Modeling mindless choice

Seo Wook Choi, Peter Craigmile, Mario Peruggia and Trisha Van Zandt

The Ohio State University

Modeling mindless choice

Abstract

Experiment participants sometimes do not pay attention to the task that experimenters provide. To manage this problem, researchers often implement data preprocessing techniques to discriminate "good data" from "bad data". The most commonly used data preprocessing technique is a removal of data points. Such preprocessing is problematic for two reasons. First, deciding what is "bad data" is based on a researcher's subjective criterion. Second, "bad data" might be generated by an underlying process which is theoretically interesting, and knowing about that process might help to understand the data better. Discarding data without deep consideration, then, wastes time, money and potentially useful information. In this project, we are suggesting an alternative to discarding data. Instead of making rid of unexpected data, the "bad data" generating process should be modeled explicitly. The key concept behind this modeling is a Markov process which can describe choice dependency between trials. Using Bayesian hierarchical modeling, we are suggesting that stimulus-unrelated response processes can be separated from the stimulus-based response process, instead of discarding outlier data points.

Introduction

Determining causality between two variables requires a controlled experimental design in which one variable is independent, the other is dependent, and potentially confounding variables are controlled. Psychological experiments do not easily reveal the causal structures that researchers are interested in studying because it is difficult to control for all possible confounding variables when dealing with human beings with different backgrounds, experiences, motivations, and so forth. In particular, there is no guarantee that participants in an experiment will follow the directions provided to them by the researcher or focus on the task properly. This means that human experimental data is very messy. Measurements reflect the influence of these unknown confounding variables and the willingness or unwillingness of the participant to put forth his or her best efforts.

Because psychological data often seem to contain measurements that are inconsistent, extreme or uninterpretable, there has been a great deal of effort by psychological researchers on the best ways to deal with "bad data" (Zimmerman, 1994). The techniques proposed usually involve some kind of "data preprocessing." Data preprocessing is the transformation or removal of measurements to produce a data set that can be used to extract better information about the process of interest.

For example, human performance data consist of response times and response accuracies collected during the performance of a choice reaction task. During a choice reaction task, participants are shown a stimulus and required to select one of a small number of possible responses to that stimulus. For example, during a signal detection task, a participant may be shown a random field containing some number of dots and be asked to determine if that number of dots is "high" or "low." Perhaps the most common method of data preprocessing is the removal of outliers. An outlier might be a response time that is too fast or too slow, relative to the other response times provided by that participant. Or, at the level of the participant, his or her accuracy may be very low relative to the other participants in the study.

If a response time is very fast or very slow, say, more than 4 standard deviations from the participant's mean response time, or if the participant performs no better than chance, the researcher may decide to eliminate those very fast or very slow response times, or eliminate all the measurements provided by that participant. While throwing away data in this way may result in data that seem more interpretable (or even meaningful), it is problematic for at least three reasons.

First, discarding data necessarily wastes time and money. It is not always easy to find participants, bring them into the laboratory, and train them to do the necessary tasks. Research assistants and often the participants themselves must be paid for their time.

Second, in addition to the loss of time and money, throwing away data risks a loss of information. Participants who seem to be performing badly may do so because of the experimental conditions and not because they are failing to perform the task as required. Discarding "bad data" may eliminate data that might otherwise provide valuable information about the process being studied. A major goal of the research in this paper is to extract information even from "bad data," or at least isolate "bad data" from "good data," and so arrive at better methods of analysis, which in turn result in better clarification of the phenomena of interest.

Third, determining which data are to be discarded relies on ad hoc criteria chosen after the researcher has examined the data. Ad hoc procedures leave open the possibility that researcher bias may influence the results. Removing outliers, for example, reduces the variability of the measurements that have been retained, which in turn could decrease the *p*-value of the test statistic computed for inferential tests using those data.

Instead of the removal of outliers from data, this paper will suggest a new framework for the analysis of psychological data without data preprocessing. This framework is based on the notion that it is possible to model the process that generates "bad data" simultaneously with the cognitive process of interest. Such a modeling strategy allows us to isolate data from the process of interest and so arrive at a clearer understanding of that process.

To understand how we plan to accomplish this goal, imagine a participant in an experiment who ignores the instructions from the experimenter and does something unrelated to the task and the stimuli presented. In such a situation, that participant's responses should be unrelated to the stimuli presented, and we might expect that the participant's responses would look like a random sequence. However, the key to this framework is the fact that human participants cannot generate truly random response sequences. Previous studies show that, when participants try to generate completely random binary sequences, they produce sequences that usually show negative dependencies across responses (Bakan, 1960; Wagenaar, 1972) indicating that participants tend to alternate their responses at rates greater than chance.

Consider now a participant whose performance is no better than chance. There are two reasons why this might have occurred. The first reason is that the participant was not performing the task as instructed, and in this situation, we might expect to see sequential dependencies (either positive or negative) across responses. The second reason is that the participant was performing the task as instructed, but had difficulty and so performed badly. In this situation, we might expect to see fewer sequential dependencies across responses and more responses that depend on the presented stimuli. Our goal is to model the participant's response sequence as a mixture of responses, some from a model of the task (perhaps with parameters set to very low performance), and some from a process that describes the probabilities of repeating and alternating responses in a way that is independent from the stimulus. To do this, we will use a simple stochastic process, a Markov chain, that can produce response dependencies across trials (Taylor & Karlin, 1998).

A Markov chain process specifies that the probability of a particular response on a trial only depends on the response made on the previous trial but not on any earlier trials. If the stimuli are chosen randomly from trial to trial, and so the stimulus sequence has no sequential dependencies, any observed dependencies across responses must arise from a process unrelated to the stimuli. A Markov chain process might be a good way to describe this task–unrelated response process.

In contrast to the stimulus-independent process, there must also exist a cognitive process that permits responses executed as a result of processing the stimuli. For a two-choice task, signal detection theory (SDT) is a well-known model that is commonly used to describe response accuracy and response bias in simple decision-making (Green & Swets, 1966). An SDT process can describe the generation of responses during the performance of the task, and its parameters will determine the accuracy of the participant's responses, but these responses will depend only on the stimuli and will not show any sequential dependencies.

Under the model that this paper will suggest, these two processes, the Markov chain and the SDT process, are treated as two components of a dichotomous response process. On any trial, there is some probability that the executed response comes from either the Markov chain or the SDT process. This probability will be higher or lower for different participants, permitting us to detect which participants focus on the task and which do not. In short, using SDT and a Markov chain process, the response sequence is a sample from a mixture of these two different processes, one that is based on the stimulus and the other that is based on previous responses. This mixture model will form the basis of a Bayesian hierarchical model that can be fit to data from a magnitude judgment task. The hierarchical structure permits different participants to have different response strategies, as well as different performance levels for the task.

Background

Human performance data usually consists of two primary dependent measures: response accuracy and response time (RT). In the previous chapter, We outlined the problem this project addresses: cleaning these kinds of data to remove the effects of outliers. More generally, we seek a way to isolate data arising from the cognitive process of interest from those arising from other processes. In this chapter, We will first introduce the most common ways of handling outliers. It is important to note that most of these methods concern RTs but not accuracy. This approach, in contrast to these methods, is to develop a model that can separate cognitive-based from non-cognitive-based responses. The model, a hierarchical Bayesian structure, includes components consisting of Markov chains, signal detection theory (SDT), and probability mixtures.

There are three ways to deal with an observation that is suspected to be an outlier or a contaminant: 1) ignore it, and include it with the other observations for analysis; 2) remove it on the basis of some criterion usually based on the distance of the observation from some measure of central tendency; or 3) transform it so that its distance from the center of the observations is reduced.

Outlier removal is probably the most popular method of data preprocessing (Barnett & Lewis, 1994; Hawkins, 1980). The primary concern with outlier removal is to specify an objective criterion to which observations can be compared. This criterion is usually determined by the distance from the sample mean or median in standard deviations. This "restricted mean" method is the most well-known way to identify outliers. The method specifies the number of standard deviations (e.g., 2.0, 2.5, 3.0) a data point can fall from the mean before being considered as an outlier. All observations that fall outside this restricted range are excluded, and the restricted mean is computed from all observations within the restricted range. Although the restricted mean is a very useful and easy way to preprocess data, the standard deviation criterion is nonetheless arbitrary. It is impossible to determine that extreme data points are not generated by the process of interest. Furthermore, if arbitrary criteria are selected in a post hoc manner, then the criteria can be abused to produce results favorable to a researcher's hypothesis.

For human performance data, a special problem arises from the positively skewed nature of RT distributions. It is difficult to apply the outlier treatments, devised for data from skewed distributions, to normally- or symmetrically-distributed data. However, Ulrich and Miller (1994) warn that cutoffs based on standard deviations may cause biased estimates of the mean, because truncation may exclude extreme measurements from the tail of the distribution. In the case of RT distributions, which are skewed to the right, truncation results in greater bias for mean estimates. Instead of using the mean and standard deviation to establish a criterion, the median and interquartile range might be used. However, according to Ratcliff (1993), using the median and interquartile range lowers the statistical power of hypothesis testing.

Another way to manage outliers is a modified version of the cutoff method called winsorizing (Hastings, Mosteller, Tukey, & Winsor, 1947). Winsorized observations are those exceed a quantile cutoff and are replaced with the values of that cutoff (either upper or lower). However, windsorizing is problematic when the underlying distribution is asymmetric (Rivest, 1994).

Instead of removing or transforming outlier values, the entire data set can be transformed. These include transformation techniques such as inversion, logarithm, root-based, trigonometric-based and so forth. For data X, the inverse transformation replaces all observations with $1/X$ and the log transformation replaces all observations with $\log(X)$. The goal of such transformations is to produce distributions that are as symmetric or normal as possible. For positively-skewed measurements, the inverse transformation $1/X$ compresses the right tail into the left tail, usually between 0 and 1. Because the transformation is discontinuous at 0, it is not usually applied to measurements taking on both negative and positive values. The log transformation $\log(X)$ can only be applied to positive X values. This transformation compresses the right tail of the distribution and expands the left tail. Because response times are always positive, both transformations have been applied to reduce the influence of extreme values and to produce transformed RT distributions that are more symmetric.

Ratcliff (1993) recommended using inversion. He showed that this transformation reduces the effect of outliers on the mean without a loss of statistical power. Also, because 1/RT is speed, the transformed values are easy to interpret. However, transformation techniques also have inherent problems. The transformation may diminish interpretability. The inversion of time to speed is a rare case that only applicable to chronometric data. Also, because there is no standard transformation for various kinds of human performance data, it is difficult to say that a transformation that worked in past research will work for future research. Furthermore, if a transformation technique is chosen after examining the effects of several different transformations, the false positive rate may increase. Because of these problems, no single method for dealing with outliers is completely satisfactory. In this paper, we are suggesting a modeling method that uses all the data without data preprocessing.

