." The area between C_1 and C_2 is much greater. Therefore, there is a much higher probability of the participant responding with a "2" than with a "1" or a "3."

The core of the model, θ_{ij}, is a linear combination of latent effects (Gelman et al., 2004; Morey et al., 2008).

$$\theta_{ij} = \beta_S + \gamma_{i,S} + x_{i,j}(\delta_{i,P}), \tag{3.5}$$

where β_S is the overall latent strength associated with each stimulus value, S, and is normally distributed with a mean of 0 and standard deviation, $\sigma_S^{(\beta)}$.
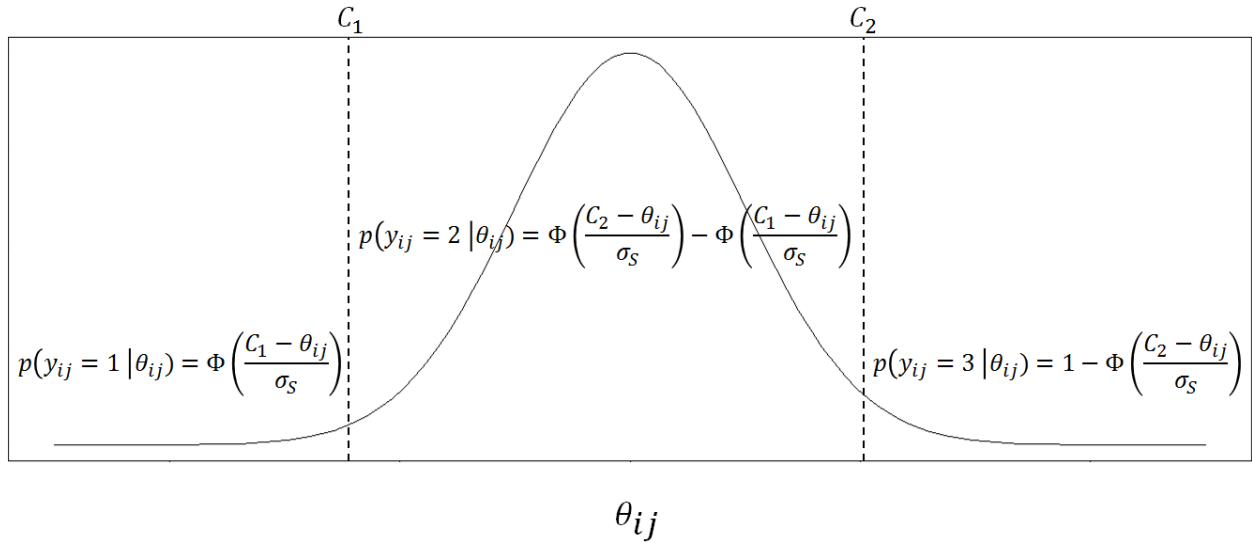
Figure 3.1. Illustrates how the continuous variable, μ_{ij}, is mapped onto an ordered response set. The vertical dotted lines represent each criterion. In this example, the area under the curve below the first criterion, C_1, (Eq. 3.2) represents the probability of making a "1" response. The area between the C_1 and C_2 (Eq. 3.1) represents the probability of making a "2" response. The area above the highest criterion C_2 (Eq. 3.3) represents the probability of making a "3" response.

I assume the effect of each stimulus, S, may vary depending on the participant and that this effect is additive. The effect, $\gamma_{i,S}$, is assumed to be drawn from a group-level normal distribution with mean, μ_S and standard deviation, $\sigma_S^{(\gamma)}$.

$$\gamma_{i,S} \sim Normal\left(\mu_S, \sigma_S^{(\gamma)}\right). \tag{3.6}$$

The final latent effect entered into the model is the effect of the previous response, $\delta_{i,P}$, where the subscript, P, represents the value of the previous response. This effect is drawn from a group-level normal distribution with mean, r_P, and standard deviation, $\sigma_P^{(\delta)}$.

As discussed above, it is important to relate measurement models to underlying processes (Malmberg, 2008). Therefore, the model is tailored to the assumptions of the process model

developed by Annis and Malmberg (2013). This process model of sequential dependencies was developed within the REM framework (described in Chapter 1) and assumes that information from the previous trial carries over to current trial. In order for the model to fit the data, we found it necessary to assume that information carryover does not occur on every trial, but only on a portion of trials. The probability of not carrying over information was referred to as the a parameter. In order to tailor the measurement model to the assumptions of the process model, I assume that information from the previous trial does not always influence the current response. In order to implement this assumption, the effect of the previous response, $\delta_{i,P}$ is multiplied on each trial by $x_{i,j}$ which follows a Bernoulli distribution with success probability $1-a_i$.

$$x_{i,j} \sim Bernoulli(1 - a_i) \tag{3.7}$$

Thus, when $x_{i,j}$ takes on a value of 1, the effect of the previous response is factored into the current response. However, when $x_{i,j}$ takes on a value of 0, the effect of the previous response does not influence the current response. The a_i parameter represents the probability of $x_{i,j}$ taking on a value of 0, and therefore represents the probability of the previous response not influencing the current response. Thus, the a parameter represents a similar construct as the a parameter in the model described by Annis and Malmberg.

The priors for the model described above are listed below. I assume a flat prior for a_i.

$$a_i \sim Uniform(0,1) \tag{3.8}$$

I place normal priors on the participant and previous response distributions with an uninformative variance suggested by Morey et al. (2008).

$$\mu_S, r_P \sim Normal(0, 100) \tag{3.9}$$

I follow Spiegelhalter, Thomas, and Best (1996) and place uninformative inverse gamma priors, on the standard deviation components.

$$\sigma_S, \sigma_S^{(\beta)}, \sigma_S^{(\gamma)}, \sigma_P^{(\delta)} \sim InverseGamma(\,.001,.001) \qquad (3.10)$$

3.3.   Simulating Multi-Interval Response Data with a Markov Model

The Markov Model easily generalizes to procedures in which there are more than 2

stimulus classes. For example, in absolute identification tasks, the participant is shown m

different types of stimuli and asked to classify these stimuli into m categories. Thus, there will be

a total of m^2 unique states corresponding to all the different stimulus-response combinations

possible. A practical difficulty associated with this model is that it becomes difficult to specify

all the different transition probabilities by hand as the number of stimulus-response combinations

grow. For example, with only 3 stimuli and 3 responses, the transition matrix has 81 different

transition probabilities that need to be specified. Obviously, specifying all these transition

probabilities is cumbersome. One way this is made more tractable is to model each row of the

transition matrix with probability distributions. Here, I model the transition matrix with a series

of beta-binomial distributions – a discrete form of the beta distribution. This distribution was

chosen because of the wide variety of forms it can take on. The beta-binomial distribution has

rate parameter, α and shape parameter, β, which were used to control the transition probabilities.

As α becomes greater than β the probability of making a low response increases. When β is

greater than α the probability of making a high response increases.

Rows of Figure 3.2 show, for each previous stimulus value, the probability of making a

"1", "2", or "3" response given a particular previous response. For example, the first row of

Figure 3.2 shows, for each previous stimulus value, the probability of making a "1", "2", or "3"

response given the previous response was "1." Each column of Figure 3.2 shows, for each

previous response value, the probability of making a "1", "2", or "3" response given a particular

previous stimulus value. For example, the first column represents, for each previous response

value, the probability of making a particular response given the previous stimulus was 1. Thus, each panel of Figure 3.2 represents the response probabilities given a particular previous stimulus and response combination. For example, the top left panel of Figure 3.2 shows the probabilities of making a particular response given the previous stimulus was 1 and the previous response was also 1. The probability of responding with a "1" was around .3. The probability of responding with a "2" or a "3" was close to 0. These values were generated with a beta-binomial distribution with $\alpha = 1$ and $\beta = 20$. The same parameters were used to generate the response probabilities for each distribution of the first row. Therefore, the probability of responding "1" was high when the previous response was "1" regardless of the previous stimulus value. Similarly, responding with a "3" was high when the previous response was also "3." This is depicted by the bottom row of probability distributions. These distributions were generated by letting $\alpha = 20$ and $\beta = 1$. Thus, this parameter set generates response assimilation in the data. As I will show in the next sections, this model can generate a wide variety of sequential dependency patterns by modifying the parameters of the probability distributions.
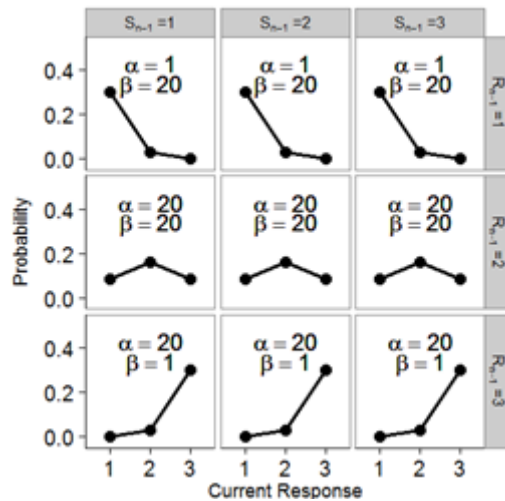
Figure 3.2. Graphical representation of a transition matrix used to simulate a data set with response assimilation.

Each row was normalized in order to maintain the condition that each row of the transition matrix sum to 1. The Markov model described above was implemented in R (see Appendix B). For each simulation, 200 chains each consisting of 500 time steps were generated.

3.4.   Tests of Multi-Interval Model R with Simulated Data

Figure 3.3 shows the results of a Markov chain simulation in which there is bias to respond high, but little to no sequential dependencies. The top left panel shows the parameter values used in order to simulate the data. The response probabilities were the same regardless of the previous stimulus-response combination. These probabilities were determined by a beta-binomial distribution with rate parameter, $\alpha = 1$ and shape parameter, $\beta = 5$. Under this distribution the probability of responding with a 3 is greater than the probability of responding with a 2, and the probability of responding with a 2 is greater than responding with a 1. Thus, there was a bias to respond with a high response and little to no sequential dependencies.

The top right panel shows accuracy plotted as a function of the current stimulus value. Accuracy increases with increases in stimulus value because of bias. The middle-left panel shows error on the current trial as a function of the previous response and current stimulus value. The absence of any positive slope indicates there are no sequential dependencies in the data. This is also true for the bottom left panel showing the relationship between error and the previous stimulus. When error is plotted as a function of the response at a given lag (middle right) or the stimulus at a given lag (bottom right), there is some noise in the data causing the graph to behave erratically; however, this is not too surprising when one considers the very small scale that the graphs are placed on.

Figure 3.4 shows the results of the model analysis using the multi-interval model described above. The top left panel plots β as a function of the current stimulus value. The lowest criterion,

C_1, was set at 0 (Morey, Pratte, & Rouder, 2008) and the highest criterion, C_2, was set at 1. The vertical dotted line in the top left panel of Figure 3.2 shows the highest criterion, C_2. For all stimuli values, the posterior density of β largely falls above the highest criterion, indicating a substantial amount of positive response bias. The top right panel shows the posterior mean of a for each simulated subject. This is the probability of the previous response *not* influencing the current response and roughly ranges from .72 to .87. This makes sense if we look closely at the response probabilities in the top left panel of Figure 3.2 There are sub-panels in which sequential dependencies are likely to occur. For example, there is a high probability of a "3" response when the previous response was 3 and the current stimulus value is 3. However, there is not an overall trend to assimilate towards the previous trial. This fact is illustrated in the bottom three panels which show the posterior densities of the overall effect of the previous response. The bottom left panel shows the posterior density for r_1 and the associated 95% Highest Density Interval (HDI) which contains 95% of area under the posterior probability curve and which follows the constraint that any density estimate inside the interval is greater than any density estimate outside the interval. The 95% HDI for r_1 is roughly centered at 0 and falls between -.69 and .60. Since 0 falls within the HDI, it is a credible value of r_1. Thus, a previous response of 1 did not have an effect on the current response. This is also true for previous responses of 2 and 3.

Given these results, the model successfully distinguished between response bias and sequential dependencies. The model was also able to capture the increase in bias to respond on the high end of the response scale. In addition, the a parameter was shown to reflect small trial-by-trial tendencies for the previous response to influence the current response. However, the model correctly identified the overall absence of sequential dependencies in the data as all r parameters were shown to have HDI's that included 0.

Figure 3.5 shows the results of a simulation in which there was a substantial amount of response assimilation. The top left panel shows the transition probabilities of the Markov chain that generated the data. When the previous response was "1" there was a high probability of the current response being "1" also. This assimilative behavior was similar when the previous response was "3." The right panel shows a bow shaped accuracy curve that is commonly observed in absolute identification tasks (Lacouture & Marley, 1995). The lowest and highest stimulus values had the highest accuracy, while the middle stimulus value had the lowest accuracy. Response assimilation is observed in the middle left panel; as the previous response increased so too did the error on the current trial. When the error on the current trial is plotted as a function of the response at a given lag, a decay in the magnitude of the error is observed with increases in lag.

The top right panel shows accuracy plotted as a function of the current stimulus value. Accuracy increases with increases in stimulus value because of bias. The middle-left panel shows error on the current trial as a function of the previous response and current stimulus value. The absence of any positive slope indicates there are no sequential dependencies in the data. This is also true for the bottom left panel showing the relationship between error and the previous stimulus. When error is plotted as a function of the response at a given lag (middle right) or the stimulus at a given lag (bottom right), there is some noise in the data causing the graph to behave erratically; however, this is not too surprising when one considers the very small scale that the graphs are placed on. For example, there is a high probability of a "3" response when the previous response was 3 and the current stimulus value is 3. However, there is not an overall trend to assimilate towards the previous trial.
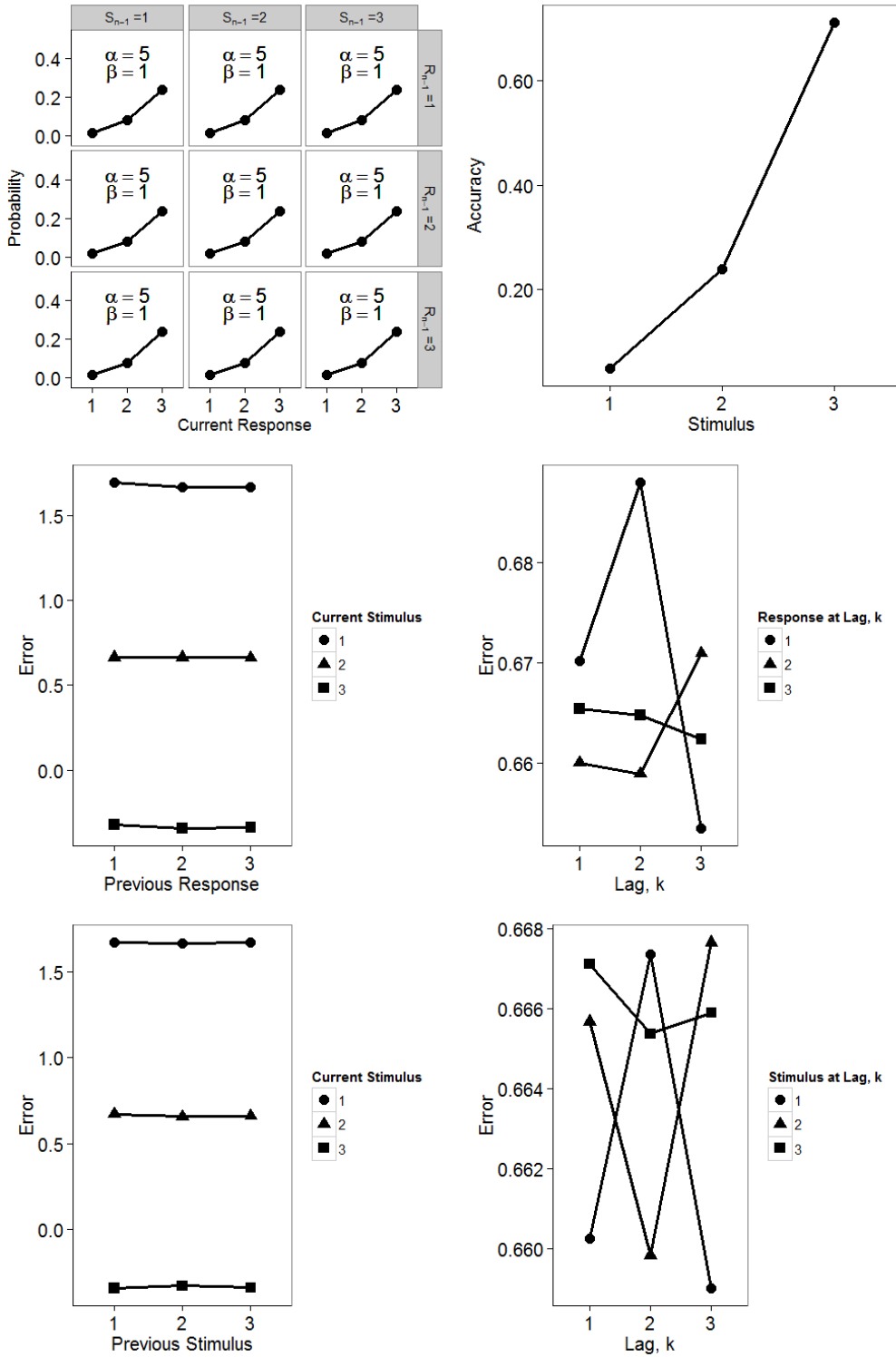
Figure 3.3. Simulated data with bias to respond high and very low sequential dependencies.
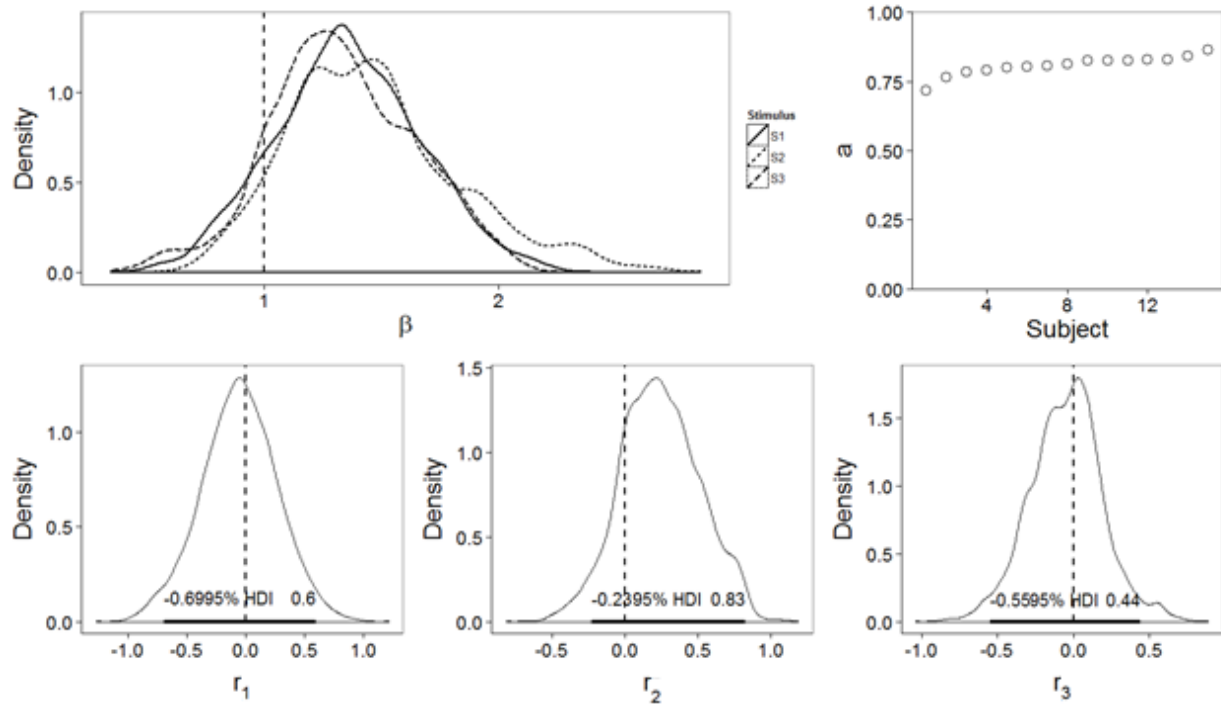
Figure 3.4. Parameters showing bias to respond high and no sequential dependencies.

However, the Markov model that generated the data does not explicitly take into account responses that occurred more than 1 trial back. Why then is a decay in assimilation being observed with increases in lag? It is simply due to the data being autocorrelated. For example, if a response of "3" is given on trial $n - 1$, then there is a high probability that the response on the current trial, n, will also be "3". Formally, this can be written as $P(R_n = 3 \mid R_{n-1} = 3)$. This probability will, on average, be lower than the probability of giving a "3" response on the current trial if a response of "3" is given on trial $n - 2$. This is due to the fact that there is a probability on trial $n - 1$ for a response other than "3" to be given. Thus, $P(R_n = 3 \mid R_{n-2} = 3) < P(R_n = 3 \mid R_{n-1} = 3)$.

This finding may be relevant for future theoretical development. For example, many models of sequential dependencies make the assumption that the representation of the current stimulus value is contaminated by previous stimulus/response representations and that as these

63

previous representations become more distant they become less relevant (Triesman & Williams, 1984; Stewart et al., 2005). This is said to lead to the pattern of decaying assimilation. The Markov model presented here demonstrates that it is possible to observe a pattern of decaying assimilation, while only taking into account the most recent trial.

The bottom left panel shows the error on the current trial as a function of the previous and current stimulus value. Note that there is no assimilation seen in the plot. In addition, when the error on the current trial is plotted as a function of the stimulus value at a given lag, no discernable pattern of assimilation is observed.

Figure 3.6 shows the results of the model analysis over the data presented in Figure 3.5. The top left panel shows the overall sensitivity associated with each stimulus value. The lowest criterion is indicated by the vertical dotted line at 0, while the highest criterion is located at 1. There is a tendency to respond with a low response, with most of the area centered between the two criteria. This may be due to the prior mean being set to 0. The top right panel shows the probability of the previous response not being factored into the current response, a. The mean probability of the previous response not being factored into the current response is near 0 for all simulated subjects.

The bottom row shows the estimates for the effect of the previous response. The bottom left panel shows the effect of the previous response on the current response when the previous response was "1." The 95% HDI ranged from -1.7 to -1.12 and therefore did not include 0. Using Kruschke's (2011) decision rule we can reject the null hypothesis that $r_1 = 0$. Thus, when the previous response was "1" there was a tendency to shift the current response negatively. The bottom middle panel shows the density over r_2 whose 95% HDI includes 0. Thus, when the previous response was "2" the current response, over all current stimuli, was unaffected. Finally,

the bottom right panel shows the density over r_3 whose 95% HDI ranges from 1.1 to 1.7. Since 0 falls outside of the 95% HDI this indicates that when the previous response was "3" there was a tendency to shift the current response positively. Thus, the model successfully detected the direction and magnitude of response assimilation in the data.

3.5. Multi-Interval Model SR: Stimulus/Response Sequential Dependencies

The previous model successfully detected assimilation and bias. In addition, it produced measures of the proportion of trials in which the previous response influenced the current response. It is possible to easily extend this model to account for sequential dependencies based on the previous stimulus. While the previous model only took into account the prior response, it is often useful to model the effect of the prior stimulus as well, especially when feedback is presented (Ward and Lockhead, 1970; Stewart and Matthews, 2009). In order to extend the model, the core of model, θ_{ij}, is modified to include the effect of the previous stimulus, $\eta_{i,Q}$, where the index, Q, represents the value of the previous stimulus. The effect is assumed to additive.

$$\theta_{ij} = \beta_S + \gamma_{i,S} + x_{i,j}(\delta_{i,P}) + z_{i,j}(\eta_{i,Q}) \tag{3.11}$$

The effect of the previous stimulus is drawn from a group-level normal distribution with mean, s_Q, and standard deviation, $\sigma_Q^{(\eta)}$.

In addition, it is assumed that the prior stimulus only effects the current response on a proportion of trials. When the previous stimulus effects the current response, $z_{i,j} = 1$, otherwise $z_{i,j} = 0$. It is assumed that $z_{i,j}$ follows a Bernoulli distribution with success rate, $1-a_i^{(s)}$. Thus, $a_i^{(s)}$ is the probability of the previous stimulus affecting the current response. The prior is assumed to be uniform on the interval from 0 to 1. The JAGS code for the model can be found in Appendix F.
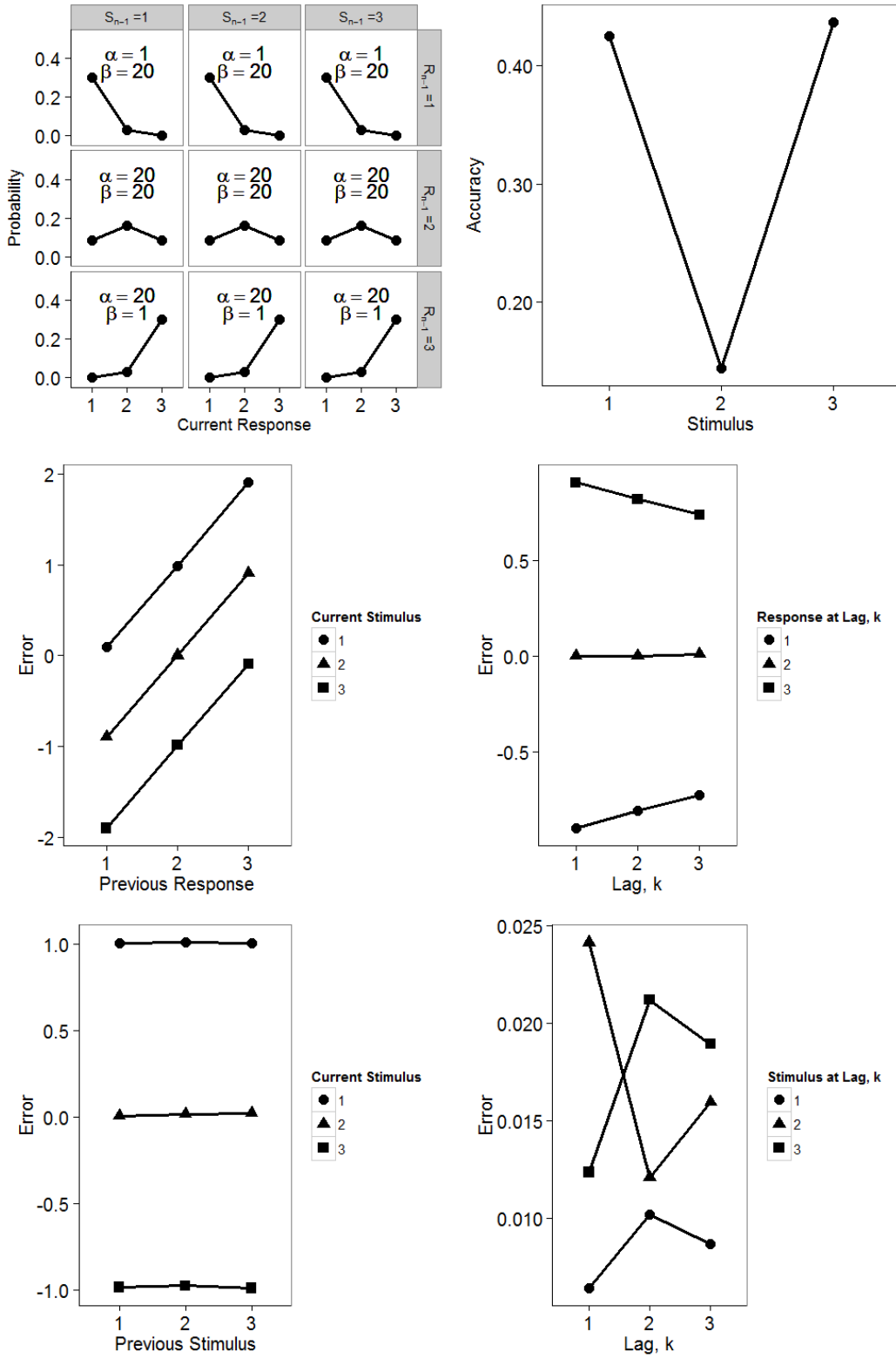
Figure 3.5. Simulated data with response assimilation.
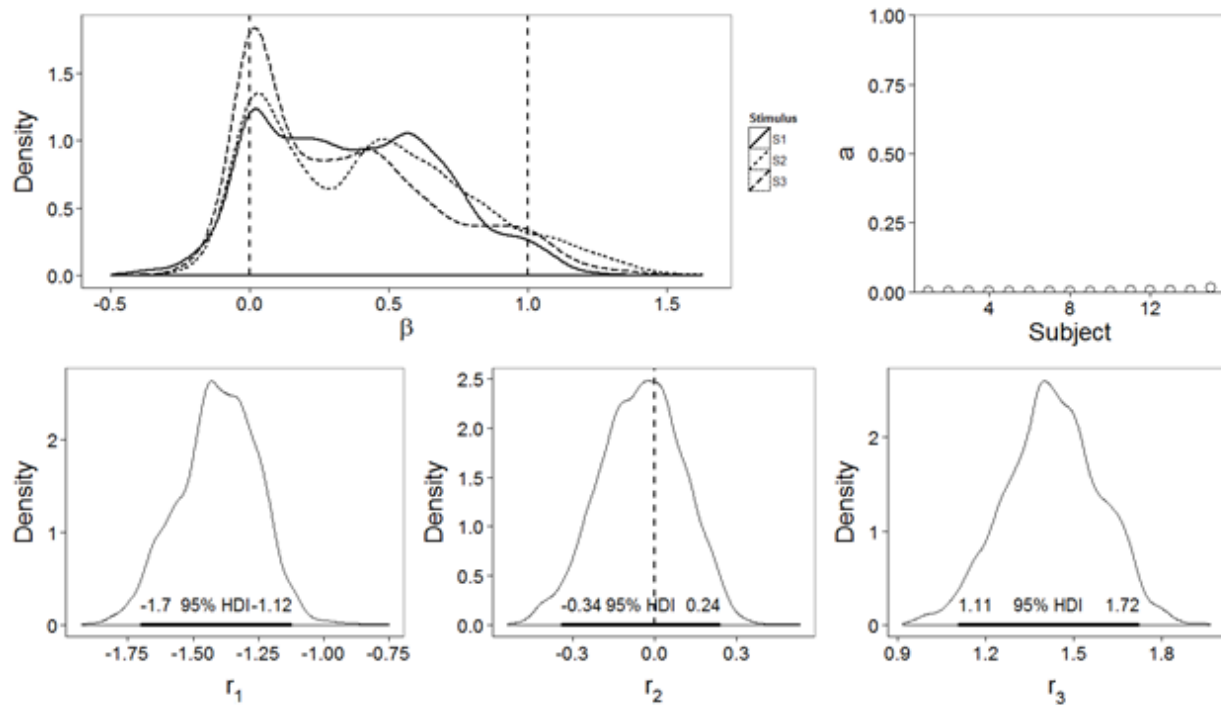
Figure 3.6. Estimated parameter values from data with response assimilation.

3.6.   Tests of Multi-Interval Model SR with Simulated Data

The top left panel of Figure 3.7 shows the transition probabilities for the Markov model

used to generate the data. There was a high probability of responding with a "1" given the

previous stimulus was also "1" compared to other previous stimulus values. There was a similar

increase in the probability to respond "3" given the previous stimulus was "3." The top right

panel shows a typical bow shaped curve in which the lowest and highest stimulus values

received the highest accuracy, while the middle stimulus value received the lowest accuracy. The

middle left panel shows the error on the current trial plotted as a function of the current stimulus

and the previous response. It is clear from the flat lines that there was no increase in the current

error with increases in the previous response. The absence of response assimilation is also

evident in the middle right panel. The graph appears noisy because of the very small scale. The

67

bottom left panel shows a clear pattern of stimulus assimilation; the error on the current trial increased with increases in the previous stimulus value. The bottom right panel shows the error on the current trial plotted as a function of the previous stimulus at a given lag. There is a clear pattern of assimilation at a lag of 1. That is, when the previous stimulus was 3, there was an overall tendency to overestimate the current stimulus value. When the previous stimulus was 1, the current stimulus was underestimated. The magnitude of this assimilative effect decays to 0 at lags greater than 1. The decay is much faster than the decay of response assimilation seen in the previous simulation (Figure 3.5). The reason for this is due to the fact that responses are no longer autocorrelated. Responses only depend on the previous stimulus and since each stimulus value has an equal probability of being presented on any given trial, correlations will only exist at lags of 1. For example, if a stimulus of 3 is presented on trial $n - 1$, there is a high probability of responding with a "3" on the next trial, n. However, if a stimulus of 3 is presented on trial $n - 2$, there is an equal probability of responding with a "1", "2", or "3" on trial n because all stimulus values have an equal probability of being presented on trial $n - 1$. Therefore the error goes to 0.

Figure 3.8 shows the analysis using the data generated above. The top left panel shows the sensitivity or mnemonic strength associated with each stimulus value. The top middle panel shows the probability of the previous response not influencing the current response for each subject and ranges from roughly .7 to .75. Given that there were 80 trials, the proportion of trials in which the previous response influenced the current response ranged from 20 to 24 trials. This is due to the fact that some previous responses were correlated with the current response. For example, Figure 3.5 shows the probability of responding with a "1" was high given the previous response was also "1." On the other hand, there is an equal probability to respond with a "3"

given the previous response was a "1". Therefore, the effect of the previous response should wash out when collapsing across all previous stimulus values. This prediction is verified in the middle row which depicts the estimated overall effect of the previous response. In all cases, 95% HDI included 0, and the distributions were generally centered at 0. Thus, the model successfully identified the absence of response assimilation in the data.

The top right panel shows the mean probability of the previous response influencing the current response was close to 1 for each subject. The direction and magnitude of the effect of each stimulus value is estimated on the bottom row. The bottom left panel shows the effect when the previous stimulus value was 1, s_1, ranged from -1.77 to -1.08. This indicates that previous stimulus values of 1 caused the current response to shift negatively over all current stimulus values. On the other hand, when the previous stimulus value was 3, there was a tendency for the current response to shift in a positive direction. Thus, the model successfully detected the presence of stimulus assimilation in the data.

The top left panel of Figure 3.7 shows the transition probabilities for the Markov model used to generate the data. There was a high probability of responding with a "1" given the previous stimulus was also "1" compared to other previous stimulus values. There was a similar increase in the probability to respond "3" given the previous stimulus was "3." The top right panel shows a typical bow shaped curve in which the lowest and highest stimulus values received the highest accuracy, while the middle stimulus value received the lowest accuracy. The middle left panel shows the error on the current trial plotted as a function of the current stimulus and the previous response. It is clear from the flat lines that there was no increase in the current error with increases in the previous response. The absence of response assimilation is also evident in the middle right panel. The graph appears noisy because of the very small scale.

Figure 3.7. Simulated data with stimulus assimilation.
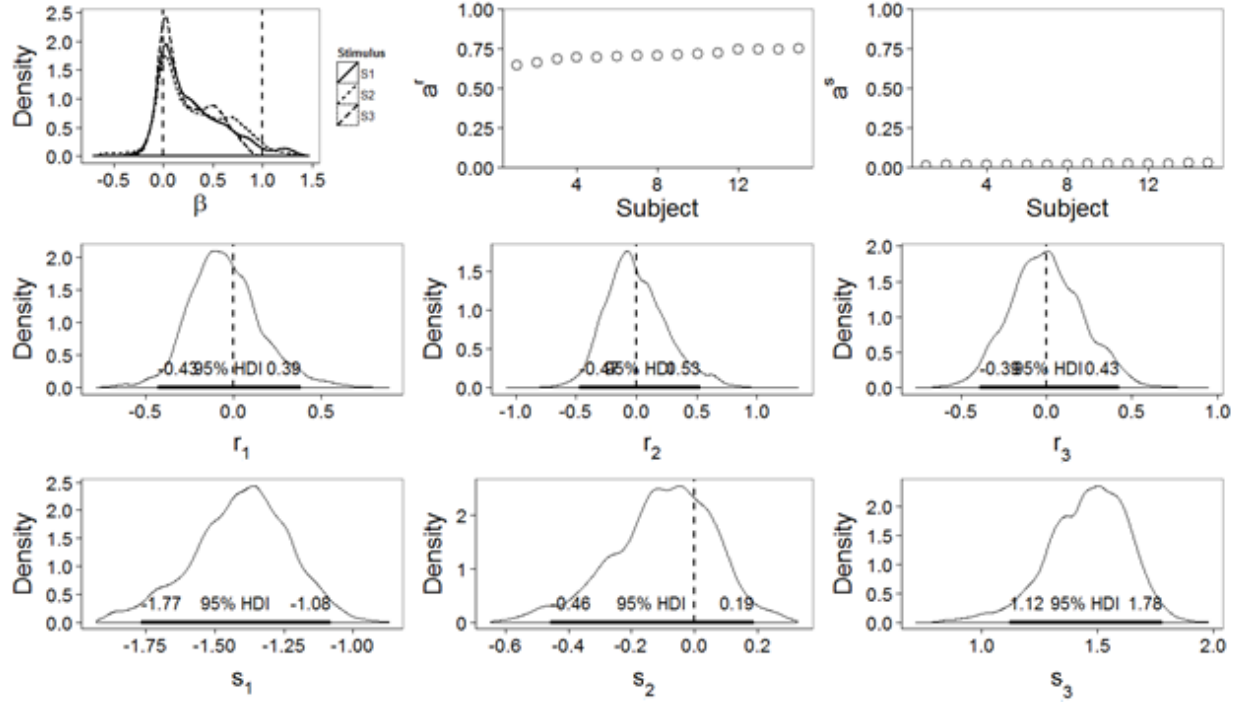
Figure 3.8. Estimated parameter values from data with stimulus assimilation.

3.7. Multi-Interval Model SRLAG: A Model for Lags ≥ 1

The prior models successfully accounted for both stimulus and response sequential dependencies in addition to measuring bias and sensitivity. In this section, the multi-interval model is extended to account for sequential dependencies that might occur at lags greater than 1. The ability to account for these types of sequential dependencies becomes especially important in absolute identification tasks in which contrast at lags greater than 1 is observed, usually in the presence of feedback (e.g. Ward & Lockhead, 1970). In addition, there is some evidence that stimuli occurring at lags of up to four may have an influence on the EEG associated with the current trial (Squires, Wickens, Squires, Donchin, 1976). Therefore, the core of the model, θ_{ij}, is again modified by adding an effect component for responses at lags of n, $\delta_{i,P_{j-n}}$, and stimuli at lags of n, $\eta_{i,Q_{j-n}}$, where $n = \{1, \ldots, \ell\}$ and ℓ is the maximum lag considered.

$$\theta_{ij} = \beta_S + \gamma_{i,S} + \sum_{n=1}^{\ell} x_{i,j}^{(n)} \left(\delta_{i,P_{j-n}} \right) + \sum_{n=1}^{\ell} z_{i,j}^{(n)} (\eta_{i,Q_{j-n}}). \qquad (3.12)$$

The effect of the prior response at lag n is drawn from a group-level normal distribution with

mean, $r_{P_{j-n}}$, and standard deviation, $\sigma_{P_{j-n}}^{(\delta)}$. Likewise, the effect of the previous stimulus is drawn

from a group-level normal distribution with mean, $s_{Q_{j-n}}$, and standard deviation, $\sigma_{Q_{j-n}}^{(\eta)}$. It is

assumed prior responses do not always influence the current response. When $x_{i,j}^{(n)}$ is 1, responses

at lag n are taken into account, otherwise the response at lag n is ignored. $x_{i,j}^{(n)}$ is assumed to

follow a Bernoulli distribution with success rate, $1 - a_{i,n}^{(r)}$. Similarly, the effect of the prior

stimulus is not always taken into account on each trial. When $z_{i,j}^{(n)} = 1$ the prior stimulus at lag

n, is taken into account, otherwise the information is discounted. $z_{i,j}^{(n)}$ follows a Bernoulli

distribution and has success rate, $1 - a_{i,n}^{(s)}$. The priors for all variables just described are the same

as in the previous models. The JAGS code for the model can be found in Appendix G.