Recent research by Craigmile, Peruggia, Van Zandt and colleagues has provided some alternatives to data preprocessing (Craigmile, Peruggia, & Van Zandt, 2010; Kim, Potter, Craigmile, Peruggia, & Van Zandt, 2016). Craigmile et al.'s research focused on human performance data consisting of accuracy and response times (RTs). RT data, in general, contains fast and slow responses that are far from the bulk of the observations and so they might be considered outliers; that is, observations that arise from processes different from the one under study. However, RT distributions are positively skewed (Van Zandt, 2002), so it is difficult to determine which slow RTs are outliers. Similarly, it can be difficult to determine which fast RTs are outliers. Rather than trying to identify contaminant observations with the goal of discarding them, Craigmile et al. have proposed to model them explicitly. Craigmile et al. (2010) presented a descriptive framework within which log RTs were modeled as an autoregressive process of order 1 (AR(1)). Each RT was assumed to arise from this process, but there was some probability that the RT was incremented or decremented by an amount that was exponentially distributed. Fluctuations in the log RTs over time were described using a wavelet regression model. With this structure, outliers and sequential effects across trials could be captured and separated from the (normal) log RT distribution. Embedding this structure within a Bayesian hierarchical model allowed them to explain both RTs from the cognitive process of interest and the outliers generated by a different process with a single framework.

The major problem with this approach was that the model on the log RTs was not theoretically motivated. In part to address this problem, Kim et al. (2016) analyzed RT and accuracy data from a (two-choice) recognition memory task as the output of a Poisson race model (Audley & Pike, 1965; Townsend & Ashby, 1965; Van Zandt, Colonius, & Proctor, 2000). The Poisson race model, an information accumulation model, has parameters for a rate of information accumulation and response thresholds.

Like Craigmile et al. (2010), Kim et al. assumed that fast and slow contaminant responses were mixed with the responses arising from the Poisson race process. Sequential effects were modeled as an AR(1) process on the log thresholds. Kim et al. showed that this model was able to identify contaminants and, in addition, by breaking the choice process into stimulus- and response-based components, characterize participants' poor performance in different, theoretically important ways. In particular, some participants performed poorly because they had trouble with the task (which showed up as process-relevant parameters that indicated poor memory encoding or retrieval), whereas others performed poorly because they did not focus on the task at all (which showed up as contaminant process parameters that overwhelmed the memory process).

The fact that a complete model needs to identify when people are or are not performing a task leads us to this project. In this project, we will focus only on the responses that people make in two- or three-choice tasks and not the RTs. To model these choice tasks, we will use a signal detection theory (SDT) framework. This framework will be embedded in a Bayesian hierarchical mixture model in which responses arise both from the SDT process and from a Markov chain process that describes decision making when a person is not paying attention. This approach exploits the idea that dependencies across responses should only arise when the participant is not paying attention (the Markov chain process), but there should be no dependencies when s/he is paying attention (the SDT process). we will explain this logic in more detail below.

General ideas of modeling

Where do outliers come from? Why would we want to remove them? The idea is that we have one process of interest that we are trying to study (the process by which stimuli are interpreted and appropriate responses to those stimuli are selected), and other processes that generate contaminant observations and make it difficult to draw conclusions from the data about the process of interest. We often ask experiment participants to perform tasks for close to an hour (or more), and we can't expect that these participants have perfectly sustained their concentration on the task for this length of time.

In this project, we will exploit the idea that contaminant responses should come from processes that are independent of the stimuli presented in the trials for which those responses were executed. The data collected over the trials in an experiment are therefore a mixture of stimulus-based and stimulus-independent responses. we will focus only on the responses and not the response times for this project.

In a two-choice task, if the stimuli are presented randomly (i.e., an independent and identically distributed sequence of stimuli) and participants' responses only depend on the stimulus, the responses should also be an independent and identically distributed sequence. If the responses are being generated by a process that does not depend on the presented stimulus, then they may or may not be independent and identically distributed. However, we have empirical evidence to suggest that responses from trials where participants are not paying attention are not independent and identically distributed. This evidence comes from two sources. First, in almost all experiments there are response dependencies between trials. For example, responses that are repeated are generally faster than responses that are not (Logan, 1990; Remington, 1969). Second, and the most important, people are unable to generate random sequences when they are asked to do so (Bakan, 1960; Wagenaar, 1972).

If responses to stimuli are independent from each other, and responses arising from stimulus-independent processes are not, then we can look for dependencies across trials to determine when people are paying attention and when they are not. From this idea, we are going to model two-choice response as a mixture of two different processes, one that depends on the stimulus and one that does not. Each component of the mixture will be based on two different models of the response process. we will describe each component and then integrate them into a mixture model that can be fit to data with a Bayesian hierarchical framework.

Because the stimuli are independent across trials, any dependency on the responses implies that there is some other process causing the dependence that is different from the process of interest.

This paper will focus on detecting dependency between trials under the assumption that such dependency cannot be attributable to the process of interest. Such dependencies can be modeled by a simple Markov chain process without requiring elaborate theories about the structure that produces the responses. Consider a sequence of responses $\{R_1, R_2, \ldots\}$ where R_i is a binary variable that takes on values 0 or 1. A Markov chain is a random process such that the probability that $R_i = r$ depends only on R_{i-1} but on no other preceding states $\{R_1, R_2, \ldots, R_{i-2}\}$ (Taylor & Karlin, 1998). The behavior of a two-state Markov chain is determined by a transition probability matrix such as

$$P_i = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} p_i & 1 - p_i \\ 1 - p_i & p_i \end{bmatrix} \end{array}, \tag{1}$$

where the subscript i indicates the individual. The number "0" represents a "noise" and "1" represents a "signal" response. The columns represent the current response state, and the rows represent the previous response state. If the previous state was 0, the probability that the next state is also 0 is p_i, and the probability that the next state is 1 is $1 - p_i$. If the previous state is 1, the probability that the next state is 1 is p_i, and the probability that the next state is 0 is $1 - p_i$. Therefore p_i is the probability that previous response is repeated. Modeling response selection with the Markov chain defined by Equation 1 assumes that the stimulus has no influence on the response, or that the participant is not paying attention to the stimulus. The symmetry of the matrix P_i represents a very simple model that explains only response repetitions and alternations. The symmetry of P_i reduces the number of model parameters so that the model can be fit to the data more easily.

To model response selection when participants are paying attention, we will implement Signal Detection Theory (SDT; Green and Swets (1966)). SDT provides an explanation of how people distinguish signals from noise. For example, suppose a doctor has to make a diagnosis of disease (leukemia) from a blood sample containing some number of white blood cells. She does not know whether the patient is sick or not. We call a blood sample from a sick patient a signal and a blood sample from a healthy patient noise. There is always variability in the sample. The number of white blood cells varies across people, so it is possible that a sample of healthy blood has a high number of white blood cells or sample from a sick person's blood may have a low number of white blood cells. So, signals and noise are not always discriminable.

The SDT model has two parameters to describe the response selection process. The difference between the means of the signal and noise distributions is d', or discriminability. High values of d' mean that a doctor can discriminate healthy blood from sick blood very well. If d' is close to zero it means that a doctor cannot diagnose disease from blood at all, so the decision is no different from a random guess. The second parameter c, the criterion, describes the intensity of the stimulus required to choose a "signal" response. High values of c mean that more evidence is needed to make a "signal" decision. A doctor with a high criterion will only make a diagnosis of sickness when the number of white blood cells is very high. Such high criteria are referred to as "conservative." Low criteria are referred to as "liberal."

We will write the presented stimulus as $S_{ik} = 0$ (noise) or 1 (signal) for individual i on trial k. Assuming that the distributions of noise and signals are Gaussian, and that the individual has criterion c_i and perceived discriminability d_i. we assume that the mean of the noise distribution is 0 and that the variance of both distributions is 1. Then, given the stimulus S_{ik}, the probability of making a noise response is $\Phi\left(c_i - S_{ik}d_i\right)$ and the probability of making a signal response is then $1 - \Phi\left(c - S_{ik}d_i\right)$ where Φ is the normal cumulative distribution function (CDF). We can put these probabilities into a transition matrix like that of Equation 1, yielding

$$\phi_{ik} = \begin{bmatrix} \Phi\left(c_i - S_{ik}d_i\right) & 1 - \Phi\left(c_i - S_{ik}d_i\right) \\ \Phi\left(c_i - S_{ik}d_i\right) & 1 - \Phi\left(c_i - S_{ik}d_i\right) \end{bmatrix}. \tag{2}$$

Comparing the two transition probability matrices in Equation 1 and 2, we see that the probability of a noise response in Equation 1 depends on the previous response whereas in Equation 2 it does not; the columns of the transition matrix ϕ_{ik} are identical.

Assuming that both processes contribute to the responses collected from an individual, for each trial there is a binary selection between the stimulus-based and stimulus-independent processes. The stimulus-based process is selected with probability ρ_i and the stimulus-independent process is selected with probability $1 - \rho_i$. This Bernoulli process integrates the two different models. The parameter ρ_i can be interpreted as the probability of properly responding to the stimulus as the experimenter intended, and $1 - \rho_i$ as the probability of doing something else. High ρ_i means that it is more likely that a decision is generated by the stimulus-based response process, and low ρ_i means that it is more likely that a decision is generated by the stimulus-independent response process.

In this study, a distinguishing point of the research is the mixture of the probabilities of two different response strategies. Bayesian hierarchical modeling can offer a framework for fitting this model to data. Bayesian hierarchical modeling is used to estimate the posterior distributions of multilevel parameters. In Bayesian statistics, because the parameters are also considered random variables, it is possible to estimate their posterior distributions using Bayes' theorem,

$$P(\theta \mid X) = \frac{P(X \mid \theta)\, P(\theta)}{P(X)}, \tag{3}$$

for data X and parameters θ. The probability P(θ) is the prior distribution of θ, P$(X|\theta)$ is the model likelihood, and P$(\theta|X)$ is the posterior distribution of θ, for individual i, $\theta_i = \{p_i, d_i, c_i, \rho_i\}$. All inferences about the parameters, both at the individual and the group-level, arise from the posterior distributions. Hierarchical models are used to analyze data that arise from variability at different levels of a process. For example, in this study, participants make decisions on each trial. Each decision is determined by individual traits. Those traits are determined by the group to which each individual belongs. Trial-level variability is governed by individual-level parameters, and individual-level variability is governed by group-level parameters (hyperparameters).

The posteriors of the parameters of complex hierarchical models rarely have analytic solutions. However, estimates of the posteriors can be obtained using Markov Chain Monte Carlo (MCMC) methods (Gelman, Carlin, Stern, & Rubin, 2014). MCMC methods are very flexible, and so very useful, especially because, mixture models can be very difficult to estimate. MCMC methods allow me to easily add more parameters and modify the model as needed.

Suggested Models

The model for Study 1

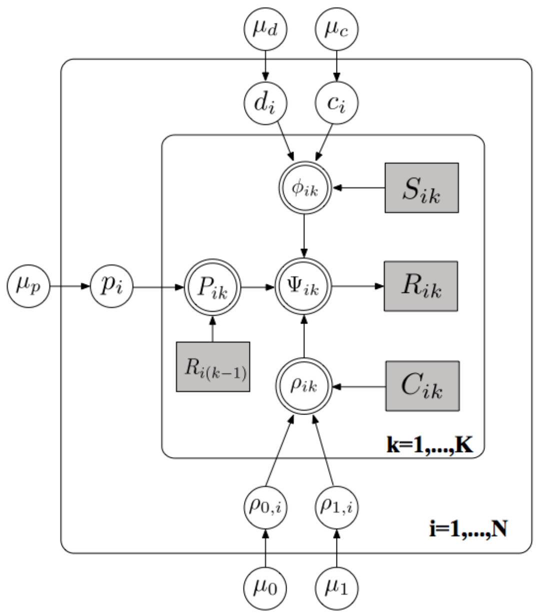

Figure 1. Bayesian hierarchical model of a mixture of stimulus-based and stimulus-independent response processes for a simple two-choice task.

The model assumes that there are two processes that generate the behavioral response. One is the stimulus-based response process and the other is the stimulus-independent response process described above. These two processes represent the cases when participants focus on the stimulus and when they fail to pay attention to the stimulus. The stimulus-based response process is modeled using SDT and the stimulus-independent process is modeled with the simple two-state Markov Chain. These two models will be integrated into a single Bayesian hierarchical framework.

This model has three levels that compose the hierarchy: the trial level k, the subject level i, and the group-level hyperparameters. In Figure 1, to distinguish between levels, there are two plates that separate the parameters. The innermost plate represents the trial level ($k = 1, ..., K$). Stimulus (S_{ik}) and condition (C_{ik}) are independent variables selected by the computer program that runs the experiment. Generation of the stimulus and condition will be explained in Chapter 4. The response R_{ik} follows a Bernoulli distribution with success probability Ψ_{ik}. The success probability Ψ_{ik} is a quantity computed from the trial-level quantities ϕ_{ik}, P_{ik} and ρ_{ik}.

The parameters between the inner plate and outer plate are the individual-level parameters d_i, c_i, p_i, $\rho_{0,i}$ and $\rho_{1,i}$. Note that ρ_{ik} is determined by condition C_{ik} and that, depending on condition, ρ_{ik} takes on either value $\rho_{0,i}$ or $\rho_{1,i}$. The parameters outside of the outer plate are the group-level hyperparameters μ_d, μ_c, μ_p, μ_0 and μ_1. We denote a "signal" decision as 1 and a "noise" decision as 0. The stimulus S_{ik} is either a signal (1) or noise (0) and the condition C_{ik} will be explained in the Methods section of Chapter 4. The probability of a signal response Ψ_{ik} is a mixture of the stimulus-based and stimulus-independent processes, so that

$$\Psi_{ik} = \rho_{ik}\phi_{ik} + (1 - \rho_{ik})P_{ik}. \tag{4}$$

The stimulus-based response probability ϕ_{ik} arises from signal detection theory, and so

$$\phi_{ik} = 1 - \Phi(c_i - S_{ik}d_i). \tag{5}$$

The value of ϕ_{ik} depends on the criterion (c_i), discriminability (d_i) and stimulus type (S_{ik}), and does not depend on the previous decision. Due to the identifiability issue, we decided to constrain $c_i = \frac{d_i}{2}$. This also applies to Study 3. However, the probability of a stimulus-independent response,

$$P_{ik} = 1 - p_i + R_{i(k-1)}(2p_i - 1), \tag{6}$$

depends on the previous response. The quantity P_{ik} is probability of making a "signal" decision whereas p_i is the probability of repeating the previous response. We will parameterize the mixture probability as

$$\rho_{ik} = \frac{1}{(1 + \exp(\rho_{0,i} + C_{ik}\rho_{1,i}))}. \tag{7}$$

As we will explain in the next chapter, the experimental condition will be designed to differently motivate participants to pay attention. As we motivate participants more or less with different C_{ik}, we expect to see ρ_{ik} increase or decrease.

For computational convenience we choose normal priors for the individual-level parameter

$$d_i \sim N(\mu_d, 1). \tag{8}$$

For the stimulus-independent Markov process, we choose logit-normal priors so that

$$\begin{gathered} p_i \sim logitN(\mu_p, 1),\ \rho_{0,i} \sim N(\mu_0, 1),\ \rho_{1,i} \sim N(\mu_1, 1), \\ \rho_{0,i} = \frac{1}{(1 + \exp(\rho_{0,i}))} \text{ and } \rho_{1,i} = \frac{1}{(1 + \exp(\rho_{0,i} + \rho_{1,i}))} \end{gathered} \tag{9}$$

Finally, for all the hyperparameters,

$$\mu_1,\ \mu_0,\ \mu_p,\ \text{and } \mu_s \sim N(0, 1). \tag{10}$$

From decades of previous work with the SDT model, the value of the hyperparameter μ_d is expected to be anywhere from slightly below zero to a maximum around 3 or 4. The $N(0, 1)$ prior provides values that are consistent with this prior knowledge. The parameters μ_p, μ_0 and μ_1 will be transformed by the logit function, so the $N(0, 1)$ prior gives more probability weight to values close to 0.5.

Theoretically, there is no reason to use normal priors: any distribution that has support consistent with our prior knowledge could be used. However, the normal distribution is computationally convenient and promotes good convergence of chains produced using Markov chain Monte Carlo (MCMC) techniques.

We will use the estimates of the posterior distributions of these hyperparameters to measure the effects of experimental conditions across different participants. Parameters of interest are the tendency to repeat the previous response μ_p, the mean discriminability μ_d, and the probability of using the stimulus-based response process ρ, which is a function of μ_1 and μ_0. Each hyperparameter controls the behavior of the individual-level parameters p_i, d_i, c_i, $\rho_{0,i}$ and $\rho_{1,i}$.

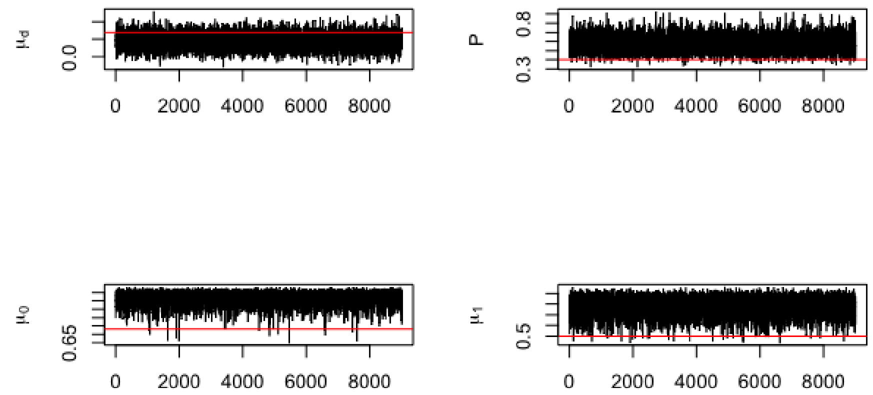

Figure 2. The trace plots for of 9000 samples (3000 iterations with 3 chains each) from posterior distribution of hyper-parameters. The red are parameters used to generate the simulation data.

Parameter Recovery. Parameter recovery is conducted similar manner with Study 1. Same parameters are used to generated simulation data ($\mu_0 = 0.7311$, $\mu_1 = 0.5$, $d = 0.7$, and $p = 0.4$). Similar to Study 1, the parameter P, mu_0 and mu_1 does not accurately recovered the parameters. Still there is a pattern that μ_0 is higher than μ_1 which indicates the higher probability of using stimulus based response process in high motivation condition. (Figure 2)

The model for Study 2

The task in Study 2 is a three-choice task. Therefore, stimulus-based response process was modeled differently from Study 1. The three possible responses ($R_{ik} = 1$, 2 and 3) are distributed as

$$R_{ik} \sim Multinomial(\Psi_{ik1}, \Psi_{ik2}, \Psi_{ik3}), \tag{11}$$

where the three parameters Ψ_{ikr} for the multinomial probabilities are mixtures of the stimulus-based response process ϕ_{ikr} and the stimulus-independent response process P_{ikr}. Regardless of the modification in the model both in the stimulus-based and stimulus-independent components, the mixture model works in the same way as that of

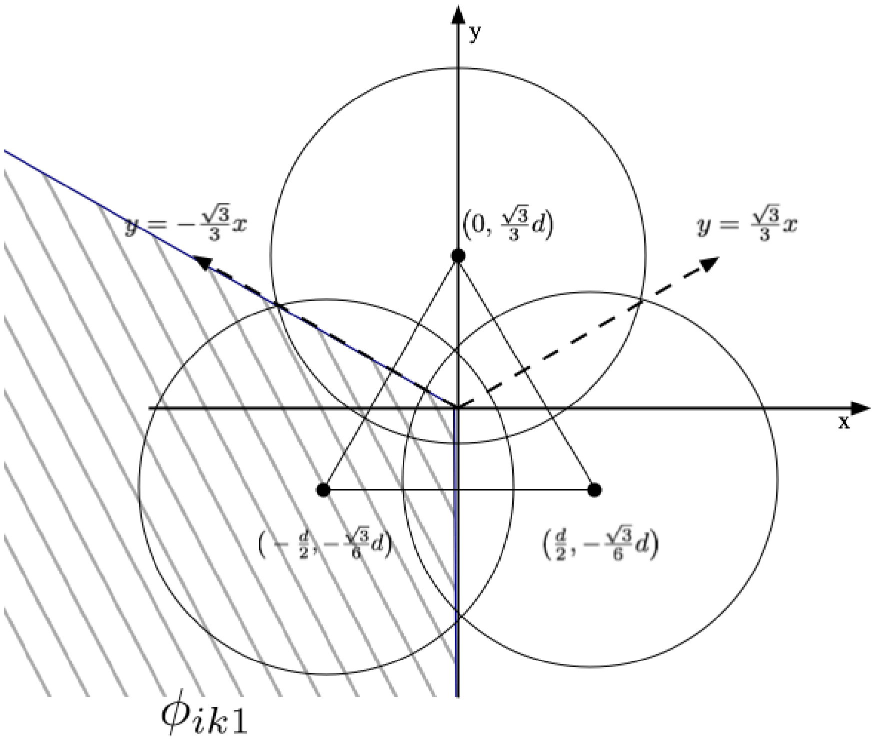

Figure 3. The circles represent bivariate normal distributions of the perceived stimuli. The probability ϕ_{ikr} is the volume under the bivariate normal distribution up to the appropriate linear criteria. For example, the shaded volume is ϕ_{ik1}.