3.8. Tests of Multi-Interval Model SRLAG

The top left panel of Figure 3.9 shows the probability of the current response equaling

"3" was highest when the previous stimulus was 3 and the previous response was "1." In

addition, the probability of responding with a "1" was highest when the previous response was

"3" and the previous stimulus was 1. This resulted in positive sequential dependencies between

the current response and previous stimulus, and negative sequential dependencies between the

current response and previous response. The negative dependency between the current response

and previous response is graphically depicted in the middle left panel. The positive dependency

between the current response and previous stimulus is shown in the bottom left panel.

Interestingly, this pattern of responding resulted in contrast. The middle right panel shows the

negative dependency between the current response and previous response switches at lags of 2 and becomes a positive dependency. The bottom right panel shows the positive dependency between the current response and previous stimulus also switches at lags of 2 and becomes a negative dependency. This is due to an oscillatory response pattern. Table 3.1 illustrates this pattern.

Table 3.1. Positive dependencies at lags of 1.

Trial	n	n-1
Stimulus	X	3
Response	3	1
Error	+	

Consider a stimulus of 3 appears on trial n - 1, and the model responds with a "1." Given the previous response was a "1" and the previous stimulus was a 3, on the next trial, n, the most likely response according to the model is a "3." Thus, a positive dependency is created at lags of 1 between the current response and previous stimulus. Continuing with this example, Table 3.2 illustrates how the sign of the dependency switches at lags of 2. Consider a response of "1" and a stimulus of 3 occurring on trial n - 2.

Table 3.2. Negative dependencies at lags of 2.

Trial	n	n-1	n-2
Stimulus	X	1	3
Response	1	3	1
Error	-		

Given the response on trial n - 2 was a "1" and the stimulus was a 3, according to the model, there is an increased probability to respond with a "3" on trial $n - 1$ when the stimulus on trial $n - 1$ is a 1. This in turn increases the probability of a "1" response on trial n. Thus, the

overall error on the current trial will be negative given the response on trial $n - 2$ was a "1" and the stimulus was a 3. This is exactly what is depicted in the middle-left and bottom-left panels of Figure 3.9. The middle left panel shows that the error on the current trial will be negative given the response at lag 2 was "1." The bottom left panel shows the error will be negative given the stimulus on trial $n - 2$ was 3. Thus, the seemingly complex pattern of results depicted in Figure 3.9 is reduced to two simple rules:

1. If the previous response was "3" and previous stimulus was 1, respond with a "1."

2. If the previous response was "1" and the previous stimulus was 3, respond with a "3."

Therefore, it is not necessary to assume responses or stimuli more than 1 trial back influence the current response if contrast is present in the data. That is, dependencies at lags greater than 1 may not always indicate a direct influence of the stimulus or response at that given lag. This is important for many models of sequential dependencies which explicitly take into account information at lags greater than 1 (e.g. Stewart, Brown, Chater, 2005; Brown, Marley, Donkin, & Heathcote, 2008).

Figure 3.10 shows the results of the model analysis over the data depicted in Figure 3.9 in which there were positive sequential dependencies between the current response and previous stimulus and negative dependencies between the current response and previous response. The top left panel shows the posterior probability density for β which can be conceived of as a representation of the stimulus strength. The vertical dotted lines show the decision criteria. The middle panel in the top row shows the probability of the previous response not influencing the current response for each simulated subject. This probability was close to 0 for each subject. The right panel in the top row shows the probability of the response on trial $n - 2$, not influencing the current response. This probability ranged from approximately .35 to .75.
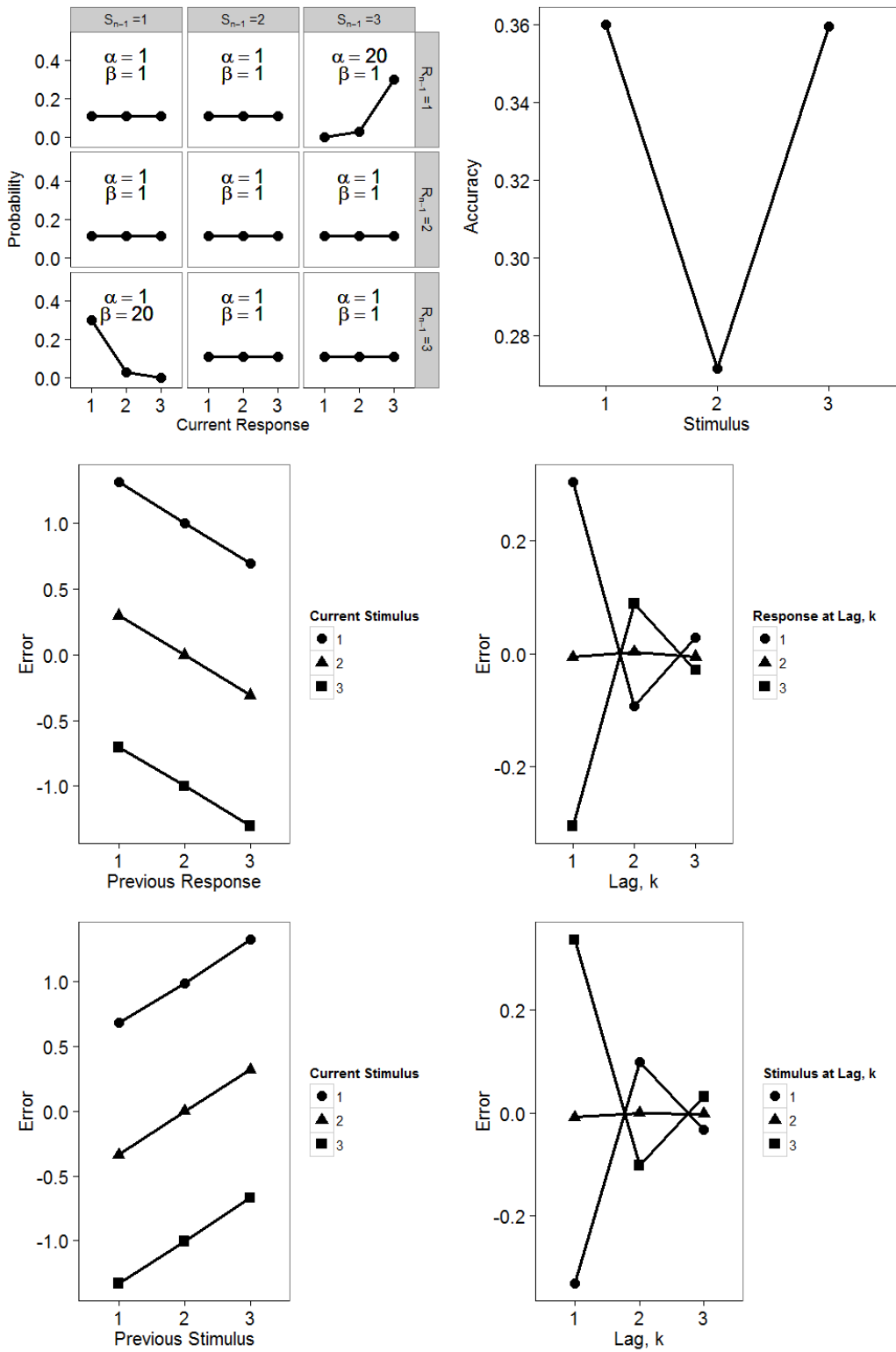
Figure 3.9. Result of simulation in which there was positive dependencies between the current response and previous stimulus, and negative dependencies between the current response and previous response.

The middle row shows the magnitude and direction of the effect for each previous response type. When the previous response was equal to 1, the overall effect on the current response was positive. When the previous response was equal to 3, the effect on the current response was negative. Thus, the model successfully detected the negative sequential dependencies between the current response and previous response.

The bottom rows show the direction and magnitude of the effect of the response at lags of 2. When the response on trial $n - 2$ was "1" there was a slight negative effect on the current response, however, 0 lies within the 95% HDI. For responses of "3" on trial $n - 2$, there was a slight positive effect, but again, 0 lies within the HDI.
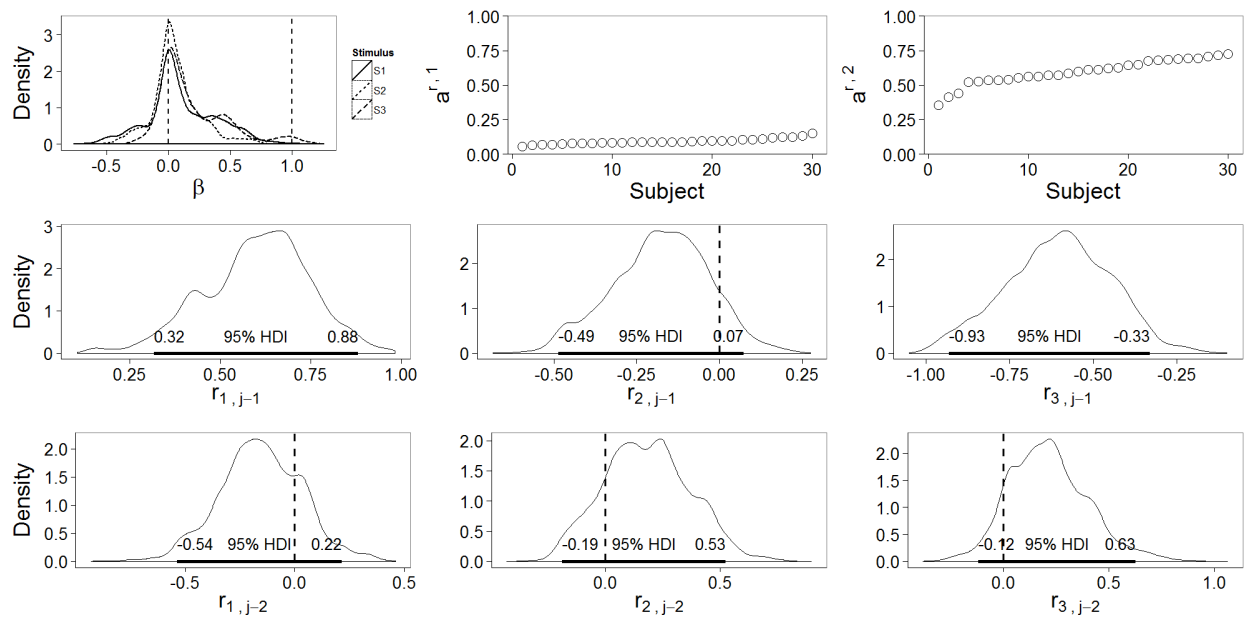
Figure 3.10. Posterior probability estimates over data in which the current response was negatively correlated with the previous response and positively correlated with the previous stimulus.

Figure 3.11 shows the estimated effects of the previous stimulus. The top left panel is the same as in Figure 3.10 and is there for reference. The top middle panel shows the probability of the current response being influenced by the previous stimulus. This probability was high for all subjects. The probability of the stimuli at lags of 2 influencing the current response showed a

76

decreased effect. The middle row shows the effect of the previous stimulus on the current response. When the previous stimulus was 1, there was an overall negative effect on the current response. When the previous stimulus was 3, the model shows a positive effect on the current response. The bottom panels show the effect of the stimulus at lags greater than 1 was neutral. In summary, the model was able to correctly detect the negative dependency between the previous response and current response and the positive dependency between the previous stimulus and current response. Although the data were generated by considering only the most recent trial, the model indicated that there was approximately a 50% chance the response/stimulus on trials $n - 2$ would influence the current response. This is due to low accuracy in the data. Because the response on the current trial is not influenced by the current stimulus value, the model is able to predict the current response based on the response on trial n - 2. To test the model under more real world conditions, the next simulation shown in Figure 3.12 took accuracy into account.
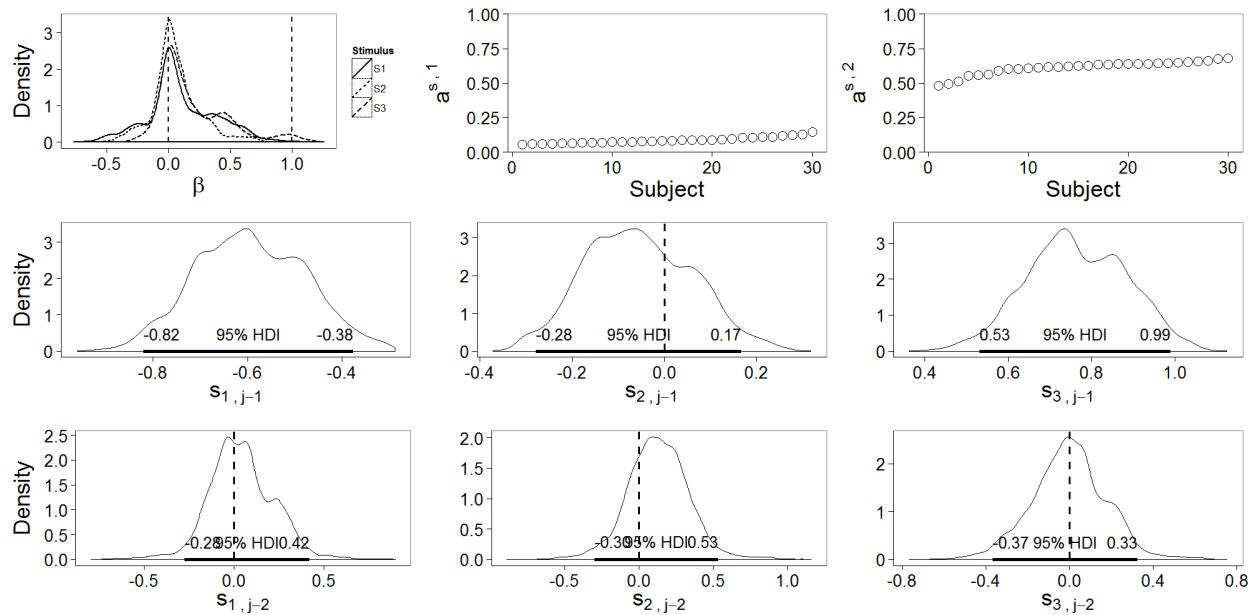
Figure 3.11. Posterior probability estimates over data in which the current response was negatively correlated with the previous response and positively correlated with the previous stimulus.

The simulation details were very similar to the previous simulation in which there was a high probability to respond with a "3" given the previous response was "1" and the previous stimulus was 3. In addition, there was a high probability to respond "1" given the previous response was "3" and the previous stimulus was 1. In order to increase accuracy, a constant value was added to the appropriate cells in the transition matrix described in section 3.4. Thus, the data were generated from a Markov chain in which the previous response was negatively correlated with current response, while the previous stimulus was positively correlated with the current response. In addition, relatively high accuracy caused the previous stimulus and previous response to be correlated with one another. However, upon inspection of Figure 3.12 we find the standard pattern of results for absolute identification experiments in which feedback is provided (Malmberg & Annis, 2013; Ward & Lockhead, 1971); it appears as though there was a slight amount of response assimilation, a large amount of stimulus assimilation, and that the stimuli and responses at lags of 2 and greater were negatively correlated with the current response. Thus, these measurements, as they are commonly interpreted, grossly mischaracterize the underlying model used to generate the data.

Figure 3.13 shows the results of the model analysis of the data depicted in Figure 3.12. The top left panel shows the posterior probability for β which can be conceptualized as the stimulus strength for each stimulus type. The dotted vertical lines show the decision criteria which clearly separate the posterior probability estimates associated with each stimulus type. Thus, accuracy was much higher in the current simulation relative to the previous simulations described above. The middle panel in the top row shows the probability of the previous response not influencing the current response and ranged from .25 to .75. The right panel of the top row shows the probability of the response on trial $n - 2$ not influencing the current response for each

subject. This probability was close to 1 for all subjects. Thus, the model correctly identified the influence of the previous response and the absence of influence of the response on trial $n - 2$.

The middle row of Figure 3.13 shows the magnitude and direction of the effect of the previous response. When the previous response was "1" there was a slight positive effect on the current response. When the previous response was "3" there was a negative effect on the current response. Thus, the model correctly identified the negative dependency between the current response and the previous response. This result is contrasted with the middle left panel of Figure 3.10 in which a clear pattern of response assimilation is depicted. In addition, the bottom panel of Figure 3.13 shows no effect of the response on trial $n - 2$ on the current response, while the middle right panel of Figure 3.10 shows contrast at lags greater than 1.

Figure 3.13 shows the results of the model analysis for the previous stimuli. The middle panel in the top row shows the probability that the previous stimulus will not influence the current response is close to 0 for all subjects. The left panel in the top row shows the probability that the stimulus on trial $n - 2$ will not influence the current response is close 1. Thus, the model successfully detected the influence of the previous stimulus on the current response, and correctly identified the absence of the effect of stimuli at lags greater than 1.

The middle row shows the magnitude and direction of the effect of the previous stimulus. When the previous stimulus was 1, there was an overall positive effect on the current response. When the previous stimulus was 3, the effect of the previous stimulus on the current response was positive. The bottom row shows the effect of stimuli two trials back did not have an overall effect on the current response. Thus, the model correctly identified the direction of the effect of the previous stimulus and the absence of any effect of the stimuli at lags greater than 1.

In summary, the standard plots in Figure 3.12 proved to be insufficient for accurately describing the actual pattern of sequential dependencies present in the data, while the Bayesian model was able to correctly identify all of the sequential dependencies. The standard plots showed both response and stimulus assimilation when, in fact, there were negative sequential dependencies between the current response and previous response. The model, on the other hand, successfully detected this. In addition, the standard plots may lead to researcher to conclude the stimulus/response at lags greater than 1 may influence the current response, while the data may have been generated by a system that only takes into account the most recent trial.
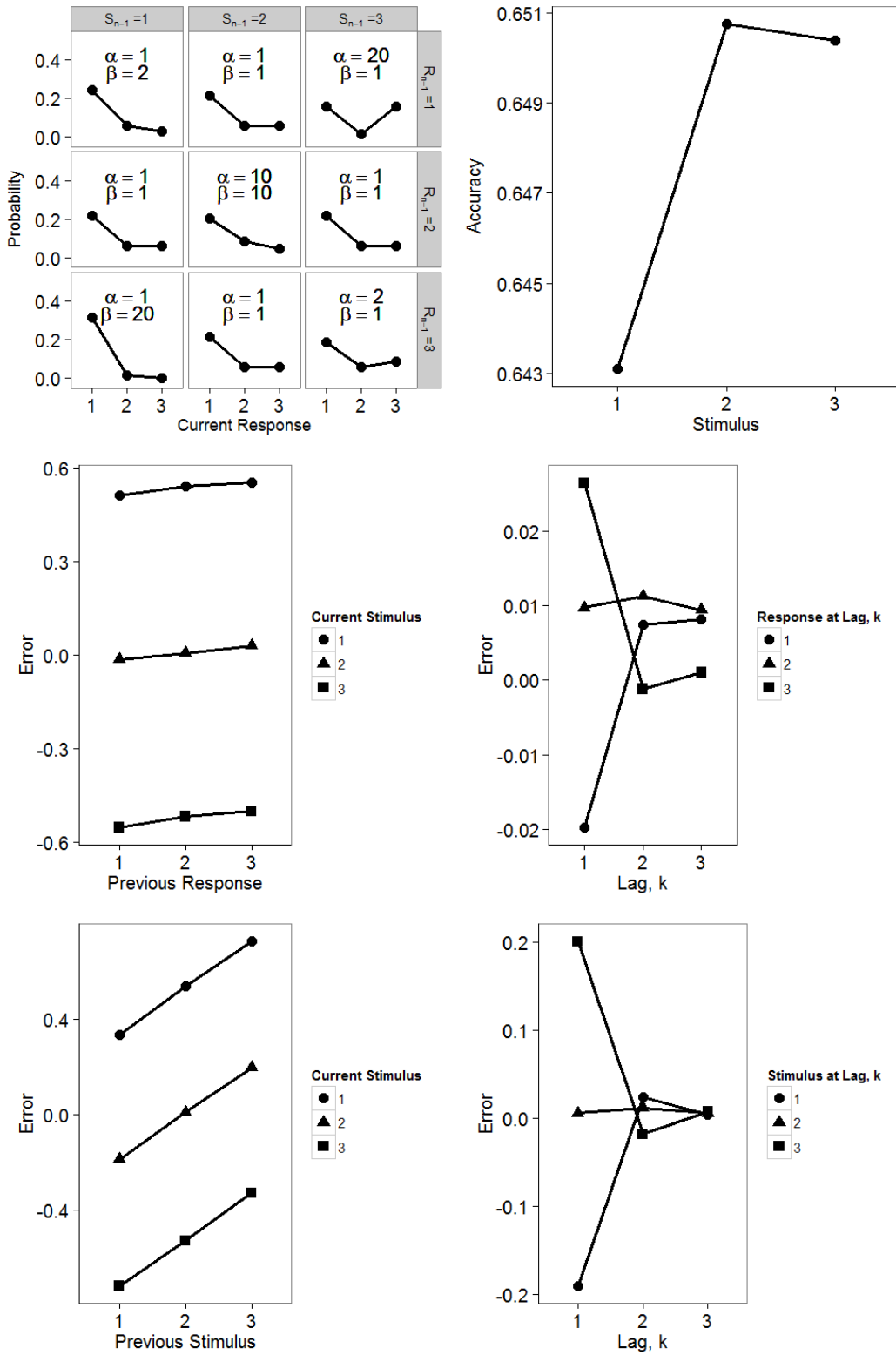
Figure 3.12. Data generated by a Markov chain in there was tendency to respond towards the previous stimulus and away from the previous response. In addition, high accuracy caused the previous stimulus to be correlated with the previous response.
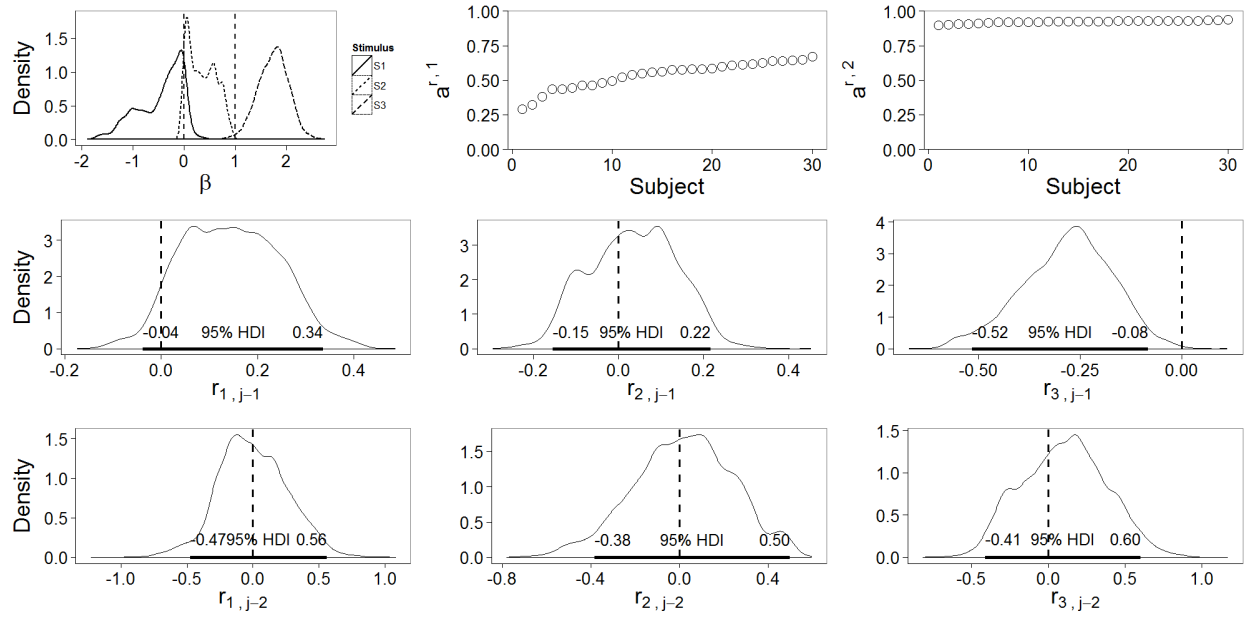
Figure 3.13. Analysis of data containing negative sequential dependencies with the previous response and positive dependencies towards the previous stimulus.
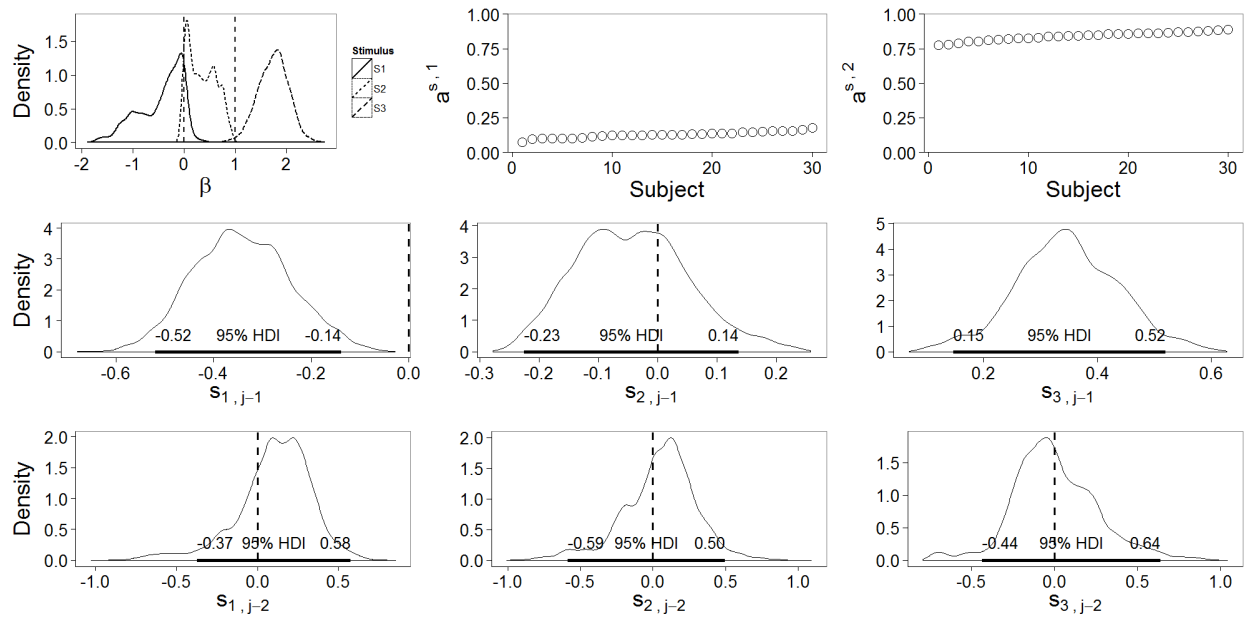
Figure 3.14. Analysis of data containing positive sequential dependencies towards the previous stimulus and negative dependencies with the previous response.

Chapter 4

A Bayesian Analysis of The Effect of Inter-trial Tasks on Sequential Dependencies in

Yes/No Recognition

Recognition memory models have traditionally assumed independence among responses during testing (e.g. Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1982; Shiffrin & Styevers, 1997; also see Chapter 1). This assumption is known as the *independence assumption*. When this assumption is violated, a sequential dependency is said to have occurred. For example, a positive correlation between the current response and previous response is known as *assimilation*. Assimilation has been observed in perception tasks including absolute identification (Stewart, Brown, & Chater 2005) categorization (Jones, Love, & Maddox, 2006), and perceptual detection (Howarth & Bulmer, 1956).

Assimilation has recently been observed in a variety of recognition tasks including yes/no recognition, confidence ratings, and judgments of frequency (Malmberg & Annis, 2012; Annis & Malmberg, 2013). Specifically, Malmberg and Annis (2012) observed the probability of a false alarm is greater when following a false alarm than when following a correct rejection, the probability of a hit is greater when preceded by a hit than by a miss, and the probability of hit is greater when followed by high confidence responses than when followed by low confidence responses.

Annis and Malmberg (2013) modeled assimilation in recognition within the Retrieving Effectively from Memory framework (REM; Shiffrin & Steyvers, 1997). (For an in-depth

explanation of the model see section 1.7.) REM represents memory traces as vectors of feature

values that are assumed to be geometrically distributed. When an item is studied, REM assumes

that a noisy and incomplete representation of the studied item is stored in episodic memory.

When an item is presented during a recognition test trial, the subject generates a representation of

that test item called the *retrieval cue*. The retrieval cue is then compared to an activated set of

memory traces stored in episodic memory. The more similar the retrieval cue is to the contents

stored in memory, the more likely the subject will endorse the test item as being studied.

In order to model assimilation, Annis and Malmberg (2013) assumed that on each

recognition test trial, there was a probability that features from the previous retrieval cue would

carry over to current retrieval cue. The carryover model assumes on each trial there is a

probability, *a*, that carryover will occur. This mechanism was hypothesized to be associated with

vigilance during testing and the nature of the stimuli affected this parameter. For example, Annis

and Malmberg (2013) found that when participants were presented with similar stimuli during a

recognition test, the *a* parameter was higher than when the stimuli were only randomly similar

(c.f. DeCarlo, 2002, 2007; Maddox & Estes, 1997; Howard, Bessette-Symons, Zhang, Hoyer,

2006; Malmberg & Murnane, 2002). Thus, it might be possible to decrease sequential

dependencies by manipulating attention at test. One way that this might be achieved is through a

task switching procedure. There is some evidence that attentional demands increase when the

participant must alternate between two tasks rather versus repeating the same task on each trial

(for a review see Monsell, 2004). Thus, I carried out an experiment in which either a lexical

decision task or a blank screen was inserted between recognition test trials. If the lexical decision

task does indeed increase attentional vigilance, this should result in decreased sequential

dependencies compared to the condition in which there is no task inserted between recognition test trials.

4.1. Experiment 1

Annis & Malmberg's model of sequential dependencies in recognition memory testing predicts that sequential dependencies should be reduced or eliminated when carryover is reduced or eliminated. According the model, reductions in vigilance may be related the tendency to carryover information from one trial to the next. In order to test the carryover model predictions, a yes/no recognition task was used in which the interpolated task was varied in two conditions. The assumption is that task switching involving tasks that rely on the use of different information should make carryover less likely. In the first condition, after each recognition test, participants were presented with a lexical decision task. In the second condition, a blank ISI was presented following each recognition test trial. Sequential dependencies should be higher in the condition in which no task was presented after each recognition test trial.

4.1.1. Method

4.1.1.1. Participants. Sixty-two undergraduate students from the University of South Florida participated in exchange for course credit.

4.1.1.2. Design and Materials. Participants completed one study-test test cycle of each inter-trial task condition. The inter-trial condition involved either a lexical decision (LD) task interpolated between each test trial or a blank ISI following each test trial. Thus, the inter-trial task was manipulated within subjects and between lists. Each study list was composed of 80 words from the Kucera and Francis (1983) word pool with normative frequencies between 20 and 50 occurrences per million. The test list was composed of 80 words from the study list and 80 foils. The LD trials contained 80 words, different from

those used in the test list, drawn from the Kucera and Francis (1983) word pool and 79

non-words.

4.1.1.3. Procedure. Immediately following the instructions of the experiment,

participants were presented with a study list in which words were presented for .75 s each

at the center of the screen. After each word, a .15 s blank ISI followed. After the study

list was presented, a 30 s math task was performed. During the math task an integer

ranging from 1 to 9 was presented on the center of the screen for 3 s. The task was to add

each integer to the previous sum. Following the math task, the test list was presented.

For each yes/no recognition trial, participants were presented with either a target or foil.

The task was to indicate whether the word was studied by typing a "1" or not studied by typing a

"0." Following each yes/no recognition test trial, there was an inter-trial interval. For the lexical

decision condition, a letter string was presented at the center of the screen during the inter-trial

interval. The task of the participant was to respond "1" if the letter string was an English word

and "0" if the letter string was a non-word. For the condition in which a blank ISI was presented,

no task was specified.

4.1.2. Results

4.1.2.1. Accuracy. Hit rates were higher when lexical decision trials were interpolated at

test ($M = .65$, $SD = .16$) than when a blank ISI followed each recognition test trial ($M =$

.59, $SD = .16$), $t(61) = 3.66$, $p < .01$. False alarm rates were also higher in the lexical

decision condition ($M = .41$, $SD = .20$) than in the blank ISI condition ($M = .27$, $SD =$

.16), $t(61) = 6.57$, $p < .01$. Thus, subjects responded "old" more frequently in the lexical

decision condition. Subject's accuracy measured by d' tended to be lower in lexical

decision condition ($M = 1.23$, $SD = .99$) than the blank ISI condition, ($M = 1.52$, $SD =$

1.07), however, this difference was not reliable, $t(61) = -1.82$, $p = .074$. However, since bias and d' are independent for recognition memory, it is quite possible that lower d' in the lexical decision condition is due to the use a more lax decision criterion.

4.1.2.2. Sequential Dependencies. A 2 (Task: lexical decision vs. blank ISI) x 2 (Previous Response: "Old" vs. "New") repeated measure ANOVA revealed a significant main effect of Task, $F(1,61) = 35.32$, $p < .01$, on the probability to respond "Old". This reflects an overall tendency to respond "old" in the lexical decision condition as described in the previous section. There was a significant main effect of Previous Response on the probability to respond "Old", $F(1,61) = 30.29$, $p < .01$, such that the probability of an "Old" responds was greater when following a "Old" response than when following a "New" response. However, the main effect of the Previous Response is qualified by a significant Task x Previous Response interaction, $F(1,61) = 15.12$, $p < .01$. In the blank ISI condition, the probability of an "Old" response was greater when following a "Old" response ($M = .49$, $SD = .13$) than when following a "New" response ($M = .38$, $SD = .13$), $t(61) = 6.87$, $p < .01$. In the lexical decision condition, the probability of an "Old" response did not significantly differ when preceded by an "Old" response ($M = .54$, $SD = .16$) or a "New" response ($M = .51$, $SD = .17$), $t(61) = 1.97$, $p = .053$. These results suggest the effect of interpolating lexical decision trials between recognition trials is twofold; there is an increase in bias to respond "old" and the effect of the previous response is reduced. The later finding is consistent with the Annis & Malmberg model, but additional analyses based on the model that I have developed for independently measuring bias and sequential dependencies are obviously necessary to make any strong conclusions based on these data.

4.1.2.3. Bayesian Analyses. The model described in Section 2.8 was used to evaluate the

differences in sequential dependencies across conditions. The model was implement

using the open-source JAGS softeware (Plummer, 2003). To run this model, three chains,

each consisting of 53,000 samples, were generated. The chains were visually checked for

convergence and the first 3000 burn-in samples were discarded. Table 4.1 shows the

results of the analysis.

Table 4.1. Bayesian analysis of differences in sequential dependencies between conditions. BF_{10} is the Bayes Factor for the effect size. Bayes Factors greater than 1 indicate support for the hypothesis that the effect size is greater than 1. Bayes Factors less than 1 indicate evidence in favor of the null hypothesis that the effect size is 0. The lower and upper bounds of the 95% HDI for the differences in sequential dependencies between conditions is also shown.

Comparison	BF_{10}	μ_α 95% HDI	
		Lower Bound	Upper Bound
Experiment 1			
Blank ISI - LD	1.32	-0.29	-0.08
Experiment 2			
Blank ISI - LD	0.05	-0.21	-0.02

The 95% HDI for the differences in sequential dependencies between the Lexical

Decision condition and the Blank ISI condition, μ_α, does not include 0. Following Kruschke

(2011), we can reject the null hypothesis that there is no difference in sequential dependencies

between conditions, on the grounds that 0 is not a credible value of μ_α. This result is in line with

the frequentist analysis in which a significant Task x Previous Response interaction was found.

The 95% HDI is lies on a negative interval, indicating that sequential dependencies were reduced

in the Lexical Decision condition. In addition to computing the HDI for the difference between

conditions, the Bayes Factor for the effect size was also computed. The height of the posterior

distribution over zero of the effect size, δ, and the height of prior distribution over zero are

computed. The ratio of the height of the posterior density and prior density over zero gives the

Bayes Factor, of the effect size, δ, at zero. The Bayes Factor for the effect size is small, 1.32.

This indicates that the alternative hypothesis, $\delta_\alpha \neq 0$, is only 1.32 times as likely as the null

hypothesis. Taken together, these results indicate that interpolating lexical decision trials does

indeed cause a reduction in sequential dependencies, but the effect of this phenomenon is small,

which makes sense, in that the effect of sequential dependencies is already a small effect.

Therefore, any reduction in the already small effect would in itself be small.

I next analyzed the differences in bias between the Lexical Decision and Blank ISI

conditions using the model described in Section 2.8 above. Table 4.2 shows the results.

Table 4.2. Bayesian analysis of the differences in bias between conditions.

Comparison	BF$_{10}$	μ_α 95% HDI	
		Lower Bound	Upper Bound
Experiment 1			
Blank ISI - LD	147.57	0.16	0.35
Experiment 2			
Blank ISI - LD	2.02E+07	0.16	0.31

The 95% HDI for the differences in bias between conditions fell between .16 and .35.

Zero was not found to be a credible value of μ_β indicating that there was an increase in the

probability of responding "Old" in the Lexical Decision condition. This is in line with the

ANOVA in which there was a significant effect of Task. The Bayes Factor for the effect size

parameter, δ_β, indicates that the alternative hypothesis, $\delta_\beta \neq 0$ is roughly 148 times as likely as

the null. These results indicate interpolating lexical decision trials during recognition testing

increases the probability of an "Old" response and given the Bayes Factor, the effect is

"decisive" (Jeffreys, 1961).

Having analyzed the differences in bias between conditions, the next analysis looks at

sequential dependencies in each condition. Table 4.3 shows the results of the analysis using the

model described in Section 2.7. The first row shows the results for the Blank ISI condition. Zero

did not lie within the 95% HDI of μ_α and the Bayes Factor for the effect size indicated a

substantial effect of the previous response. This is contrasted with the Lexical Decision condition

in which 0 lies only slightly outside of the 95% HDI. The null hypothesis that the effect size,

$\delta_\beta = 0$, was 25 times more likely than the alternative hypothesis. These results suggest that there

was a strong effect of the previous response in the Blank ISI condition, but little if any effect of

the previous response in the Lexical Decision condition.