Study 1. The response probabilities ϕ_{ikr} for the stimulus-based response process are given by the volume under a bivariate normal density up to the perceived boundary lines (criteria) between three different perceived stimulus representations (Figure 3). For the stimulus $S_{ik} =$ 1,2 or 3, $P(R_{ik} = 1, s)$ can be written as $\phi_{ik1,s}$, calculated as

$$\phi_{ik1,s} = \int_{-\infty}^{0} \int_{-\infty}^{-\frac{\sqrt{3}}{3}x} BVN\Big((c_{xs}, c_{ys}), I_2\Big) dy dx. \tag{12}$$

where $-\frac{\sqrt{3}}{3}x$ represents one of the boundary lines oriented of 120° and BVN is a bivariate normal distribution with mean vector (c_{xs}, c_{ys}). The matrix I_2 is a 2×2 identity matrix which represents variances of 1 and covariance of 0. For each $S_{ik} = s$, the mean vectors are

$$(c_{x1}, c_{y1}) = (-\tfrac{d}{2}, -\tfrac{\sqrt{3}}{6}d), \; (c_{x2}, c_{y2}) = (0, \tfrac{\sqrt{3}}{3}d), \text{ and } (c_{x3}, c_{y3}) = (\tfrac{d}{2}, -\tfrac{\sqrt{3}}{6}d), \tag{13}$$

where the parameter d represents the distance between the mean vectors of each perceived stimulus representation. However, there is no closed form solution for Equation 12. To approximate this integral, I used a a sampling technique. I sampled pairs of x and y coordinates (a_{xs}, a_{ys}) such that

$$(a_{xs}, a_{ys}) \sim BVN\Big((c_{xs}, c_{ys}), I_2\Big). \tag{14}$$

After obtaining N samples, I counted the number of coordinate pairs falling in each region defined by the linear criteria. The proportions of the number of coordinate pairs falling in each region is an approximate value of the integral in Equation 12. So, for example, for the case of $P(R_{ik} = 1|S_{ik} = s)$, $\phi_{ik1,s}$ can be approximated as,

$$\phi_{ik1,s} \xrightarrow{p} \tfrac{1}{N}\sum_{j=1}^{N}(\mathbb{1}_{a_{xsj}<0})(\mathbb{1}_{a_{ysj}<-\frac{\sqrt{3}}{3}a_{xsj}}), \tag{15}$$

where $\mathbb{1}$ is an indicator function. In Figure 3, the shaded area (actually volume) is equivalent to the probability of the randomly sampled coordinates falling inside the boundary lines. As N increases, the calculation becomes more accurate, but also increases computational load. Calculating this function value for every iteration in MCMC procedure is too slow to implement. There is an approximation of polynomial function with N=1000000 and $d = 0$ to $d = 3$ by 0.01 for ϕ. The approximate function which is used in the analysis is

$$\phi_{ik1,s} \approx 0.33333 + 0.39455(d/2) + 0.11330(d/2)^2 - 0.11426(d/2)^3 + 0.01773(d/2)^4. \tag{16}$$

Computing of $\phi_{ik2,s}$ and $\phi_{ik3,s}$ can be done in a similar manner to Equation 15. If the value of d is high, then there is high discriminability between the different stimuli. The stimuli S_{ik} 1,2 or 3 are randomly selected with equal probability on each trial.

The stimulus-independent process is modeled similarly to Study 1. This model has two possible outcomes, repeating the previous decision or not. Assume that the transition probability P_{ik} is given by

$$P_{ik}|(R_{ik} = R_{i(k-1)}) = p_i \text{ and } P_{ik}|(R_{ik} \neq R_{i(k-1)}) = \tfrac{1-p_i}{2}, \tag{17}$$

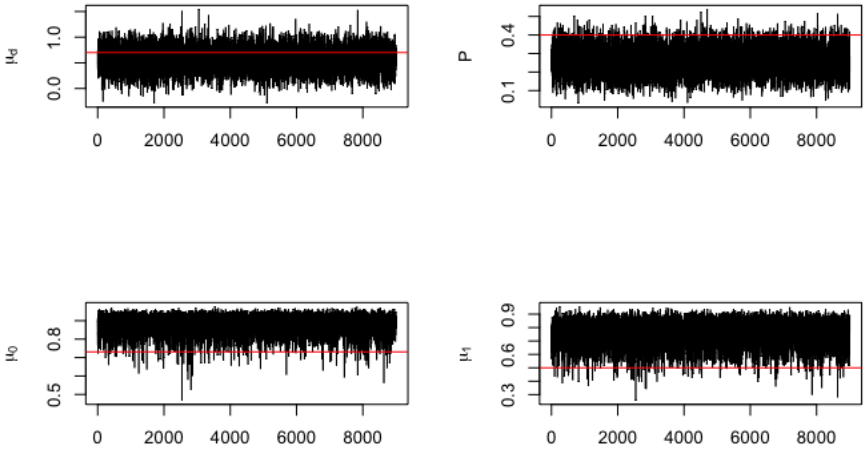

Figure 4. The trace plots for of 9000 samples (3000 iterations with 3 chains each) from posterior distribution of hyper-parameters. The red are parameters used to generate the simulation data.

so that the two decisions that are different from the previous decision have the same probability. The three parameters for the multinomial probability parameters, Ψ_{ikr}, are then mixtures of the stimulus-based response probability and the stimulus-independent response probability. They can be written as

$$\Psi_{ikr} = \rho_{ik}\phi_{ikr} + (1 - \rho_{ik})P_{ik} \tag{18}$$

which is very similar structure with that of Study 1.

Parameter Recovery. Parameter recovery is conducted similar manner with Study 1. Same parameters are used to generated simulation data ($\mu_0 = 0.7311$, $\mu_1 = 0.5$, $d = 0.7$, and $p = 0.4$). The parameter that we want to isolate d' is successfully recovered. However, the parameter P, mu_0 and mu_1 does not accurately recovered the parameters. Still there is a pattern that μ_0 is higher than μ_1 which indicates the higher probability of using stimulus based response process in high motivation condition. (Figure 4

To test the identifiability of the model, parameter recovery has beend conducted. The data to fit the model is generated with fixed parameters, $\mu_0 = 0.7311$, $\mu_1 = 0.5$,

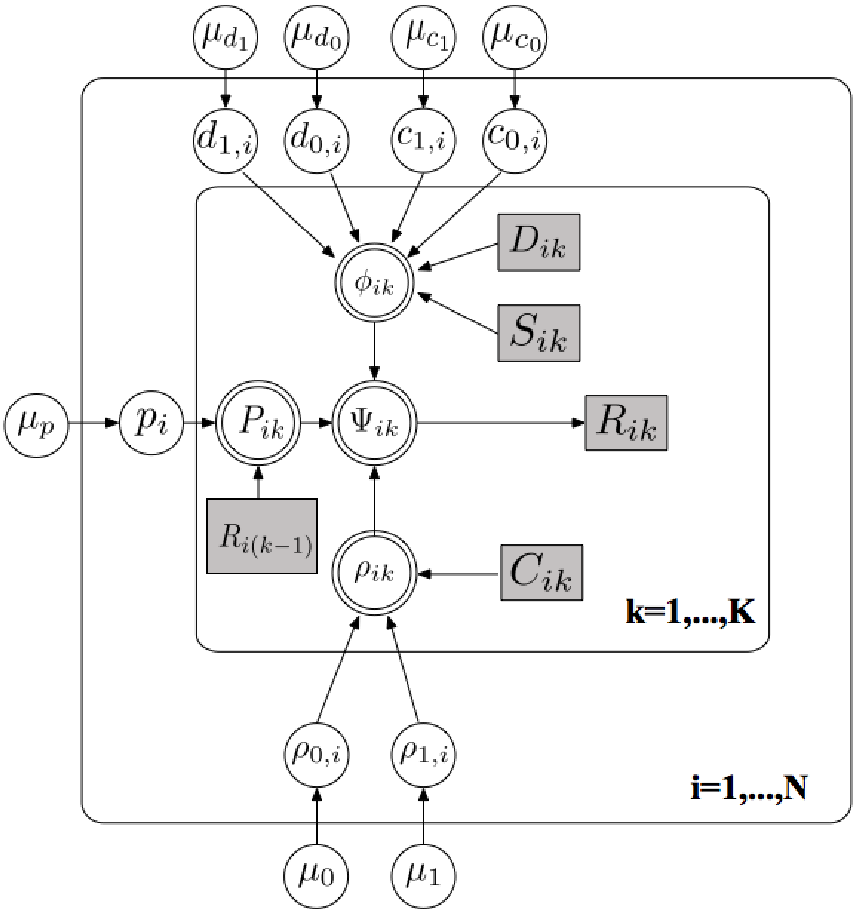

Figure 5. Bayesian hierarchical model for Study 3. The variable D_{ik} indicates task difficulty, where $D_{ik} = 1$ is difficult and $D_{ik} = 0$ is easy.

$d = 0.7$, and $p = 0.4$. Figure 4 is posterior using same data generating function. The most important parameter in this model mu_d seems to be recovered correctly. However, estimation in mu_0 and mu_1 has some bias. Still there is pattern of high and low μ which indicates the probability of using stimulus based response process.

The model for Study 3

In Study 3, we added a component to the stimulus-based response process to account for level of difficulty. Because there are two levels of difficulties in Study 3, we added two new parameters representing discriminability in the second difficulty condition (see Figure 5). The modified parameters from Study 1 are

$$d_{1,i} \sim N(\mu_{d_1}, 1) \text{ and } d_{0,i} \sim N(\mu_{d_0}, 1) \quad (19)$$

for individual i and hard (0) easy (1) discriminability conditions. The other components of the model are identical to that of Study 1. We expect the model to discriminate different difficulty levels by producing different mean discriminabilities μ_{d_0} and μ_{d_1}. Simultaneously, regardless of task difficulty, we should see the same effects of motivation on the mixture probability ρ. As in Study 1, we expect to see high ρ in the high motivation condition, suggesting that people tend to pay more attention to this condition than in the low motivation condition, regardless of any change in the task difficulty over trials. As the model is very similar to Study 1, there is no parameter recovery process for Study 3.

Experiments and results

To test the models that we introduced in Chapter 3, we designed three experiments to fit the models to data. We told participants that they would make numerosity or categorization judgments under two conditions that would be signaled by displays of different colors. Green "experimental" trials would be used to measure their performance and red "recalibration" trials would be used to adjust the computer. We asked the participants to do the task in the same manner in both conditions. We expected participants to pay more attention to the experimental condition than in the recalibration condition. This, in turn, should result in a higher probability of using the stimulus-based response process in the experimental condition.