Table 4.3. Bayesian analysis of sequential dependencies for each condition.

Experiment	BF$_{10}$	μ_α 95% HDI	
		Lower Bound	Upper Bound
1			
Blank ISI	2804.73	0.22	0.39
Lexical Decision	0.04	0.01	0.20
2			
Blank ISI	20.27	0.22	0.38
Lexical Decision	193.07	0.10	0.25

Table 4.4. Bayesian analysis of bias in each condition.

Experiment	BF$_{10}$	μ_α 95% HDI	
		Lower Bound	Upper Bound
1			
Blank ISI	28.91	-0.25	-0.08
Lexical Decision	0.05	-0.02	0.23
2			
Blank ISI	0.14	-0.05	9.22
Lexical Decision	492212.20	0.25	0.49

Table 4.4 shows the results of the Bayesian analysis of bias for each separate condition.

There was a bias to respond "New" in the Blank ISI condition indicated by the 95% HDI of μ_α

not including 0. In addition, there was strong evidence for the effect size being greater than 0. In

the Lexical Decision condition, zero was a credible value of μ_α. The Bayes Factor for the effect size was in favor of the null hypothesis.

4.1.2.4. Modifying the Multi-Interval Model. In addition to modeling the data with the binary model described in Chapter 2, these data can also be modeled using the multi-interval model described in Section 3.2. The multi-interval model can be applied to tasks in which there are *n* responses by assuming *n* - 1 criterion locations, where *n* > 1. Therefore, for binary data the model would assume a single criterion. Following Morey et al. (2008) the criterion was fixed at zero.

The results are shown in Figures 4.1 through 4.3. Figure 4.1 shows the group-level estimates for the proportion of trials in which the previous response did *not* influence the current response. This parameter is referred to as *a* in the model. The 95% HDI of the differences ranges from -.08 to .12. Given that zero is a credible value of the difference in the *a* parameter between conditions, this result suggests interpolating Lexical Decision trials between recognition test items did not decrease the proportion of trials in which carryover occurred. The right panel shows the individual-level estimates for the *a* parameter. The analysis revealed the proportion of trials in which the previous response influenced the current response was greater in the Blank ISI condition than in the Lexical Decision condition for 36 out of 52 participants. These results are surprising insofar as both the previous Bayesian and frequentist result suggested an overall decrease in the degree of sequential dependencies in the Lexical Decision condition. Given these prior results, one would expect to observe a decrease in the *a* parameter in the Lexical Decision condition.

However, the driving force behind the decrease in sequential dependencies in the Lexical Decision condition observed in the prior analyses may be due to another parameter in the model.

91

The left panel of Figure 4.2 shows the estimated difference in the effect of the previous response between conditions is not credibly different from zero although zero lies just within the 95% HDI. The middle panel shows a clear positive effect of "Old" previous responses in the Blank ISI condition.
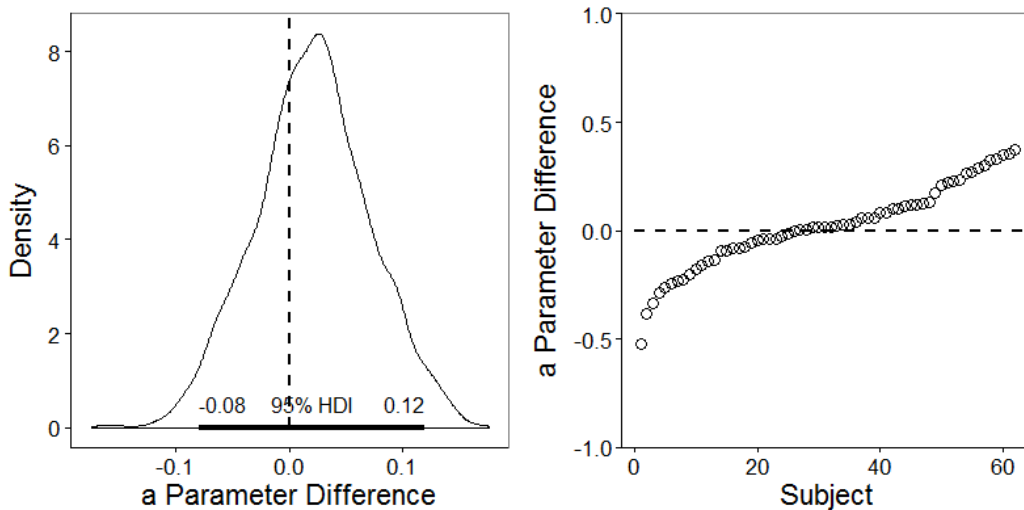
Figure 4.1. The difference in the *a* parameter across conditions.

The right panel shows the effect of the previous response in the Lexical Decision condition is not credibly different from zero. These results suggest an effect of the previous response in the Blank ISI condition, but not in the Lexical Decision condition, however, the analysis suggests that the difference in the effect of the previous stimulus between conditions is very small. While this is consistent with the previous Bayesian analysis in which a small effect size was found, this is slightly at odds with the significant Task x Previous Response interaction revealed by the ANOVA.

Figure 4.3 shows the parameter estimates for each condition upon which the difference estimates in the Figures above were based. The top row of Figure 4.3 shows the posterior probability estimates for β which represents the mnemonic strength associated with each stimulus type. The dotted vertical line represents the criterion upon which the recognition

decisions were based. There was a clear positive shift in the estimates for both old and new stimuli in the Lexical Decision condition compared with the Blank ISI condition.
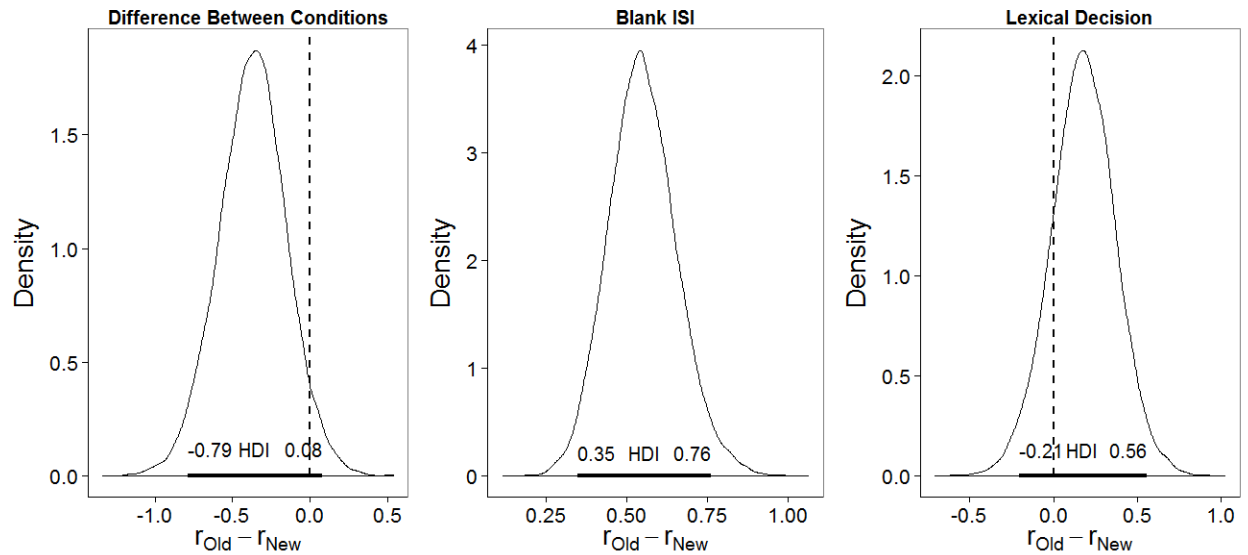
Figure 4.2. The left panel shows the difference in the effect of the previous response between conditions. The middle panel shows the effect of the previous response in the Blank ISI condition. The right panel shows the effect of the previous response in the Lexical Decision condition.

This indicates an increased bias to respond "Old" in the Lexical Decision condition. This result is corroborated with the frequentist analysis that showed an increase in the hit and false alarm rates in the Lexical Decision condition.

The second row from the top in Figure 4.3 shows the estimates for the proportion of trials in which the previous response influenced the current response for each condition. In the Blank ISI condition the 95% HDI for *a* ranged from .45 to .59 indicating 45% to 59% of responses were influenced by the previous response. This was similar to the Lexical Decision condition in which the percentage of responses that were influenced by the previous response ranged from 47% to 61%.

The third row shows zero is a credible value for the effect of previous "New" responses in both conditions. This suggests previous "New" responses did not have a systematic effect on

the current response. However, the bottom row shows zero is not a credible value for the estimated effect of "Old" previous responses in the Blank ISI condition. On the other hand, credible values for the effect of previous "Old" responses in the Lexical Decision condition ranged from -.22 to .67.
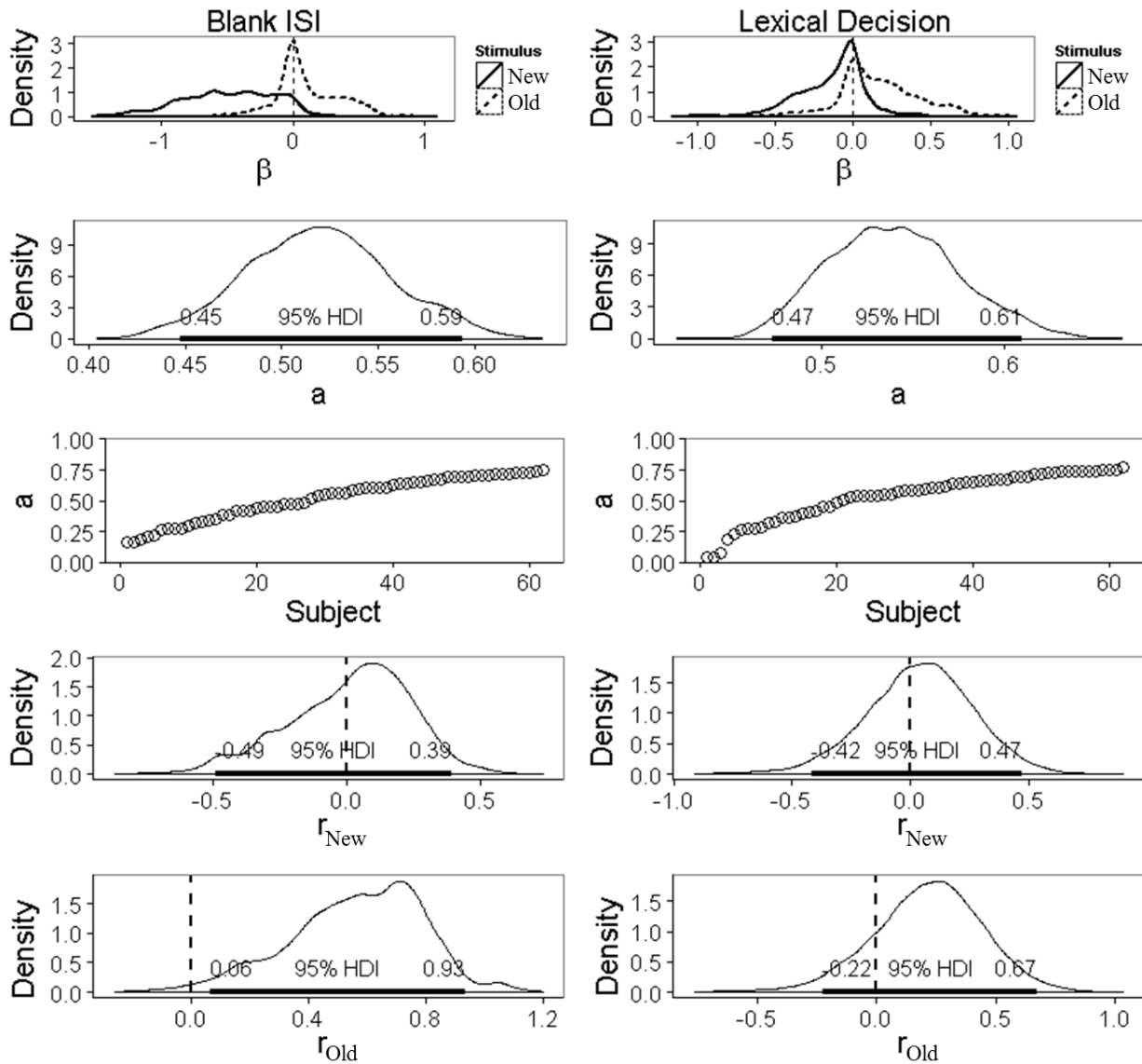
Figure 4.3. Bayesian analysis of the binary data. The left column shows the application of the model to the Blank ISI condition, while the right side shows the analysis of the Lexical Decision condition.

4.2. Discussion

The model described by Annis and Malmberg (2013) posits that vigilance may be related to the proportion of trials in which information is carried over from the previous trial. In order to test the carryover model predictions, a yes/no recognition task was used in which the interpolated task was varied in two conditions on the assumption that task switching increases vigilence. The inter-trial task was manipulated during recognition testing by either interpolating lexical decision trials or a blank ISI in which there was no task. The frequentist analysis revealed a significant increase in the tendency to respond "Old in the Lexical Decision condition. More importantly, a reduction in sequential dependencies in the Lexical Decision condition was observed, confirming the hypothesis.

In addition to analyzing the data with the frequentist approach. Several Bayesian models were also applied. By analyzing the data with models containing different assumptions, a conclusion can be reached by triangulating the results from the different analyses. In this manner, I hope to avoid basing my conclusions on the idiosyncrasies of a particular model. The Bayesian models I developed to distinguish between bias and sequential dependencies, described in section 2.6 and 2.9, were applied to the data in order to more fully test the hypothesis and draw stronger conclusions.

The Bayesian analysis revealed a difference in the amount of sequential dependencies between conditions, and an increase in the tendency to respond "Old", consistent with the frequentist result. In addition, the Bayes Factor for the effect size was small indicating only slight support for the alternative hypothesis that the effect size was not 0.

Next, the Bayesian model developed in section 3.2 was modified in order to analyze the binary data set. The development and application of this model is theoretically important insofar

as it makes assumptions that are linked to the assumptions of the process model developed by Annis and Malmberg (2013). For example, Annis and Malmberg assumed that the carryover of information from trial to trial may only occur on a proportion of trials. The parameter governing the probability of carryover occurring on a given trial was referred to as a. The Bayesian model, developed here, also makes a similar assumption. The model assumes a linear combination of normally distributed latent effects is mapped onto a decision via a set of criteria, where the latent effect of the previous response is assumed to contribute to the current decision on only some trials. For consistency, the parameter governing the proportion of trials in which this latent effect influences the current response is also referred to as a. Thus, the model allows for a very direct test of the hypothesis that interpolating inter-trial tasks will lead to a reduction in the amount of carryover. The analysis revealed, the a parameter did not differ between the Blank ISI and Lexical Decision conditions. However, a difference in the magnitude of the effect of the previous response was slightly larger in the Blank ISI condition than in the Lexical Decision condition.

Within the framework of the Annis and Malmberg model, this might be explained by positing that the representation of the interpolated Lexical Decision trials may have interfered with the representation of the retrieval cue generated on the preceding trial, while the task-switching demands placed on the subjects may not have been great enough to observe any difference in the a parameter between conditions. In order to test this hypothesis, the next experiment seeks to increase the dissimilarity between the recognition procedure and the interpolated task.

4.3.   Experiment 2

The absence of any difference in the a parameter might be due to the difference in tasks not being great enough. For example, both the lexical decision and recognition tasks involve a

binary choice of either "1" or "0." In order to increase the difference between the inter-trial task

and the recognition decision, the next experiment uses a confidence ratings procedure instead of

a yes/no task. By changing the recognition decision, this design has the added benefit of

extending the results in the previous experiment to multiple recognition procedures in addition to

increasing the dissimilarity of the inter-trial task. The ratings task will also provide a platform for

a comprehensive test of the multi-interval model. Thus, Experiment 2 was identical to

Experiment 1 except participants gave a ratings response instead of yes/no response.

4.3.1. Method

4.3.1.1. Participants. Sixty-four undergraduate students from the University of South

Florida took part in the study in exchange for course credit.

4.3.1.2. Design, Materials, and Procedure. Experiment 2 was identical to Experiment 1

except participants gave a ratings response instead of yes/no response. Subjects

responded with a "1" if they were "very confident the word was studied," "2" if they

were "less confident the word was studied," "3" if they were "less confident the word

was not studied," and "4" if they were "very confident the word was not studied."

4.3.2. Results

4.3.2.1. Accuracy. Ratings responses were analyzed in terms of hit and false alarm rates

in which P(Response < 3 | Target) = HR, and P(Response < 3 | Foil) = FAR. There were

higher hit rates in the LD condition ($M = .75$, $SD = .14$) than in the ISI condition ($M = $

$.69$, $SD = .02$), $t(63) = 4.38$, $p < .01$. The false alarm rate was also higher in the LD

condition ($M = .53$, $SD = .21$) than in the blank ISI condition ($M = .40$, $SD = .21$), $t(63) = $

6.48, $p < .01$. Thus, the bias to respond "old" was greater in the LD condition than in the

blank ISI condition. Accuracy measured by d' in the LD condition ($M = 1.13$, $SD = 1.02$)

did not significantly differ from the blank ISI condition ($M = 1.24$, $SD = 1.03$), $t(63) = .75$, $p = .454$.

4.3.2.2. Sequential Dependencies. Ratings responses were analyzed in terms of "Old" and "New" responses in which P(Response < 3) = "Old", and P(Response > 2) = "New." There was a significant main effect of Task on the probability of responding "Old", $F(1,63) = 43.95$, $p < .01$, reflecting an overall bias to respond "Old" in the condition in which a lexical decision task was interpolated between recognition test trials. There was a significant main effect of the Previous Response such that the probability of an "Old" response was greater when preceded by an "Old" response than by a "New" response, $F(1,63) = 58.95$, $p < .01$. Most importantly, the Task x Previous Response interaction was significant, $F(1,63) = 7.83$, $p < .01$. In the Blank ISI condition, there was an increased probability to respond "Old" given the previous response was "Old" ($M = .59$, $SD = .16$), rather than "New" ($M = .49$, $SD = .16$), $t(63) = 7.28$, $p < .01$. Sequential dependencies were also present in the Lexical Decision condition in which the probability of responding "Old" was greater following an "Old" response ($M = .66$, $SD = .60$) than when following a "New" response ($M = .60$, $SD = .16$), $t(63) = 4.49$, $p < .01$. Thus, according to the frequentist statistical analysis assimilation was observed in the blank interval condition and reduced in the lexical decision condition, but not eliminated.

4.3.2.3. Bayesian Analysis with the Binary Model . The Bayesian analysis shown in Table 4.2 revealed a difference in sequential dependencies between conditions in which sequential dependencies were reduced in the lexical decision condition. This is indicated by zero not lying in the 95% HDI. However, this effect is quite small as indicated by the Bayes Factor of .05 for the effect size. Table 4.3 shows that there were sequential

dependencies in each condition. However, the magnitude of the dependencies was somewhat lower in the Lexical Decision 95% HDI[.10, .25] condition than in the Blank ISI condition 95% HDI[.22, .38]. Table 4.4 shows a very large bias effect in the Lexical Decision condition compared to the Blank ISI condition.

4.3.2.4. Bayesian Analysis with the Multi-Interval Model. In this section, I apply the Multi-Interval model, outlined in section 3.3, to the current data set. Figure 4.4 through 4.7 shows the results of the model analysis. Figure 4.4 shows the posterior probability estimates for the mnemonic strength associated with each stimulus type. The dotted vertical lines shows the decision criteria. An overall positive shift was observed for strengths associated with new stimuli.
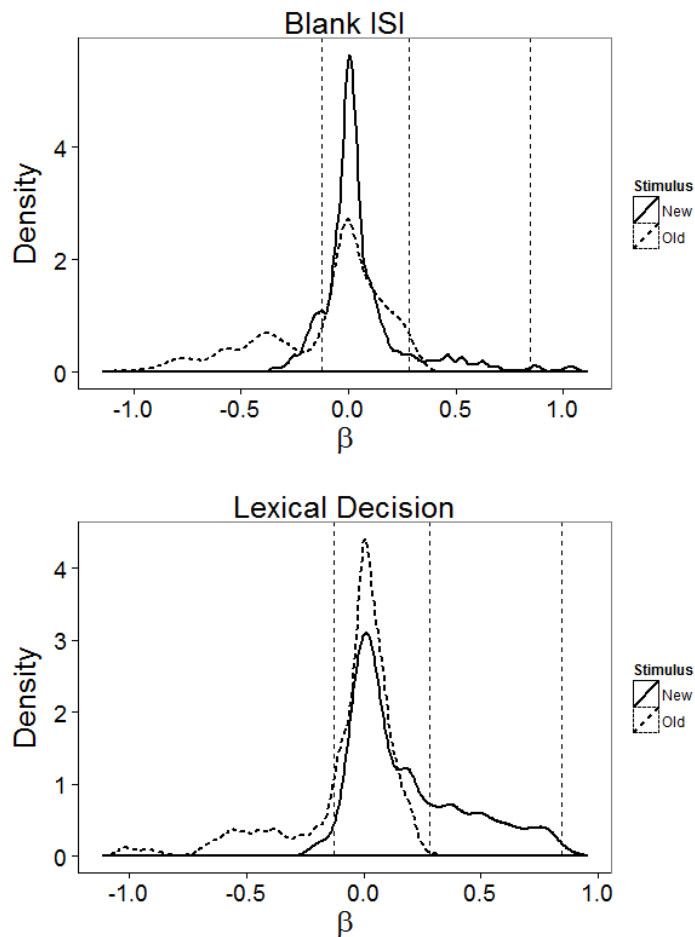
Figure 4.4. Posterior probability estimates for the mneumonic strengths associated with each stimulus type and the criteria upon which the confidence ratings were based.

The top left panel of Figure 4.5 shows the posterior probability estimate for the group-level difference in the *a* parameter between the Blank ISI and Lexical Decision conditions. The difference ranged from .03 to .10. Given that zero is outside the 95% HDI, this suggests that interpolating lexical decision trials during a ratings task decreases sequential dependencies. The right panel of Figure 4.5 shows forty-three out of sixty-two of the subjects showed some positive increase in the *a* parameter when moving from the Blank ISI condition to the Lexical Decision condition.
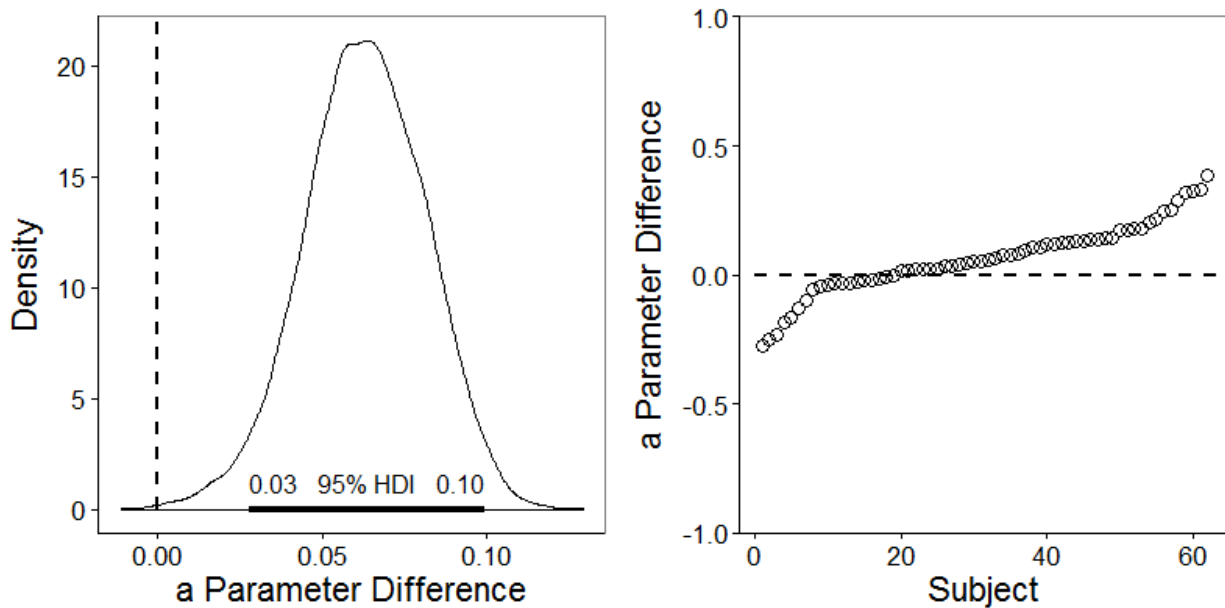
Figure 4.5. The left panel shows the posterior probability estimate for group-level difference in the *a* parameter between the Blank ISI and Lexical Decision conditions. The right panel plots these differences at the individual level.

The top row of Figure 4.6 shows the group-level posterior probability estimate for the *a* parameter in the Blank ISI condition. The 95% HDI ranged from .68 to .73. In other words, the previous response influenced the current response on 27% to 32% of trials. This probability increased in the Lexical Decision condition in which the 95% HDI of the *a* parameter ranged

100

from .73 to .80. Therefore, the previous response influenced the current response on 20% to 27% of trials in the Lexical Decision condition. These estimates are highly consistent with the estimates obtained from our process model of assimilation in judgments of frequency described above (Annis & Malmberg, 2013). We estimated the percentage of trials in which carryover occurs to be between 20% and 30%. The bottom row in Figure 4.6 shows the average estimate of the *a* parameter for each subject. The range of estimates is highly similar.
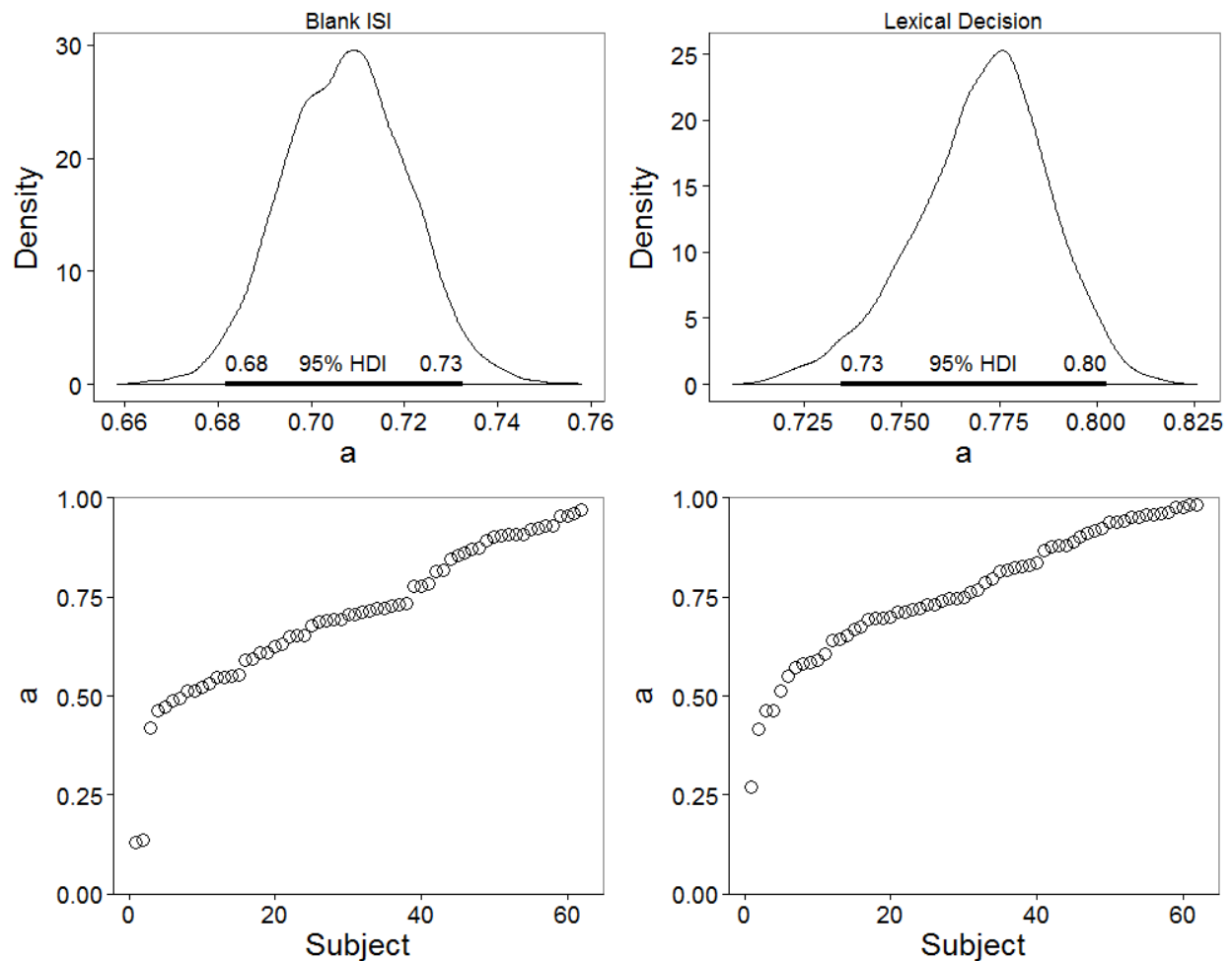
Figure 4.6. The top panels show the group-level posterior probability estimates for the *a* parameter in the Blank ISI and Lexical Decision conditions. The bottom panels show these estimates at the individual level.

The left panel of Figure 4.7 shows the difference in the effect of the previous response between conditions was not credibly different from zero. This result is consistent with the previous Bayesian analysis found in section 4.2.2.3, however, it is inconsistent with the frequentist analysis which revealed a significant Task by Previous Response interaction. The middle panel shows the difference between the effect of the lowest and highest previous response trended positively, but was not credibly different from zero. The far right panel panels shows the difference between the effect of the lowest and highest previous response was also not credibly different from zero in the Lexical Decision condition.
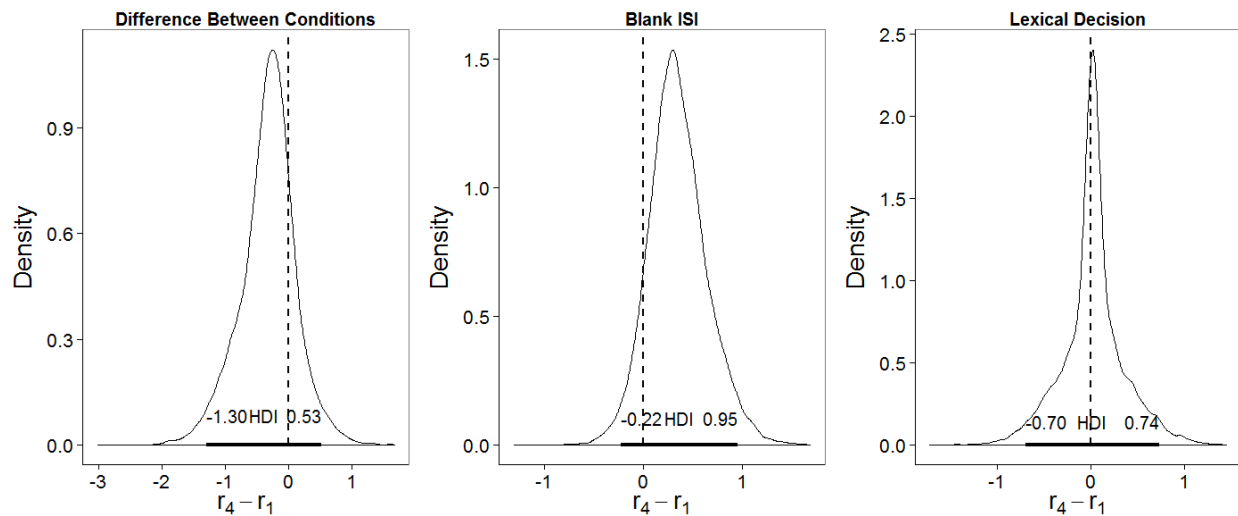
Figure 4.7. The left panel shows the difference in the effect of the previous response between conditions. The middle panel shows the effect of the previous response in the Blank ISI condition. The right panel shows the effect of the previous response in the Lexical Decision condition.

The left column of Figure 4.8 shows the posterior probability estimates of the effect of each previous response type on the current response in the Blank ISI condition. As the previous response increases the associated effect also tends to increase. For example, the 95% HDI of r_1 ranges from -.69 to .24. This range slightly increases for r_2, 95% HDI[-.30, .55], and r_3, 95% HDI[-.34, .48]. Another increase in the associated effect can be seen for the highest previous

response, r_4, 95% HDI[-.22, 1.01]. In the Lexical Decision condition, the posterior probability estimates for the effect of the previous response are relatively similar for all previous response types.
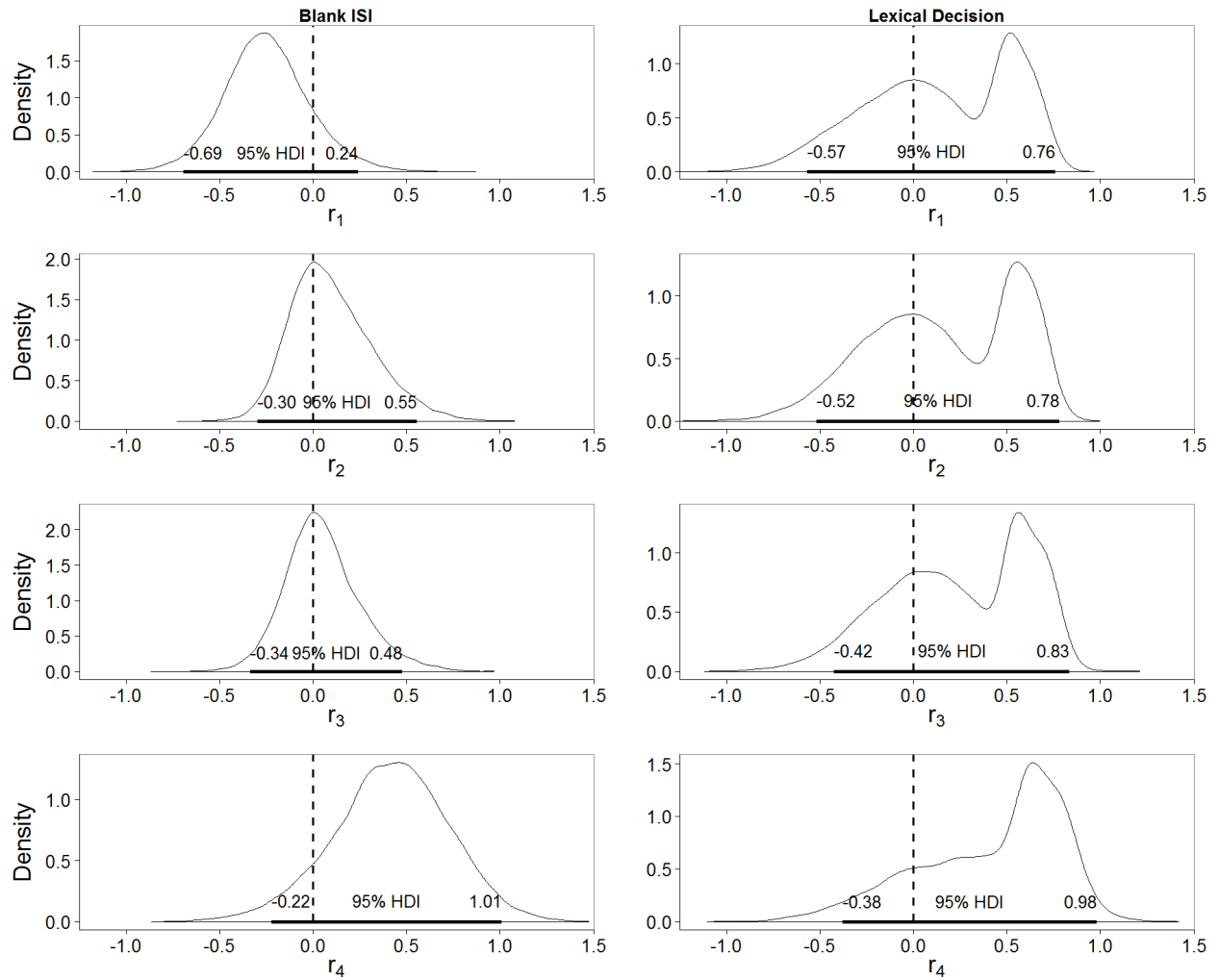
Figure 4.8. Posterior probability estimates of the effect of the previous response.

4.4.   Discussion

I hypothesized that the failure to observe a difference in the proportion of trials in which carryover occurs in Experiment 1 was due to the dissimilarity between the recognition task and the interpolated trials not being great enough. For example, both tasks were binary choice tasks. Therefore, the dissimilarity between the inter-trial and the recognition test procedure was

increased in Experiment 2 by changing the yes/no recognition task to a ratings task. This had the added benefit of generalizing the results to other recognition procedures and providing a full test of the multi-interval model developed in section 3.2.

The frequentist analysis revealed a reduction in sequential dependencies and an increased bias to respond "Old" in the Lexical Decision condition. The Bayesian analysis using the model developed in section 2.9 also revealed similar trends, however, the Bayes Factor for the effect size suggests the reduction in sequential dependencies was very small.

The multi-interval model developed in section 3.2 was also fit to the data and revealed the effect of the previous response was slightly diminished in the Lexical Decision condition. This is similar to the result found in Experiment 1 and suggests that the representation of the interpolated task may interfere with the information carried over from trial to trial.

The multi-interval model also provides a more direct test of the hypothesis that reductions in vigilance are related the tendency to carry over information from one trial to the next by explicitly assuming that the current response can be modeled by a linear combination of latent effects, where the effect of the previous response is considered on only a proportion of trials. The proportion of trials in which this carryover effect occurs is governed by the a parameter. Increases in the a parameter result in decreases in the number of trials influenced by the previous response. In Experiment 2, the a parameter was shown to increase in the Lexical Decision condition, thus supporting the hypothesis that increases in vigilance are related to decreases in the carryover of information.

Chapter 5

General Discussion

In this paper, I described and tested several Bayesian models of sequential dependencies in binary and multi-interval tasks. There were several challenges associated with measuring sequential dependencies that these models were shown to overcome. The main challenges for these models included separately measuring bias and sequential dependencies, measuring the magnitude and direction of assimilation and contrast, measuring sequential dependencies at lags greater than 1, measuring the proportion of trials on which sequential dependencies occur, and decorrelating the previous response from the previous stimulus. In order to test the validity of the measurement models I simulated data with known properties and determined whether the model was able to successfully detect those properties.