Study 1

Participants. Thirty-seven participants were recruited from the Ohio State University Research Experience in Psychology participant pool and received partial course credit for their involvement. Data from one participant who left the experiment in the middle of the session was eliminated, so data from 36 participants were used in

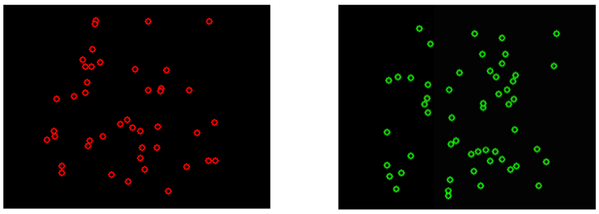

Figure 6. Stimuli used in the numerosity judgment task. The green and red stimulus conditions inform participants whether the current trial is a recalibration trial (red) or an experimental trial (green).

the analysis. All participants were naive to the purposes of the experiment and all reported normal or corrected-to-normal vision.

Materials. The experiment was run on a PC-style microcomputer running Ubuntu 14.04. Stimuli, stimulus presentation, timing and response collection were controlled by scripts written using the OpenSesame application (Mathot, Schreij, & Theeuwes, 2012). Stimuli were arrays of opaque circles of radius 4 pixels presented within a square measuring 1280×1024 pixels on each side. The number of circles presented on each trial was determined by selecting a random number from a normal distribution rounded to the nearest integer. The location of each circle was determined by selecting from continuous bivariate uniform distributions defined over the 1280 by 1024 pixels in each square resulting in a display of dot patterns similar to those shown in Figure 6. Participants responded to each stimulus by pressing "z" or "/" on a computer keyboard.

Design. The experiment was a 2×2 (Condition by Stimulus) repeated measures factorial design. The two conditions (C_{ik}), high and low motivation corresponding to the "experimental" and "recalibration" trials, respectively, were blocked over sets of 24 trials. The condition alternated between blocks, with the first block consisting of low motivation trials ($C_{ik} = 1$) followed by a second block of high motivation trials ($C_{ik} = 0$). There was a total of 48 blocks, 24 blocks in each condition, for a total of 1056 trials for each participant.

The stimulus types (S_{ik}) were the noise and signal stimuli of SDT. Letting $S_{ik} = 1$

represent signals and $S_{ik} = 0$ represent noise, the numbers of dots were generated from normal distributions with means equal to 47.5 and 52.5 and standard deviation equal to 5. The stimulus types were decided on each trial by randomly selecting a value for S_{ik} with probability 0.5. On average, there were 12 signal and noise trials in each block, and 528 signal and noise trials for each participant.

Procedure. Participants first gave their informed consent. Then they read the experiment instructions on the monitor while the most important parts of the instructions were read aloud by the experimenter. The participants were instructed to determine if the number of circles displayed on each trial was more than 50 or less than 50, and to press "/" on the computer keyboard if the number was greater than 50 and "z" otherwise. The participants were informed that some displays would be red and some would be green. They were told that red trials would be used to recalibrate the computer, whereas green trials would be used to measure their performance, but that they should perform the task the same way in both conditions.

Each trial began with the presentation of the dot display, which was terminated by a response. If the response was incorrect, an 'X' was presented in the center of the screen for 100ms, followed by a blank (black) screen for 200ms. If the response was correct, a blank screen was presented for 300ms. The trial series began with 40 practice trials. During the practice trials, the experimenter stayed with the participants to answer questions as needed. After the practice trials, the experimenter gave final instructions and the participant was left alone to perform the remaining 1056 trials.

Calculation of d'. From each participant, we obtained measures of stimulus (S_{ik}), condition (C_{ik}) and response (R_{ik}) over the 1056 trials of the experiment. We performed no data preprocessing. We first examined the effect of the motivation manipulation by computing the discriminability statistic d' for each participant and motivation condition. Discriminability is estimated as

$$d' = \Phi^{-1}(\text{hit rate}) - \Phi^{-1}(\text{false alarm rate}), \tag{20}$$

where Φ^{-1} is the inverse of the normal cumulative distribution function. The hit rate is the probability of making a correct decision when a signal was presented. The false

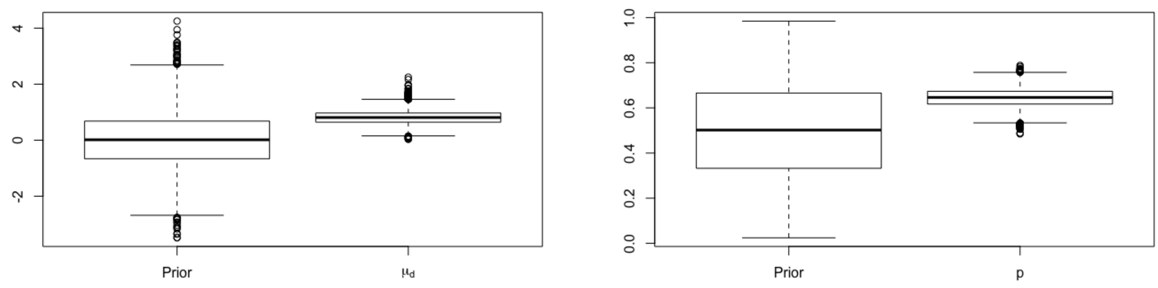

Figure 7. Prior and posterior distributions of model parameters. The left panel shows the prior and posterior distributions mean discriminability parameter μ_d. The right panel is the prior and posterior distributions of the probability of repeating the previous response.

alarm rate is the probability of making an incorrect decision when a noise was presented. Discriminability was higher in the high-motivation condition (mean $d' = 0.224$, $s.e. = 0.031$) than in the low motivation condition (mean $d' = 0.172$, $s.e. = 0.027$), a difference that was not significant at α equal to 0.05, [$t(35) = 1.5055$, p=0.1412]. The difference is, however, in the expected direction.

Hierarchical Bayesian Analysis. The model described in Suggested Models section were implemented in Stan (Gelman, Lee, & Guo, 2015). Stan is a probabilistic programming language that uses No U-Turn Sampling (NUTS) algorithm (Hoffman & Gelman, 2014). After a burn-in period of 300 samples, we obtained 3 chains of 3000 samples, total of 9000 samples from the posteriors of each parameter. Convergence was confirmed by the Gelman-Rubin convergence statistics $\hat{R}$ (Gelman & Rubin, 1992): every set of parameter chains had $\hat{R} \approx 1$.

The estimated posteriors of the hyperparameters are shown in Figure 7. The left panel shows the estimated posteriors of the mean discriminability parameter together with their prior. The right panel shows the mean probability of repeating the previous decision together with its prior distribution. The mean discriminability parameter μ_d has a posterior mean of 0.8175 and posterior 95% highest density interval (HDI) of [0.357 ,1.339]. This is much higher than calculation of d' without mixture model of

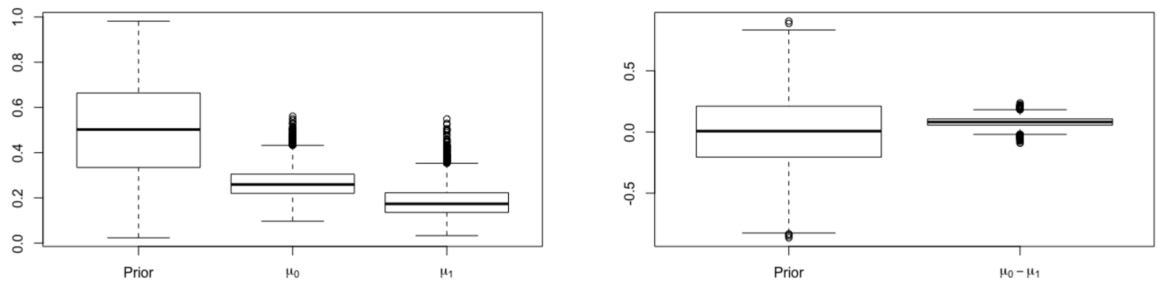

Figure 8. The left panel shows the posterior distributions for probability of using stimulus-independent response in high motivation condition (μ_0) and low motivation condition (μ_1) with its prior distribution. The right panel shows the posterior distribution of $\mu_0 - \mu_1$ in Study 1.

stimulus based and stimulus independent response process. The probability of repeating the previous response, $1/(1 + \exp(\mu_p))$ has an estimated a posterior mean of 0.645 and 95% HDI of [0.563, 0.722].

The posterior mean of μ_0 in the high motivation condition is 0.2656 and the 95% HDI is [0.150, 0.395]. The posterior mean of μ_1 in the low motivation condition is 0.184 and the 95% HDI is [0.07, 0.318]. More importantly, the posterior distribution of the difference between values of μ for the two conditions has a mean of 0.0812 and the 95% HDI of [0.001, 0.156].

The beauty of Bayesian hierarchical modeling is that it is possible to analyze effects at both the group level and the individual level. Figure 9 shows the estimated posterior distributions of $\rho_{0,i} - \rho_{1,i}$, which reflects the effect of motivation for each participant i. All but two participants had estimated posterior difference distributions with medians greater than zero.

For demonstration purposes, the analysis of individual participant was conducted Figure 10 show the estimated posterior distributions for Participant 15. Participant15 was greatly influenced by the motivation manipulation. The estimated posterior distribution of $\rho_{0,30} - \rho_{1,30}$ had a mean of 0.5472 with 95% HDI of [0.3934, 0.6901] and d_{30} had a mean of 0.5021 with HDI of [.1900, 0.8347]. The participant tend to focus

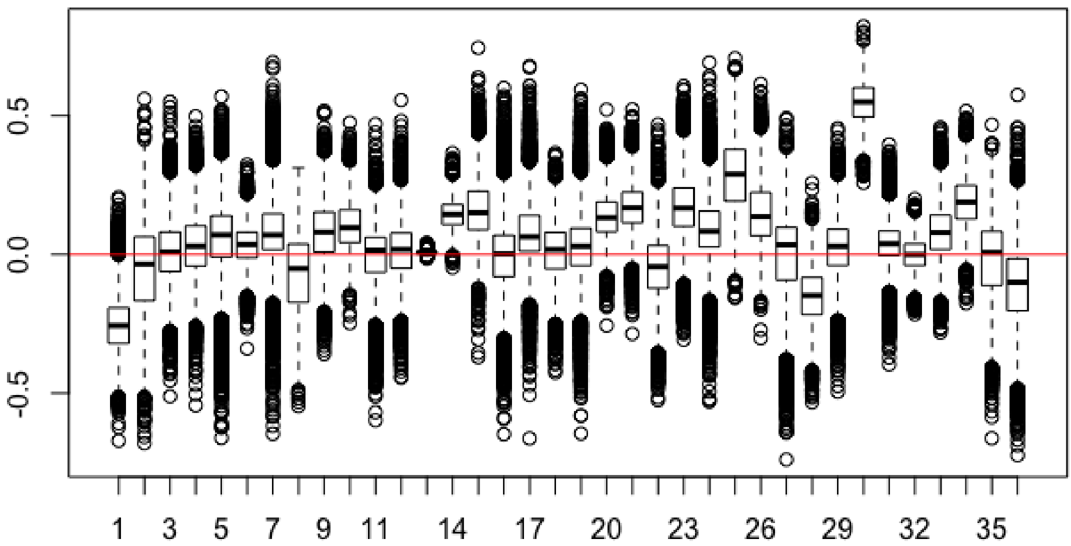

Figure 9. Posterior distributions of $\rho_{0,i} - \rho_{1,i}$ for each of 36 participants in Sutdy1.

more on motivation condition with okay discriminability.