5.1.   Binary Models

The first issue that was identified was the problem of distinguishing between bias and sequential dependencies. In Chapter 1, I showed the formal relationship between sequential dependencies and bias. Holding all else constant, the relationship shows that as bias to respond "Old" increases, so too will the probability of repeating an "Old" response. Therefore, I argued that it is necessary to develop models of sequential dependencies that explicitly take into account both bias and sequential dependencies. The ability to distinguish between bias and sequential dependencies is also classically motivated by models such as Signal Detection Theory which was developed in order to distinguish between bias and sensitivity.

The first model that was tested was a Bayesian version of a *t*-test developed by Wagenmakers et al. (2010). While this model was shown to be able to distinguish between bias and sequential dependencies, it did not explicitly take bias into account. Because of this, the model was inefficient in that it took a large number of observations to reach a Bayes Factor that would decisively indicate whether sequential dependencies were absent or present in the data. The reason for this is that some of the variance in the rate of responding "Old" following "Old" and "Old" following "New" is due to bias as well as sequential dependencies. When bias is taken into account this variance is better accounted for. By including a separate parameter that modeled bias, the model showed a drastic improvement in its ability to decisively identify the presence or absence of sequential dependencies and bias.

A limitation of the model was that it only took into account sequential dependencies between the current response and previous response, but not between the current response and the previous stimulus. For example, in yes/no recognition the data are usually analyzed in terms of hits, false alarms, misses, and correct rejections. Thus, there are 16 different sequential dependencies that can arise such as hits following hits, hits following misses, false alarms following correct rejections, misses following false alarms and so forth. In order to develop a complete theory of sequential dependencies it will be necessary to know whether there are differences that exist between these different dependencies. Therefore, a generalized version of the model that takes into account all 16 different stimulus-response combinations that exist in a yes/no recognition task needs to be developed. A Bayesian model is almost essential in such a situation as data sparsity will almost certainly be an issue (see Chapter 1).

Another limitation of the model was that it could only be used to analyze data from a single condition in an experiment. However, the researcher might be interested in how sequential

dependencies might differ across conditions of an experiment. For example, in Chapter 4 I tested the hypothesis that task switching would decrease the amount of sequential dependencies during recognition testing. I extended the model by duplicating the model for each condition and then modeled the differences in bias and sequential dependency between conditions. If there are many different conditions of the experiment, this model could be applied to each condition pair. On the other hand, if the researcher is interested in the overall effect of some manipulation, then a Bayesian ANOVA might be more appropriate (see Kruschke, 2011).

5.2.   Multi-Interval Models

The models outlined above, are only capable of analyzing sequential dependencies in binary data. Some of the most interesting and complex patterns of sequential dependencies come from perception tasks like absolute identification and memory tasks like judgments of frequency, both of which are multi-interval response tasks. For example, in absolute identification, positive sequential dependencies are observed between the current response and previous stimuli, but in the presence of feedback, this pattern reverses at lags greater than 1 and negative sequential dependencies are observed (Ward & Lockhead, 1970). In a judgment of frequency task, negative sequential dependencies are associated with previous stimulus, but positive sequential dependencies are associated with the previous response (Annis & Malmberg, 2013). Given the complexity of the phenomenon it should not come as a surprise that there were many different challenges to measuring sequential dependencies in multi-interval tasks. In the following sections I will briefly review these challenges and how the model handled each one.

5.2.1. Response Bias and Sensitivity

The first challenge was to independently measure response bias and sensitivity (Luce, Nosofsky, Green, & Smith, 1982). Often times, participants will bin their responses on extreme

ends of the response scale. Because of this bias, the accuracy for the stimuli associated with such responses will increase. The multi-interval model handled this issue by assuming criterion and sensitivity parameters related to Signal Detection Theory (Morey, Pratte, & Rouder, 2008). To test the model, I simulated a data set containing a bias to respond at high ends of the response scale. The model was able to correctly distinguish between bias and sensitivity. In addition, the model did not misclassify this bias as a sequential dependency.

5.2.2. Decorrelating the Previous Stimulus and Previous Response

Another challenge was to decorrelate the effect of the previous stimulus from the previous response. In order to test the model, I simulated a data set that contained negative sequential dependencies between the current response and previous response and positive sequential dependencies between the current response and previous stimulus. In addition, I correlated the previous response with previous stimulus by increasing the overall accuracy. While the standard method of plotting the error on the current trial as a function of the previous response showed positive sequential dependencies (caused by the correlation of the previous response with the previous stimulus), the model correctly identified the dependency as negative. The model also correctly identified the positive sequential dependency between the current response and previous stimulus.

5.2.3. Relating Measurement and Process

Modeling sequential dependencies at both the process level and the measurement level is critical for future model development (Malmberg, 2008; Estes, 1975). This dissertation represents some of the first steps towards such a goal.

Annis and Malmberg (2013) developed a process model of sequential dependencies in the REM framework (Shiffrin & Steyvers, 1997) by assuming that feature values from the previous

retrieval cue carries over to the current retrieval cue. The model further assumes that carryover may not occur on every trial. In order to test the model, we conducted a judgment of frequency task and estimated, via REM simulations, the number of trials in which carryover occurs to be between 20% and 30% of trials. This estimate turned out to be very similar the measurement model estimate obtained from the ratings task described in Chapter 4. Thus, the first tentative link between a process model and measurement model of sequential dependencies in recognition memory has been made.

The consistency is surprising considering that the estimates came from different models that were applied to different recognition memory tasks with different stimuli and different subjects. In order to bolster the link between the process model and the measurement model, REM should be applied to the current data set and the *a* parameter estimated. Future work should also include a rigorous study of the relationship between the *a* parameter in our process model and the *a* parameter in the current measurement model with simulated data. This could be achieved through simulating data sets with known parameter values via processes outlined in this manuscript.

5.3. Insights Gained By Simulating Data

The Markov model proved to be incredibly powerful tool that was capable of generating a wide array of sequential dependency patterns and yet, at its core, is very simple. For instance, the Markov model only took into account the most recent trial. Given the very limited memory of the model, I did not expect it to generate some of the results that it did. For example, it was able to generate the decaying pattern of assimilation observed at lags greater than 1. The most surprising result was that it was capable of generating positive sequential dependencies at lags of 1 and then reversing this pattern to create negative sequential dependencies at lags of 2 and

greater. This was unexpected given the model only considered the most recent trial to generate

its current response. The reason behind such a baffling result was due to an oscillatory response

pattern of which I present the details in Chapter 3. This result is theoretically relevant as many

models of sequential dependencies posit the representation of the current stimulus is

contaminated by representations occurring more than 1 trial back. However, these simulations

show this assumption may be sufficient, but not necessary in order to generate contrast.

Although the Markov model was used to *simulate* sequential dependences, it would be

straight forward to estimate the transition probabilities from the data. Thus, the Markov model in

itself can be used as a measurement model of sequential dependencies. Given its flexibility and

its lack of theoretical assumptions, it could be used to relate a wide range of tasks across memory

and perception. There were many serendipitous insights that were gleaned from this model and it

may hold many more.

5.4.   Insights Gained By Applying the Model to Real-World Data

In Chapter 4, the models described above were applied to real-world data from two

experiments in which the inter-trial task was manipulated during recognition testing. The first

experiment was a simple yes/no recognition experiment in which either a blank ISI or a lexical

decision task was interpolated between recognition test trials. Both the Bayesian and frequentist

analysis suggested sequential dependencies decreased in the Lexical Decision condition. When

analyzing the simple effects, the frequentist approach revealed the probability of an "Old"

following an "Old" response did not significantly differ from the probability of an "Old"

response following a "New" response. Often times, non-significant results are misinterpreted as

evidence in favor of the null hypothesis. On the other hand, the Bayesian analysis does not suffer

from such limitations. The Bayes Factor revealed an effect size of zero was 25 times more likely than an effect size not equaling zero.

Experiment 2 was the same as Experiment 1 with the exception that subjects made confidence ratings instead of yes/no responses. Both the frequentist and Bayesian analysis indicated sequential dependencies decreased in the Lexical Decision condition. In addition, the multi-interval model found the proportion of trials that the previous response influenced the current response was lower in the Lexical Decision condition by 3% - 10%.

It should be noted, that obtaining such a result would not be possible with standard regression models as they would assume the effect of the previous response is factored into every current response (Jones, Love, & Maddox, 2006; Jesteadt, Luce, & Green, 1977; Lockhead & King, 1983; Mathews & Stewart, 2009). This measurement assumption is rather inflexible and is not compatible with our process model described in Chapter 1.

In addition to modeling the group parameter values, the multi-interval model showed there were large variations in individual differences in how often the previous response influenced the current response. This model could be used in future investigations into why these individual differences exist and what cognitive factors might be correlated with sequential dependencies.

5.4.1. Limitations of the Measurement Model. Although the multi-interval model showed to be very helpful in increasing our theoretical understanding of sequential dependencies, there are several issues the present analysis did not address. In this section I will describe these issues.

Output interference in recognition occurs when accuracy decreases over the course of testing and is thought to be the result of interference from the storage of item information at test

(Criss, Malmberg, & Shiffrin, 2011; Malmberg, Criss, Gangwani, Shiffrin, 2012; Annis, Malmberg, Criss, & Shiffrin, 2013). In a similar experiment to the one presented here, Annis et al. (2013) varied the inter-trial task during recognition testing and found the interpolation of lexical decision trials does not result in increased output interference compared to a blank ISI. Annis et al. hypothesized that traces associated with lexical decision trials do not produce additional interference because the task context of the lexical decision trials is considerably different from the task context associated with recognition trials.

Given the strong role of task context during recognition testing, could task context also be used to explain the present set of results? For example, given a recognition test on trial n, task context features might carryover from the lexical decision task on trial $n-1$. Given that task context differs between tasks, the task context from the interpolated lexical decision trial might decorrelate the retrieval cues associated with the adjacent recognition test trials and a decrease in sequential dependencies would result.

Another type of context that might be used to explain the present results is known as *temporal context* which is assumed to gradually change over the course of recognition testing (e.g. Annis et al., 2013). Thus, the "drifting" of temporal context could be used as a surrogate for the carryover of item information. For example, the retrieval cue on the current recognition test trial might be correlated with the previous retrieval cue due to the similarity of the temporal contexts associated with each trial rather than the carryover of item information. This explanation might be sufficient to explain the presence of sequential dependencies during the blank ISI condition, but it would be difficult to explain why sequential dependencies decreased in the Lexical Decision condition. Context and item representations are defined at the process

level, and therefore the multi-interval measurement model would not be able to differentiate between the two.

Although the measurement model cannot offer explanations for sequential dependencies at the process level, it has the advantage of including assumptions closely tied to those at the process level. Measurement models that are not linked to process models often times suffer from circularity. For example, Treisman & Williams (1984) proposed a model of sequential dependencies within the Signal Detection Theory (SDT; Green & Swets, 1966) framework in which sequential dependencies were the result of the criterion shifting as a function of the response on the current trial. An "old" response would cause the criterion to shift negatively. Thus, on the following trial it would be more likely that an "old" response would be repeated. In addition, a decay function caused the shifted criterion to gradually move back to a stationary point on the axis during the interstimulus interval. Therefore, the results can be explained within the Treisman and Williams framework either by hypothesizing that the decay function differs between conditions or that the magnitude of criterion shifting differs between conditions. However, the model does not explain why these differences between conditions would exist. Therefore, explaining the present results within this framework would be circular. Because the multi-interval measurement model is theoretically linked to a process model of sequential dependencies, it is far less likely to suffer from such motivational issues. There is a clear theoretical explanation as to why the *a* parameter changes between conditions.

5.5.   Future Directions

An overarching goal of the research presented herein is to develop a measurement model of sequential dependencies that can be related to neural models of sequential dependencies. Recently, the technical aspects of this goal have been greatly elucidated by Turner et al. (2013)

who developed a Bayesian framework for simultaneously fitting cognitive models to behavioral data and neural models to neural data such that each model constrains the other's fits. This represents a significant methodological step forward in the quest to relate brain to behavior and has immediate implications for the measurement models of sequential dependencies I described Chapter 2 and 3.

For example, Kondo and Watanabe (2012) presented stimuli varying in brightness and asked participants to provide a brightness judgment. They found increases in activation in the left-occipto-temporal region with increases in the brightness of the preceding stimulus. This is consistent with the model described by Annis and Malmberg (2013) in which increases in the familiarity strength of the preceding stimulus cause increases in the overestimation of the current stimulus. Given the measurement model described in Chapter 3, would this increased activation be related to the parameter associated with the effect of the previous stimulus?

Squire, et al. (1976) were interested in how variations in the sequence of stimuli affect the amplitude of the P300, an event related potential (ERP) known to be elicited upon the presentation of low probability target items in the oddball paradigm. To investigate this, Squire et al. presented various sequences of high and low pitched tone bursts. They found that the amplitude of the P300 in response to a high pitched tone increases as the number of low pitched tones preceding it increases. In addition, they observed that stimuli as far as four trials back had influence on the P300 associated with the current trial, and this influence decayed as lag increased. This is consistent with the Annis and Malmberg model as well as many models of absolute identification which assume stimuli further back usually have less influence on the representation of the current stimulus than immediately preceding stimuli. Further, it might be

possible to relate the decaying influence of the previous stimulus to parameters associated with stimuli at varying lags.

Finally, attention appears to be a critical component involved in modulating information carryover. In Chapter 4 I showed that increases in task vigilance are related to decreases in the tendency to carryover information from one trial to the next. Additional evidence for this hypothesis comes from cognitive neuroscience. Recently, Payne et al. (2013) showed that brain oscillations related to attention, referred to as alpha-band oscillations, are related to the interference of previous stimuli. They found that as alpha power increased, the ability to filter task-irrelevant information also increased. Thus, the relationship between the a parameter in the measurement model and alpha power might provide the basis for future investigations of the underlying mechanisms of sequential dependencies.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409-424.

Annis, J., & Malmberg, K. J. (2013). A model of positive sequential dependencies in judgments of frequency. *Journal of Mathematical Psychology,* 57, 225-236.

Annis, J., Malmberg, K.J., Criss, A.H., & Shiffrin, R.M. (2013). Sources of interference in recognition testing. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 39*, 1365-1376.

Brown, S.D., Marley, A. A. J., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review, 115*, 396-425.

Criss, A.H., Malmberg, K. J., & Shiffrin, R.M. (2011). Output Interference in Recognition Memory. *Journal of Memory and Language, 64*, 316-326.

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review*, *109*, 710-721.

DeCarlo, L. T. (2007). The mirror effect and mixture signal detection theory. Journal of *Experimental Psychology: Learning, Memory, and Cognition*, *33*, 18-33.

Dennis, S., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list length effect. Journal of Memory and Language, 59, 361-376.

Dickey, J. M. (1971). The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location

    Parameters. *The Annals of Mathematical Statistics, 42,* 204-223.

Estes, W. K. (1975). *Handbook of learning and cognitive processes*. Hillsdale, N.J: L. Erlbaum

    Associates.

Egan, J. P. (1958). Recognition memory and the operating characteristic (No. AFCRC-TN-58–

    51). Bloomington, IN: Indiana University Hearing and Communications Laboratory.

Gelman, A., Carlin, J. B., Stern, H.S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.).

    London. Chapman & Hall/CRC.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall.

    Psychological Review, 91, 1-67.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York:

    Wiley.

Griffiths, T. L., Sanborn, A. N., Canini, K. R., Navarro, D. J., & Tenenbaum, J. B. (2011).

    Nonparametric Bayesian models of categorization. In Pothos, E., & Wills, A. (Eds.)

    Formal Approaches in Categorization.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace

    memory model. *Psychological Review, 95,* 528-551.

Howard, M. W., Bessette-Symons, B., Zhang, Y., Hoyer, W. J. (2006). Aging selectively impairs

    recollection in recognition memory for pictures: Evidence from modeling and receiver

    operating characteristic curves. *Psychology and Aging*, *21*, 96-106.

Howarth, C. I., & Bulmer, M. G. (1956). Non-random sequences in visual threshold experiments.

    Quarterly Journal of Experimental Psychology, 8, 163-171.

Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Ruys, K. I. (2001). Perception and preference in short-term word priming. *Psychological Review, 108*, 149-182.

Jaynes, E. T. (2003). *Probability theory: The logic of science* (G. L. Bretthorst, Ed.). New York: Cambridge University Press.

Jeffreys, H. (1961). *Theory of probability*. Oxford: Clarendon Press.

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology. Human Perception and Performance, 3*, 92-104.

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32*, 316-332.

Kass, R. E., & Wasserman, L. (1995). A Reference Bayesian Test for Nested Hypotheses and its Relationship to the Schwarz Criterion. *Journal of the American Statistical Association, 90*, 928-934.

Kemeny, J. G., & Snell, J. L. (1976). *Finite Markov chains*. New York: Springer-Verlag.

Kruschke, J. K. (2011). *Doing Bayesian data analyses: A tutorial with R and BUGS.* Burlington, MA: Academic Press.

Kucera, H., & Francis, W. (1983). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

Lacouture, Y. & Marley, A. A. J. (1995). A mapping model of bow effects in absolute identification. *Journal of Mathematical Psychology, 39,* 383–395.

Lee, M. D. (2008). BayesSDT: Software for Bayesian inference with signal detection theory. *Behavior Research Methods, 40*, 450-456.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course.* Cambridge University Press.

Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review, 12*, 605-621.

Lockhead, G. R., & King, M. C. (January 01, 1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology. Human Perception and Performance, 9*, 461-73.

Luce, R. D., Nosofsky, R. M., Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. *Perception & Psychophysics, 32,* 397-408.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 539–559.

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57,* 335-384.

Malmberg, K. J., & Annis, J. (2012). On the relationship between memory and perception: Sequential dependencies in recognition memory testing. *Journal of Experimental Psychology: General, 141*, 233-259.

Malmberg, K. J., Criss, A.H., Gangwani, T. & Shiffrin, R.M. (2012). Overcoming the Negative Consequences of Interference that Results from Recognition Memory Testing. *Psychological Science, 23*, 115-119.

Malmberg, K. J. & Murnane, K. (2002). List composition and the word- frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 616-630.

Matthews, W. J., & Stewart, N. (2009). The effect of interstimulus interval on sequential effects in absolute identification. *Quarterly Journal of Experimental Psychology, 62,* 2014-29.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 63, 81-97.

Monsell, S. (2003). Task switching. *Trends in Cognitive Science, 7*, 134-140.

Morey, R. D., Pratte, M. S., & Rouder, J. N. (2008). Problematic effects of aggregation in zROC analysis and a hierarchical modeling solution. *Journal of Mathematical Psychology, 52*, 376-388.

Murdock, B. B., Jr. (1983). A distributed memory model for serial-order information. *Psychological Review, 90*, 316-338.

Navarro, D. J., & Griffiths, T. L. (2008). Latent features in similarity judgment: A nonparametric Bayesian approach. *Neural Computation, 20*, 2597-2628.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Payne, L., Guillory, S., & Sekuler, R. (2013). Attention-modulated alpha-band oscillations protect against intrusion of irrelevant information. *Journal of Cognitive Neuroscience, 9*, 1463.

Petrov, A., & Anderson, J. R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review, 112*(2), 383-416.

Pitt, M. A., Myung, I., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*, 472-491.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Paper presented at the *Proceedings of the 3rd International Workshop on Distributed Statistical Computing,* 20-22.

Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior, 17*, 403-417.

Rice, J. A. (2007). *Mathematical statistics and data analysis*. Belmont, CA: Thomson/Brooks/Cole.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review, 12*, 573-604.

Rouder, J. N., Lu, J., & Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review, 12*, 195-223.

Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory, *Psychological Science, 16*, 898-904.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145-166.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General,* 117, 34-50.