Study 2

In this experiment, we generated categorization data to test a version of the model expanded to account for performance in a three-choice task.

Design. Study 2 is a three-choice expansion of Study 1. The new task asks participants to identify stimuli that arise from one of three different regions in stimulus array (Figure 11). A bird appears from one of the regions, Z, C or X. These regions are not identified for the participant except during the instruction phase of the experiment. The stimuli are constructed in such a way that the geographic centers of each region are equidistant and the angles of separation as defined by optimal response criteria are also equal. These equivalences allow me to retain the same parameterization of the stimulus-independent response process as in Study 1.

On each trial, the stimulus is selected with equal probability (1/3) from one of the three regions. Using the central point of each region as a mean, the stimulus's X and Y

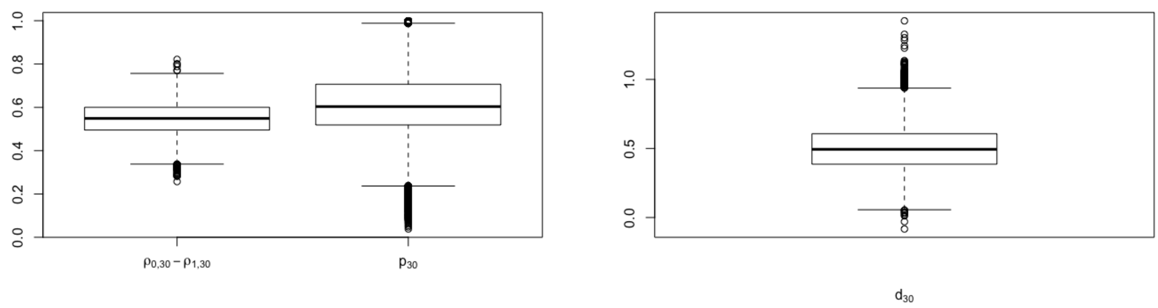

Figure 10. Posterior distributions of model parameters for Participant 15 in Study 1.

coordinates are generated from bivariate normal distributions with variance of 1. The mean vectors of each distribution are

$$(x_1, y_1) = (-2.037\cos(\pi/4), -2.037\cos(\pi/4)), \tag{21}$$

$$(x_2, y_2) = (2.037\cos(15\pi/180), 2.037\sin(15\pi/180)), \text{ and} \tag{22}$$

$$(x_3, y_3) = (-2.037\cos(75\pi/180), -2.037\sin(75\pi/180)). \tag{23}$$

Because a stimulus can be presented closer to the center of another region than to the center of its own, there is a possibility of error. The trigonometric terms in Equation 21∼23 works to keep angles of 120° between the mean vectors of each pair of stimulus. We selected the radius of 2.037 is selected to maintain approximately the same error rate as in Study 1 (about 30 percent). We conducted this experiment with two different motivation conditions (bird colors): red birds indicated recalibration (low motivation) trials and green birds indicated experimental (high motivation) trials. Including the motivation condition, the experiment was conducted in the same way as that of Study 1.

Method. Forty-two participants were recruited from participant pool at the Ohio State University. Four of the participants were eliminated because they did not finish the task. So, data from 38 participants were used for analysis. Other details were same as Study 1.

Calculation of d'

The d' is calculated by using correct response rate using polynomial. It is the discriminability without consideration of sequential dependency Mean d' is 0.9320

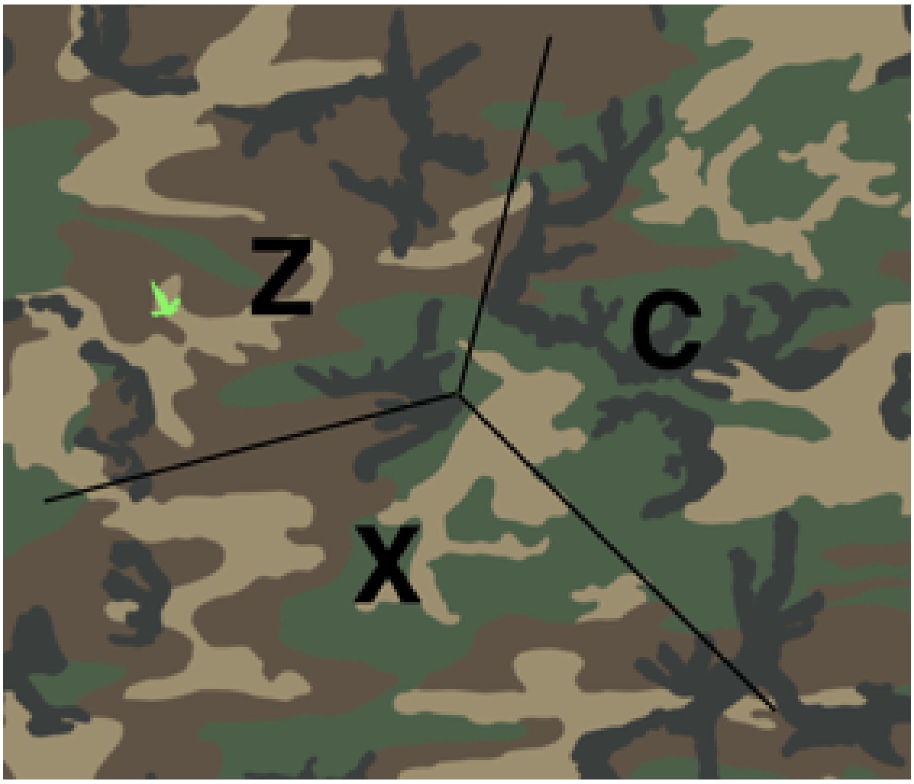

Figure 11. The task instruction for Study 2. This figure is presented to participants to understand the task. In this case, 'Z' will be the best decision. Area name Z, C, X and boundary lines are presented only for the instruction.

(s.e.=0.0532). The maximum of discriminability among subject was 1.3384 and minimum was 0.1829

Hierarchical Bayesian Analysis. The posterior mean of discriminabilty is μ_d=1.3695 and posterior 95% HDI is [1.027, 1.843]. The probability of repeating the previous response, $1/(1 + \exp(\mu_p))$'s posterior means is 0.526 and HDI of [0.426, 0.630]. There is no condition effect found that the ρ for high and low condition effect were similar [μ_0=0.723 μ_1=0.69]. It seems that condition effect did work in the study 3. The individual effect of participants 3 shows better performance on the task.

The Participant 14 is picked for individual analysis. The posterior mean of d_{14} is 1.0241 with HDI of [0.3995, 1.9701]. The probability of repeating previous decision, $p_1 4$ is 0.5867 with HDI of [0.4531, 0.8235]. The difference of $rho_{01}4$ and $rho_{11}4$ has posterior mean of 0.3201 with HDI[0.0608, 0.6316].

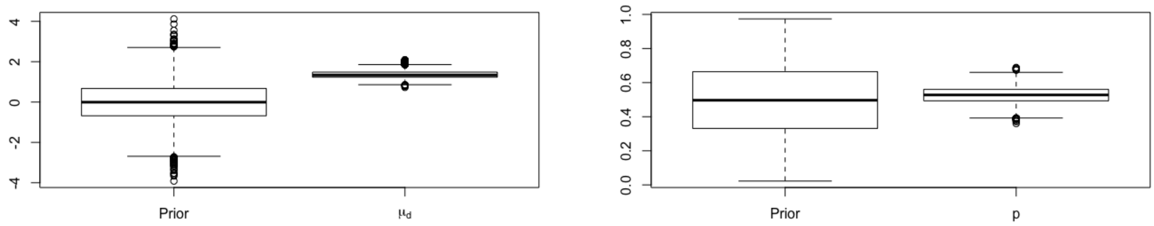

Figure 12. Prior and posterior distributions of model parameters μ_d (left panel) and p (right panel).

Study 3

The purpose of Study 3 was to verify that the model behaves appropriately under manipulations of task difficulty.

Design. We modified the design of Study 1 using two levels of difficulty (easy and hard). Blocks of 24 low and high motivation trials were nested within difficulty. Easy and hard super-blocks of 48 trials each repeated in turn 11 times, giving 1056 trials in total. The distance between the mean of signal and noise distribution of the easy blocks was increased to 2 standard deviations. The difficult blocks have the same distance (1 standard deviation) as Study 1. In the easy blocks, the number of dots on the screen for the noise stimulus was generated from a $N(45, 5)$ distribution, and for the signal stimulus was generated from a $N(55, 5)$ distribution.

Method. Fifty participants were recruited the same way as in Study 1. Data from one participant who did not finish the task was eliminated, so data from 49 participants were used for analysis. Participants were not informed of the difficulty manipulation.

Calculation of d'. From each participant, we obtained measures of stimulus (S_{ik}), condition (C_{ik}), response (R_{ik}) and difficulty (D_{ik}) over the 1056 trials of the experiment. We performed no data preprocessing. First, we conducted a two-way repeated-measures analysis of variance to examine the two main effects (motivation and difficulty) and their interaction on d'. The analysis showed no significant interaction between motivation and difficulty ($F(1, 192) = 0.0026$, $p = 0.9592$) and no main effect

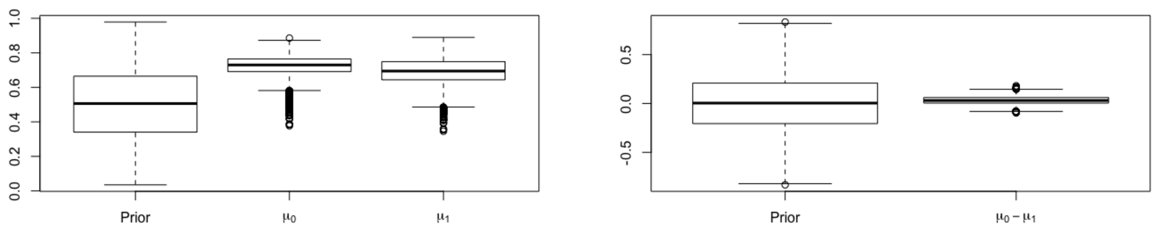

Figure 13. The left panel shows the posterior distributions for probability of using stimulus-independent response in high motivation condition (μ_0) and low motivation condition (μ_1) with its prior distribution. The right panel shows the posterior distribution of $\mu_0 - \mu_1$ in Study 2.

of motivation ($F(1, 192) = 1.8950$, $p = 0.1702$). There was, however, a significant effect of difficulty ($F(1, 192) = 48.1937$, $p < 0.001$). Table 4.2, shows the mean d' for the difficulty and motivation conditions. There was a tendency for participants to perform better in the high motivation condition, but the effect did not reach significance.