Spiegelhalter, D. J., Thomas, A., & Best, N. G. (1999). *WinBUGS Version 1.2 User Manual*. Cambridge, UK.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The Effect of Stimulus Sequence on the Waveform of the Cortical Event-Related Potential. *Science, 193,* 1142-1146.

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: a meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17*, 99–139.

Stewart, N., Brown, G. D. A., Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review, 112*(4), 881-911.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91*, 68-111.

Turner, B. M., Forstmann, B. U., Wagenmakers, E. J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *Neuroimage, 72*, 193-206.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology, 60*, 158-189.

Ward, L. M., & Lockhead, G. R. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology, 84*, 27-34.

Appendix A Binary Decision Model JAGS Code

```
model
{
    for(i in 1:NSUBJ){
        #counts of yes given yes responses
        #are binomially distributed
        #Eq. 2.7
        YY[i] ~ dbin(theta1[i],Y_[i])
        #counts of yes given no responses
        #are binomially distributed
        NY[i] ~ dbin(theta2[i],N_[i])
        #transformation to rate scale
        theta1[i] <- phi(phi1[i])
        theta2[i] <- phi(phi2[i])
        phi1[i] ~ dnorm(muphi,tauphi)
        phi2[i] <- phi1[i]+alpha[i]
        #alpha is the difference between
        #the rate of responding yes following yes
        #and the rate of responding yes following no
        #for each individual
        alpha[i] ~ dnorm(mualpha,taualpha)
        }
        # Group-Level Priors:
        muphi ~ dnorm(0,1)
        #uninformative group level standard deviation
        sigmaphi ~ dunif(0,10)
        #transformation to precision
        tauphi <- pow(sigmaphi,-2)
        #the group level mean of alpha
        mualpha <- delta * sigmaalpha
        #uninformative prior
        sigmaalpha ~ dunif(0,10)
        #transformation to precision
        taualpha <- pow(sigmaalpha,-2)
        #unit information prior on the effect size
        delta ~ dnorm(0,1)
    }
```

Appendix B Markov Chain R Code

```r
sim.data.markov <- function(trans.mat,chain.length){
#Generates a markov chain with transition matrix, trans.mat,
#trans.mat: transition matrix
#chain.length: length of the chain
#Author:Jeffrey Annis
#Date:3.29.14

  num.states = length(trans.mat[,1])
  init.probs = rep(1,num.states)/num.states #all states have eq. prob at t=1
  state = NULL
  state <- sample(1:num.states,1,prob=init.probs) #get initial state

  for(i in 1:(chain.length-1)){
    trans.probs <- trans.mat[state[i],] #get the transition probabilities
    state[i+1] <- sample(1:num.states,1,prob=trans.probs) #sample next state
  }

  return(state)

}
```

Appendix C Bias Model JAGS Code

```
model{

    for(i in 1:NSUBJ){
        #counts of yes given yes responses
        #are binomially distributed
        YY[i] ~ dbinom(theta.b.phi[i], Y_[i])
        #counts of yes given no responses
        #are binomially distributed
        NY[i] ~ dbinom(theta.a.phi[i], N_[i])
        #transform theta.a to rate scale
        theta.b.phi[i] <- phi(theta.b[i])
        #transform theta.b to rate scale
        theta.a.phi[i] <- phi(theta.a[i])
        #reparameterization of rates
        theta.a[i] <- bias[i] - (alpha[i]/2)
        theta.b[i] <- bias[i] + (alpha[i]/2)
        #sequential dependency parameter
        alpha[i] ~ dnorm(mu.alpha, sigma.alpha)
        #bias parameter
        bias[i] ~ dnorm(mu.bias,sigma.bias)
    }

    #bias
    mu.bias <- delta.bias * sigma.bias
    sigma.bias <- pow(sigma.bias.prec,-2)
    sigma.bias.prec ~ dunif(0,10)

    #hyperpriors on alpha
    mu.alpha <- delta.alpha * sigma.alpha
    sigma.alpha <- pow(sigma.alpha.prec,-2)
    sigma.alpha.prec ~ dunif(0,10)

    #effect size of alpha
    delta.alpha ~ dnorm(0,1)
    delta.bias ~ dnorm(0,1)

    #hyperpriors on theta.a
    sigma.phi <- pow(sigma.phi.prec,-2)
    sigma.phi.prec ~ dunif(0,10)
}
```

Appendix D Multiple Conditions JAGS Code

```
model{

    for(i in 1:NSUBJ){

        for(j in 1:2){
            YY[i,j] ~ dbinom(theta.b.phi[i,j], Y_[i,j])
            NY[i,j] ~ dbinom(theta.a.phi[i,j], N_[i,j])

            theta.b.phi[i,j] <- phi(theta.b[i,j])
            theta.a.phi[i,j] <- phi(theta.a[i,j])
            theta.a[i,j] <- bias[i,j] - (alpha[i,j]/2)
            theta.b[i,j] <- bias[i,j] + (alpha[i,j]/2)
        }

    alpha[i,1] ~ dnorm(mu.alpha, sigma.alpha)
    alpha[i,2] <- alpha[i,1] + gamma.alpha[i]
    gamma.alpha[i] ~ dnorm(mu.gamma.alpha,sigma.gamma.alpha)
    bias[i,1] ~ dnorm(mu.bias,sigma.bias)
    bias[i,2] <- bias[i,1] + gamma.bias[i]
    gamma.bias[i] ~ dnorm(mu.gamma.bias,sigma.gamma.alpha)


    }

    #hyperpriors on bias
    mu.bias <- delta.bias * sigma.bias
    sigma.bias <- pow(sigma.bias.prec,-2)
    sigma.bias.prec ~ dunif(0,10)

    #hyperpriors on alpha
    mu.alpha <- delta.alpha * sigma.alpha
    sigma.alpha <- pow(sigma.alpha.prec,-2)
    sigma.alpha.prec ~ dunif(0,10)
    #hyperpriors on gamma.alpha
    mu.gamma.alpha <- delta.gamma.alpha * sigma.gamma.alpha
    sigma.gamma.alpha <- pow(sigma.gamma.alpha.prec,-2)
    sigma.gamma.alpha.prec ~ dunif(0,10)
    #hyperpriors on gamma.bias
    mu.gamma.bias <- delta.gamma.bias * sigma.gamma.bias
    sigma.gamma.bias <- pow(sigma.gamma.bias.prec,-2)
    sigma.gamma.bias.prec ~ dunif(0,10)
    #delta priors
    delta.alpha ~ dnorm(0,1)
    delta.bias ~ dnorm(0,1)
    delta.gamma.bias ~ dnorm(0,1)
    delta.gamma.alpha ~ dnorm(0,1)

}
```

Appendix E Multi-Interval Model R, JAGS Code

```
model{

for(i in 1:NSUBJ){ #NSUBJ: number of subjects
    #the first LAG trials are not influenced by the previous trials
    for(j in 1:LAG){ #LAG is the number of previous trials considered
        #the data are categorical
        R[i,j] ~ dcat(theta[i,j,1:NRESP]) #NRESP: highest possible response
        #probability of subject i making a response between
        #criterion k and k-1, given mu[i,j]
        for(k in 2:(NRESP-1)){
            theta[i,j,k] <- max( 0 , phi((C[k]-mu[i,j])/sigma[S[i,j]])
            - phi((C[k-1]-mu[i,j])/sigma[S[i,j]]) )
        }
        #probability of subject i making the lowest response
        #given mu[i,j]
        theta[i,j,1] <- phi((C[1]-mu[i,j])/sigma[S[i,j]])
        #probability of subject i making the highest response
        #given mu[i,j]
        theta[i,j,NRESP] <- 1 - phi((C[NRESP-1]-mu[i,j])/sigma[S[i,j]])
        #b0 is the overall latent strength associated with
        #each stimulus value. gamma is subject x stimulus interaction
        mu[i,j] <- b0[S[i,j]] + gamma[i,S[i,j]]
    }
    #this section of code is the same as the last section
    #except this section takes into account the effect of the previous
    #stimulus up LAG trials back
    for(j in (LAG+1):NTRIALS){ #NTRIALS: total number of trials
        R[i,j] ~ dcat(theta[i,j,1:NRESP])
        for(k in 2:(NRESP-1)){
            theta[i,j,k] <- max( 0 , phi((C[k]-mu[i,j])/sigma[S[i,j]])
            - phi((C[k-1]-mu[i,j])/sigma[S[i,j]])   )
        }
        theta[i,j,1] <- phi((C[1]-mu[i,j])/sigma[S[i,j]])
        theta[i,j,NRESP] <- 1 - phi((C[NRESP-1]-mu[i,j])/sigma[S[i,j]])
        #if x[i,j] = 1 then the previous resposne is taken into account
        #if x[i,j] = 0 then the effect is 0
        #r[i,R[i,j-1]] is the effect of the previous response for each
        #subject i
        mu[i,j] <- b0[S[i,j]] + gamma[i,S[i,j]] + x[i,j]*r[i,R[i,j-1]]
        #the proportion of trials in which carryover occurs follows
        #a Bernoulli distribution with success parameter 1-a[i]
        x[i,j] ~ dbern(1-a[i])
    }

    for(m in 1:NSTIM){#NSTIM is the number of stimuli
        #the subject x stimulus interaction
        #has group level mean mu.gamma and
```

```
        #group level precision sigma.gamma
        gamma[i,m] ~ dnorm(mu.gamma[m],sigma.gamma[m])
        #the subject x previous response interaction
        #has group level mean mu.r and precision sigma.r
        r[i,m] ~ dnorm(mu.r[m],sigma.r[m])
    }

    #uniform prior for the proportion of trials
    #in which carryover occurs
    a[i] ~ dbeta(1,1)

}   #uninformative group level priors
    for(m in 1:NSTIM){
        b0[m] ~ dnorm(0,sigma.b0[m])
        mu.gamma[m] ~ dnorm(0,10)
        mu.r[m] ~ dnorm(0,10)
        sigma.gamma[m] ~ dgamma(.001,.001)
        sigma.b0[m] ~ dgamma(.001,.001)
        sigma.r[m] ~ dgamma(.001,.001)
        sigma.s[m] ~ dgamma(.001,.001)
        sigma[m] ~ dgamma(.001,.001)
    }

}
```

Appendix F Multi-Interval Model SR, JAGS Code

```
model{

for(i in 1:NSUBJ){
    #the first LAG trials are not influenced by the previous trials
    for(j in 1:LAG){#LAG is the number of previous trials considered
        #the data are categorical
        R[i,j] ~ dcat(theta[i,j,1:NRESP])
        #probability of subject i making a response between
        #criterion k and k-1, given mu[i,j]
        for(k in 2:(NRESP-1)){
            theta[i,j,k] <- max( 0 , phi((C[k]-mu[i,j])/sigma[S[i,j]])
            - phi((C[k-1]-mu[i,j])/sigma[S[i,j]]) )
        }
        #probability of subject i making the lowest response
        #given mu[i,j]
        theta[i,j,1] <- phi((C[1]-mu[i,j])/sigma[S[i,j]])
        #probability of subject i making the highest response
        #given mu[i,j]
        theta[i,j,NRESP] <- 1 - phi((C[NRESP-1]-mu[i,j])/sigma[S[i,j]])
        #b0 is the overall latent strength associated with
        #each stimulus value. gamma is subject x stimulus interaction
        mu[i,j] <- b0[S[i,j]] + gamma[i,S[i,j]]
    }
    #this section of code is the same as the last section
    #except this section takes into account the effect of the previous
    #stimulus up LAG trials back
    for(j in (LAG+1):NTRIALS){
        R[i,j] ~ dcat(theta[i,j,1:NRESP])
        for(k in 2:(NRESP-1)){
            theta[i,j,k] <- max( 0 , phi((C[k]-mu[i,j])/sigma[S[i,j]])
            - phi((C[k-1]-mu[i,j])/sigma[S[i,j]])   )
        }
        theta[i,j,1] <- phi((C[1]-mu[i,j])/sigma[S[i,j]])
        theta[i,j,NRESP] <- 1 - phi((C[NRESP-1]-mu[i,j])/sigma[S[i,j]])
        #if x.r[i,j] = 1 then the previous resposne is taken into account
        #if x.r[i,j] = 0 then the effect is 0
        #this same rule is applied to the effect of the previous stimulus
        #for x.s[i,j]
        #r[i,R[i,j-1]] is the effect of the previous response for each
        #subject i
        #s[i,S[i,j-1]] is the effect of the previous stimulus for each
        #subject i
        mu[i,j] <- b0[S[i,j]] + gamma[i,S[i,j]] + x.r[i,j]*r[i,R[i,j-1]]
        + x.s[i,j]*s[i,S[i,j-1]]
        #the proportion of trials in which carryover occurs follows
        #a Bernoulli distribution with success parameter 1-a[i]
        x.r[i,j] ~ dbern(1-a.r[i])
```

```
        x.s[i,j] ~ dbern(1-a.s[i])
    }


    #priors at the individual level for each stimulus
    for(m in 1:NSTIM){#NSTIM is the number of stimuli
        #the subject x stimulus interaction
        #has group level mean mu.gamma and
        #group level precision sigma.gamma
        gamma[i,m] ~ dnorm(mu.gamma[m],sigma.gamma[m])
        #the subject x previous response interaction
        #has group level mean mu.r and precision sigma.r
        r[i,m] ~ dnorm(mu.r[m],sigma.r[m])
        #the subject x previous stimulus interaction
        #has group level mean mu.r and precision sigma.r
        s[i,m] ~ dnorm(mu.s[m],sigma.s[m])
    }


    #uniform prior for the proportion of trials in which carryover occurs
    a.r[i] ~ dbeta(1,1)
    a.s[i] ~ dbeta(1,1)

}    #uninformative group level priors
    for(m in 1:NSTIM){
        b0[m] ~ dnorm(0,sigma.b0[m])
        mu.gamma[m] ~ dnorm(0,10)
        mu.r[m] ~ dnorm(0,10)
        mu.s[m] ~ dnorm(0,10)
        sigma.gamma[m] ~ dgamma(.001,.001)
        sigma.b0[m] ~ dgamma(.001,.001)
        sigma.r[m] ~ dgamma(.001,.001)
        sigma.s[m] ~ dgamma(.001,.001)
        sigma[m] ~ dgamma(.001,.001)
    }

}
```

Appendix G Multi-Interval SRLAG, JAGS Code

```
model{
for(i in 1:NSUBJ){
    #the first LAG trials are not influenced by the previous trials
    for(j in 1:LAG){#LAG is the number of previous trials considered
        #the data are categorical
        R[i,j] ~ dcat(theta[i,j,1:NRESP])
        #probability of subject i making a response between
        #criterion k and k-1, given mu[i,j]
        for(k in 2:(NRESP-1)){
            theta[i,j,k] <- max( 0 , phi((C[k]-mu[i,j])/sigma[S[i,j]])
            - phi((C[k-1]-mu[i,j])/sigma[S[i,j]]) )
        }
        #probability of subject i making the lowest response
        #given mu[i,j]
        theta[i,j,1] <- phi((C[1]-mu[i,j])/sigma[S[i,j]])
        #probability of subject i making the highest response
        #given mu[i,j]
        theta[i,j,NRESP] <- 1 - phi((C[NRESP-1]-mu[i,j])/sigma[S[i,j]])
        #b0 is the overall latent strength associated with
        #each stimulus value. gamma is subject x stimulus interaction
        mu[i,j] <- b0[S[i,j]] + gamma[i,S[i,j]]
    }
    #this section of code is the same as the last section
    #except this section takes into account the effect of the previous
    #stimulus up LAG trials back
    for(j in (LAG+1):NTRIALS){
        R[i,j] ~ dcat(theta[i,j,1:NRESP])
        for(k in 2:(NRESP-1)){
            theta[i,j,k] <- max( 0 , phi((C[k]-mu[i,j])/sigma[S[i,j]])
            - phi((C[k-1]-mu[i,j])/sigma[S[i,j]])   )
        }
        theta[i,j,1] <- phi((C[1]-mu[i,j])/sigma[S[i,j]])
        theta[i,j,NRESP] <- 1 - phi((C[NRESP-1]-mu[i,j])/sigma[S[i,j]])
        #if x.r1[i,j] = 1 then the previous resposne is taken into account
        #if x.r1[i,j] = 0 then the effect is 0
        #this same rule is applied to the effect of the previous stimulus
        #for x.s1[i,j]
        #r1[i,R[i,j-1]] is the effect of the previous response for each
        #subject i
        #s1[i,S[i,j-1]] is the effect of the previous stimulus for each
        #subject i
        mu[i,j] <- b0[S[i,j]] + gamma[i,S[i,j]]
        + x.r1[i,j]*r1[i,R[i,j-1]] + x.s1[i,j]*s1[i,S[i,j-1]]
        #if x.r2[i,j] = 1 then the resposne at lag, 2, is taken into account
        #if x.r2[i,j] = 0 then the effect of the lag 2 response is 0
        #this same rule is applied to the effect of the previous stimulus
        #for x.s2[i,j]
```

131

```
        #r2[i,R[i,j-1]] is the effect of the response at lag 2 for each
        #subject i
        #s2[i,S[i,j-1]] is the effect of the stimulus at lag 2 for each
        #subject i
        + x.r2[i,j]*r2[i,R[i,j-2]] + x.s2[i,j]*s2[i,S[i,j-2]]
        #the proportion of trials in which carryover occurs follows
        #a Bernoulli distribution with success parameter 1-a[i]
        x.r1[i,j] ~ dbern(1-a.r1[i])
        x.s1[i,j] ~ dbern(1-a.s1[i])
        x.r2[i,j] ~ dbern(1-a.r2[i])
        x.s2[i,j] ~ dbern(1-a.s2[i])
    }
    #priors at the individual level for each stimulus
    for(m in 1:NSTIM){
        #the subject x stimulus interaction
        #has group level mean mu.gamma and
        #group level precision sigma.gamma
        gamma[i,m] ~ dnorm(mu.gamma[m],sigma.gamma[m])
        #the subject x previous response interaction
        #has group level mean mu.r1 and precision sigma.r1
        r1[i,m] ~ dnorm(mu.r1[m],sigma.r1[m])
        #the subject x previous stimulus interaction
        #has group level mean mu.r1 and precision sigma.r1
        s1[i,m] ~ dnorm(mu.s1[m],sigma.s1[m])
        #the subject x lag 2 response interaction
        #has group level mean mu.r2 and precision sigma.r2
        r2[i,m] ~ dnorm(mu.r2[m],sigma.r2[m])
        #the subject x lag 2 stimulus interaction
        #has group level mean mu.r1 and precision sigma.r1
        s2[i,m] ~ dnorm(mu.s2[m],sigma.s2[m])
    }

    #uniform prior for the proportion of trials
    #in which carryover occurs
    a.r1[i] ~ dbeta(1,1)
    a.s1[i] ~ dbeta(1,1)
    a.r2[i] ~ dbeta(1,1)
    a.s2[i] ~ dbeta(1,1)

}   #uninformative group level priors
    for(m in 1:NSTIM){
        b0[m] ~ dnorm(0,sigma.b0[m])
        mu.gamma[m] ~ dnorm(0,10)
        mu.r1[m] ~ dnorm(0,10)
        mu.s1[m] ~ dnorm(0,10)
        mu.r2[m] ~ dnorm(0,10)
        mu.s2[m] ~ dnorm(0,10)
        sigma.gamma[m] ~ dgamma(.001,.001)
        sigma.b0[m] ~ dgamma(.001,.001)
        sigma.r1[m] ~ dgamma(.001,.001)
        sigma.s1[m] ~ dgamma(.001,.001)
        sigma.r2[m] ~ dgamma(.001,.001)
        sigma.s2[m] ~ dgamma(.001,.001)
        sigma[m] ~ dgamma(.001,.001)
    }

}
```