Hierarchical Bayesian Analysis. In Figure 16, the left panel shows the estimated posteriors of the mean discriminability for each difficulty level together with their prior distribution. The right panel shows the mean probability of repeating the previous decision together with its prior distribution. The mean discriminability parameter for easy task μ_{d_1} has a posterior mean of 1.862621 and a posterior 95% HDI of [1.272, 2.438]. The mean discriminability parameter for the difficult task μ_{d_0} has a posterior mean of 0.8025 and a posterior 95% HDI of [0.461, 1.168]. This demonstrates that the difficulty level had the expected effect: discriminability is higher in difficult task condition over easy task condition.

The probability of repeating the previous response p has an estimated posterior mean of 0.651 and 95% HDI of [0.586, 0.721].

The parameter of greatest interest is the mixture probability μ which should have been affected by the motivation manipulation. For the high motivation condition, μ_0 has a posterior mean of 0.333 and a 95% HDI is [0.242, 0.436]. For the low motivation condition, μ_1 has a posterior mean of 0.254 and 95% HDI of [0.156, 0.360]. More

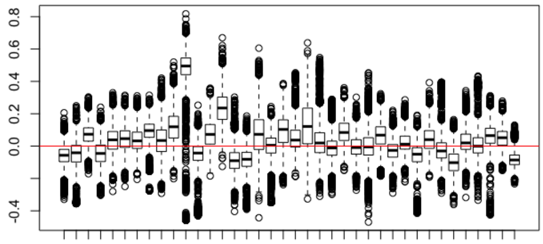

Figure 14. All individual's difference in ρ_0 and ρ_1 in Study 2.

importantly, the posterior difference between values of μ under the two conditions has a mean of 0.092 and a 95% HDI of [0.013 0.149].

Figure 18 shows the estimated posterior distributions of $\rho_{0,i} - \rho_{1,i}$, for each participant i. For demonstration, in the same way to Study 1, we picked two individuals whose results seemed interesting. Figure **??** shows the estimated posterior distributions for Participant 9's parameters. The estimated posterior distribution of the difference in ρ between high and low motivation conditions has a mean of 0.1625, and 95% HDI of [-0.0094, 0.3416]. This indicated that s/he had no big difference in two different conditions. Participant 9 showed pretty good discrimination in both conditions. (Posterior mean of $d_{0,9}$=1.8626 with HDI [1.2721, 2.4381] and $d_{1,9}$=0.8025 with HDI [0.4611, 1.6789]) Participant 9 tended to repeat his/her responses (posterior mean p_9 of 0.6513 and 95% HDI of [0.5856, 0.7209].

Discussion

Over three studies, we built mixture models for simple choice composed of two different processes. One process generates, with probability $1 - \rho$, responses that are stimulus-independent and the other generates, with probability ρ, responses that depend on the stimulus. If it is possible to model the stimulus-independent process instead of trying to identify and isolate contaminates, we could make better use of data.

A key idea is that response dependency between trials is assumed to come from the stimulus-independent response. Under this assumption, we developed a model for

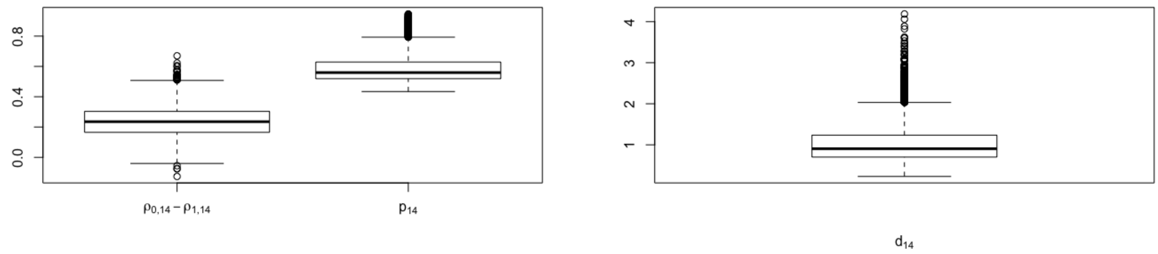

Figure 15. Posterior distributions of model parameters for Participant 14 in Study 2.

the stimulus-independent process using a Markov chain that can insert dependency between trials. The stimulus-based response process is based on signal detection theory (SDT) which is commonly used to model two-choice tasks.

We designed three experiments to test the utility of this modeling strategy. Experiments for Study 1 and Study 3 were two-choice numerosity tasks. One difference experimental design between Study 1 and Study 3 is the task difficulty. Study 1's task had only one difficulty level, but Study 3 had two different task difficulties. Study 2 used a 3-choice task. To make a 3-choice task with equal distance between the means of the stimulus distributions, we developed a location classification task. The task was to classify the origination of stimulus using its presented location.

Study 1 and Study 3 showed that we can discriminate between stimulus-based and stimulus-independent processes as indicated by a higher value of ρ in high motivation conditions. Study 3 with two levels of task difficulty, showed higher values of the discriminability parameter d in easy conditions. The results of Study 2 did not show any difference in ρ for high and low motivation conditions. Even without condition effect, the posterior distribution of d and p spanned a plausible range of values.

Although the three studies had different experimental designs, the main framework is a mixture model of stimulus-dependent and stimulus-independent processes. The probability parameter of the mixture model ρ is interpreted as the probability of using the stimulus-based response over the stimulus-independent response. In these studies, we observed a higher probability ρ of using the stimulus-based response process in high motivation conditions. This result is consistent with the idea of decomposing the

Difficulty	Low	High	High-Low
d'_0	0.502 (S.E.=0.030)	0.561 (S.E.=0.037)	0.056 (t(48)=1.6155)
d'_1	0.213 (S.E.=0.046)	0.268 (S.E.=0.052)	0.06 (t(48)=1.6336)

response process into the stimulus-based and stimulus-independent response processes. The main point is that, by allowing the model to separate the two processes, we can estimate more accurately the parameters d and c of the stimulus-based response process.

This modeling strategy can be fruitful in two ways. First, isolating the effect of the stimulus-independent process can help researchers examine an effect of interest without performing data preprocessing to remove contaminants. Second, the phenomena involved in failing to pay attention during an experiment is worth studying as a topic of research in and of itself.

In Study 2, even though participants tended toward better discriminability in high motivation conditions according to a frequentist analysis of d', the posterior distribution of the parameters of the Bayesian model did not show strong evidence for higher probabilities of using the stimulus-based response process. We assumed there is a single d, distance between the three means, with fixed criteria reflecting no bias. This assumption may not be valid. The estimates of d derived from pairwise comparisons between distributions suggested that the distances between the perceived stimuli may be a lot different from each other. Allowing d to vary might improve the model fit as might allowing variations in criteria.

The biggest issue of the model suggested in the paper might be an identifiability problem. As shown in the model section, it seems that parameter recovery is not accurate for ρ and p parameters. Still discriminability seems to recovered which is mostly the parameter of interest. Also, the pattern of difference between ρ_0 and ρ_1 tends to be consistent. In the purpose of eliminating the contamination of sequential effect, the model is doing some function. However, interpretation of ρ and p is still in question, because there is high possibility that the estimate is biased

In all three studies, there was a strong assumption of the role played by

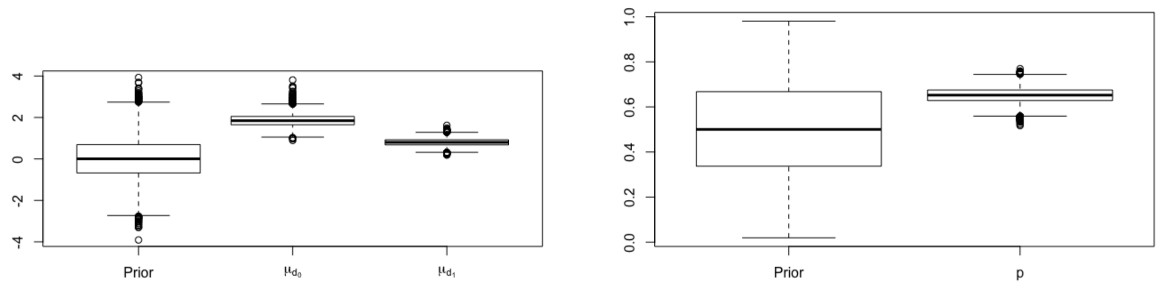

Figure 16. Prior and posterior distributions of model parameters for Study 3. The left panel shows the prior and posterior distributions of the signal detection parameters. The right panel shows the prior and posterior distributions of the probability of repeating the previous response.

dependency between trials. The reason for this assumption comes from the fact that, if the stimulus is generated as an iid sequence then the sequence of responses should be at least approximately iid if the participant responds only on the basis of the stimulus. However, there might be dependencies between trials that are relevant to the task. For example, if a participant was not good at discriminating between stimuli in the early trials, then eventually, from the successes and failures of previous trials, the participant will learn how to do better in the task. Such learning cannot be considered as either stimulus-independent responding or mind wandering. The current model ignores such long-term trends over the sequence of responses.

Future Work

Future research will need to consider the cause of failure of parameter recovery. If we can reproduce the pattern of overestimation of d then, we will be able to determine how to better estimate d. We examined the correlations between parameters with data from Study 1. However, the result does not show very high correlations between the hyperparameters, except for the ρ_0 and ρ_1 which seems okay to have the correlations, it is the same measure but only for different conditions (Figure 5.1).

We also plan to collect neural evidence for the behavioral model in this study.

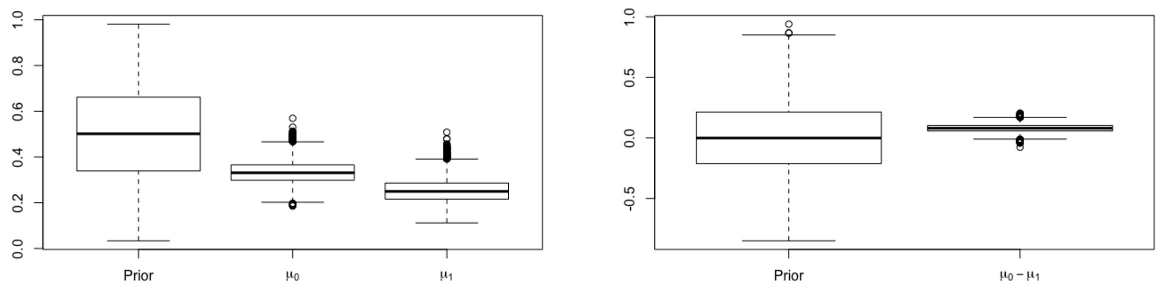

Figure 17. The left panel shows the posterior distributions for the mixture probability μ for high and low motivation (experiment and recalibration) conditions with its prior distribution. The right panel shows the posterior distribution of the μ under the difference in high and low motivation conditions.

From previous studies, we know that mind wandering or task-unrelated thoughts are related to a neural system called the Default Mode Network (DMN; Raichle, 2007). The goal is to relate the model parameters to activation of the DMN, and by so doing, further test the validity of the model. We have already collected pilot data from three participants. In the pilot study, We looked for a relationship between changes in the mixture probability parameter ρ_i between high and low motivation conditions to the brain areas implicated in the DMN (posterior cingulate cortex, medial prefrontal cortex, and angular gyrus). Only one participant showed any difference of ρ_i in high and low motivation conditions, and that participant showed significantly higher activation of those brain areas in the low motivation condition relative to the high motivation condition. This suggests that ρ_i can be related to activation of the DMN. There are three things We need to do to develop this study. First, we need to test more participants. Second, we will need to examine more carefully the connections between the areas which compose the DMN. Third, the current design of the experiment has no lags between the transition of conditions. Because the hemodynamic response is slow compared to the speed of the task, there might be an overlap of the hemodynamic responses between different conditions. We need to develop a new experimental design to avoid this problem.

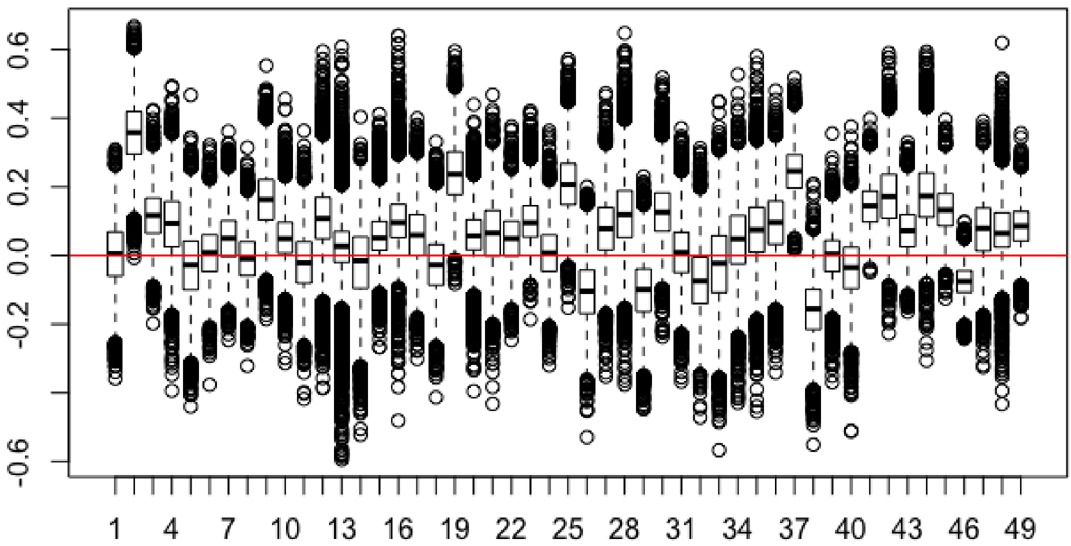

Figure 18. Posterior distributions of $\rho_{0,i} - \rho_{1,i}$ for each of 49 participants.

Conclusion

Modeling task-unrelated processes is a useful way to avoid wasting data by preprocessing. Using a mixture model, we decomposed the response process into stimulus-based and stimulus-independent processes. The results of Study 1 and Study 3 show that it is possible to detect changes in the probability of using the two different processes under different motivation conditions in a two-choice task. However, we still have problem with identifiability problem for some of the parameters in the model. It is important to modify model to have all parameters identifiable.

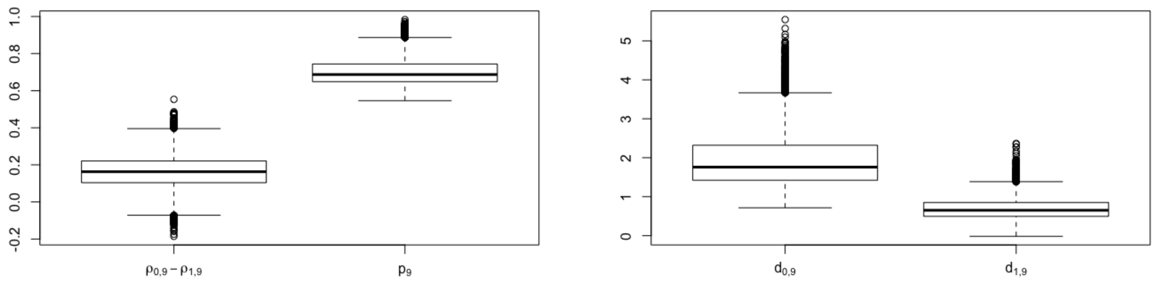

Figure 19. Prior and posterior distributions of model parameters for Participant 9 in Study 3.

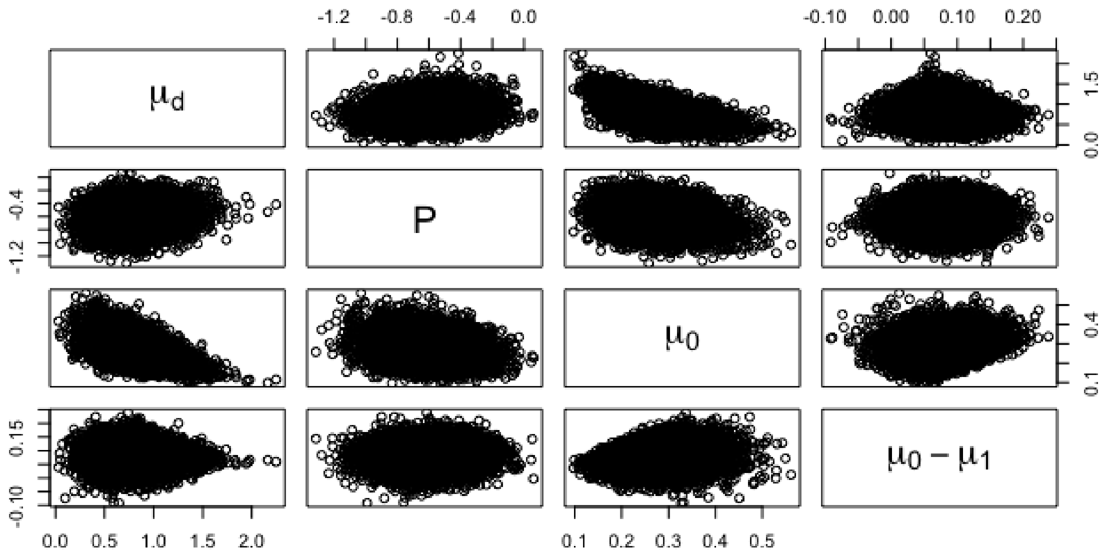

Figure 20. Correlation plot for hyper parameters in Study 1

References

Audley, R., & Pike, A. (1965). “ Some alternative stochastic models of choice.”. *British Journal of Mathematical and Statistical Psychology*(18), 207-225.

Bakan, P. (1960, March). “Response-Tendencies in Attempts to Generate Random Binary Series”. *The American Journal of Psychology*, *73*(1), 127-131.

Barnett, V., & Lewis, T. (1994). *Outliers in Statistical Data.* John Wiley.

Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2010). “Hierarchical Bayes Models for Response Time Data.”. *Psychometrika*, *75*, 613-632.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis 3.* Boca Raton, FL, USA: Chapman Hall/CRC.

Gelman, A., Lee, D., & Guo, J. (2015). Stan: A probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavior Science.*

Gelman, A., & Rubin, D. (1992). “ Inference from Iterative Simulation Using Multiple Sequences.”. *Statistical Science*(7), 457-511.

Green, D., & Swets, S. (1966). *Signal detection theory and psychophysics.* J Wiley New York.

Hastings, C., Mosteller, F., Tukey, J. W., & Winsor, C. P. (1947). “Low moments for small samples: a comparative study of order statistics”. *Annals of Mathematical Statistics*, *18*(3), 413-426.

Hawkins, D. (1980). *Identification of Outliers.* Identification of Outliers.

Hoffman, M., & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, *15*, 1351-1381.

Kim, S., Potter, K., Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2016). “A Bayesian Race Model for Recognition Memory.”. *Journal of the American Statistical Association.*

Logan, G. (1990, November). “Analysis of sequential effects on choice reaction times.”. *Journal of Experimental Psychology,*, *82*(2), 250-257.

Mathot, S., Schreij, D., & Theeuwes, J. (2012). “OpenSesame: An open-source, graphical experiment builder for the social sciences”. *Behavior Research Methods*, *44*(2), 314-324.

Ratcliff, R. (1993). “Methods for Dealing With Reaction Time Outliers”. *Psychological Bulletin*, *114*(3), 510-532.

Remington, R. (1969, January). “Repetition priming and automaticity: Common underlying mechanisms?”. *Cognitive Psychology*, *22*(1), 1-35.

Rivest, L. P. (1994). “Statistical Properties of Winsorized Means for Skewed Distributions”. *Biometrika*, *81*(2), 373-383.

Taylor, H. M., & Karlin, S. (1998). *An Introduction to Stochastic Modeling* (Third ed.). Academic Press.

Townsend, J., & Ashby, G. (1965). “ Some alternative stochastic models of choice.”. *British Journal of Mathematical and Statistical Psychology*(18), 207-225.

Ulrich, R., & Miller, J. (1994). “Effects of truncation on reaction time analysis.”. *Psychological Bulletin*, *123*(1), 34-80.

Van Zandt, T. (2002). Analysis of response time distributions. In *Stevens' handbook of experimental psychology.* John Wiley.

Van Zandt, T., Colonius, H., & Proctor, R. (2000). “ A comparison of two responses models applied to perceptual matching.”. *Psychonomic Bulletin and Review*(27), 208-256.

Wagenaar, W. A. (1972). “Generaton of Random Sequences by Human Subjects: A Critical Survey of Literature”. *Psychological bulletin*, *77*(1), 65-72.

Zimmerman, D. W. (1994). “A note on the influence of outliers on parametric and nonparametric tests.”. *Journal of General Psychology*, *121*(4), 391–401.