Item Response Models
for the Measurement
of Thresholds

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

In Partial Fulfillment
of the Requirements for the Degree
of Doctor of Philosophy

---

by

Richard D. Morey

Dr. Jeffrey N. Rouder, Dissertation Supervisor

August 2008

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

Item Response Models
for the Measurement
of Thresholds

presented by Richard D. Morey

a candidate for the degree of Doctor of Philosophy

and hereby certify that in their opinion it is worthy of acceptance.

Dr. Jeffrey N. Rouder

Dr. Nelson Cowan

Dr. Paul Speckman

Dr. Lori Thombs

Dr. Douglas Steinley

ACKNOWLEDGMENTS

- Dr. Jeff Rouder, for training, help, and support over the past six years
- Mike Pratte, for his helpful discussions regarding the perils of priming paradigms
- My parents, for instilling in me a curiosity about nature and a respect for the search for knowledge

ABSTRACT

At least since Fechner (1860) described examples of human sensory thresholds, the concept of a threshold has been foundational in psychology. Thresholds exist when a sensation can be so weak that it does not lead to detection. Recently, however, thresholds have been abandoned in psychology as a result of the advent of the Theory of Signal Detection (Green & Swets, 1966). I argue that this abandonment was premature and that the concept of a threshold is useful in psychological theory. Thresholds may be defined as the maximum stimulus intensity for which performance is equal to a chance baseline. The measurement of thresholds, however, remains a difficult problem. I present statistical models designed to allow the efficient measurement of thresholds. The models, which have much in common with Item Response Theory models, are hierarchical and are analyzed by Bayesian methods. The models perform well both in simulation and in application to data. Finally, I apply the general model to data from two subliminal priming experiments to test the phenomenon of subliminal priming.

Contents

List of Tables

List of Figures

Chapter 1

Thresholds in Psychology

The concept of a threshold has a long history in Psychology. Although Luce (1963) attributes the earliest mention of thresholds to the ancient Greeks, in modern times, the first comprehensive treatment of the threshold was Fechner's (1860) *Elements of Psychophysics*. In this seminal work, Fechner describes how stimuli that are present, such as a very dim light or a low-intensity sound, can remain unperceived. In Fechner's conception, a sensation may either rise to conscious awareness, or not. A *threshold* determines whether the sensation is intense enough to rise to awareness. In this way, Fecherian sensation is discrete; either a sensation rises to a threshold or does not. One example of a model of sensation which makes use of discrete sensation is the high-threshold model (Blackwell, 1953). In this model, on each presentation of a stimulus there is a probability that it will rise above threshold and the participant will enter a state of detection. If the stimulus is not perceived, participants enter a non-detection state and guess that the stimulus was presented with some probability. If no stimulus is presented, a participant will enter a non-detection state with probability 1.

In the mid-twentieth century, however, discrete models of perception were challenged by continuous models. In continuous models, both stimulation and the absence thereof gives rise to a range of graded perceptions. The most widely-used continuous model of perception is the theory of signal detection (SDT; Green & Swets, 1966). In SDT, there are no discrete detection and non-detection states. When a stimulus is presented, it gives rise to a sensation strength that is a random variable, typically distributed as a normal. If no stimulus is presented, the sensation strength is a also a random variable, but one with a smaller mean than when a stimulus is presented. Observers are assumed to know these distributions well enough to be able to compute the likelihood that a given sensation strength comes from either the no-stimulus distribution or the stimulus distribution. Observers place a criterion on the likelihood ratio, and they use this criterion to select a response.

This dichotomy between discrete and graded perception has characterized the field for over 40 years (see, for example, Luce, 1963a). In perception and memory, researchers have empirically differentiated these contrasting approaches through construction of *receiver operating characteristic* functions. The results seem clear—the data on whole better support continuous-perception models than discrete ones (Green & Swets, 1966; Swets, 1996, cf., Luce, 1963b). Based on this logic, Swets (1961) declared thresholds unnecessary, and the concept of a threshold has been discounted since. Thresholds today do not refer to levels of sensation needed to transition between discrete mental states; instead, they refer to levels of stimulus intensity that result in an select level of performance, say .75 or .707 (Taylor & Creelman, 1967). I use the term intensity throughout to generically to refer to any number of physical stimulus dimensions such as loudness, brightness, duration, etc. Perhaps the most poignant example of the lack of appeal of the concept of thresholds comes from the field of subliminal priming. Subliminal priming refers to the phenomenon in which a stimulus that cannot be detected can, nonetheless, influence subsequent behavior. Subliminal priming seems to necessarily imply a notion of threshold. Even in this domain, however, there is ambivalence. Greenwald, Klinger, and Schuh (1995), for example, write,

> The term *subliminal* implies a theory of perceptual threshold, or limen, that has has been largely abandoned as a consequence of the influence of signal detection theory (Green & Swets, 1966). A more theoretically neutral designation of the class of stimuli which this article is concerned is *marginally perceptible*. This article uses *subliminal* and *marginally perceptible* as interchangeable designations. (p. 22; emphasis in original)

The concept of threshold has been discounted unnecessarily. The discounting arises because thresholds have been associated with discrete processing. This association, however, is not necessary; thresholds may be treated as logically independent from the nature of processing. Figure 1.1A shows an example of a threshold in a continuous processing model. The model is signal detection, and the main graph is a hypothetical relationship between stimulus intensity and sensitivity (the difference between the distribution means, d'). As shown, there is a threshold, denoted with a triangle at an intensity of 30. Stimuli with intensities below this threshold give rise to a range of percepts, but the distribution of these percepts is exactly the same as had there been no stimulation (see left insert). Stimuli more intense than the threshold result in distributions that differ as shown in the right insert. Figure 1.1B shows the opposite case—discrete processing without a threshold. Responses are assumed to be governed by a high-threshold model (Blackwell, 1953) with d denoting the probability of detection. This probability is zero for a lack of stimulation. For any stimulus intensity value above zero, however, the probability of detection is above zero. Even though processing is discrete, there is no positive intensity value for which the distribution of mental states is equivalent to a lack of stimulation.

The demonstration in Figure 1.1 motivates the following definition of threshold. Let *baseline* refer to performance when there is a lack of stimulation. A *threshold* is the maximum intensity for which true performance does not deviate

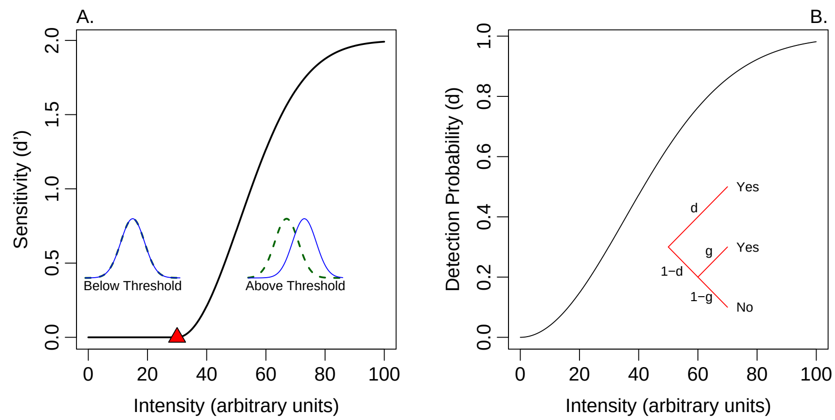

Figure 1.1: A. An example of a threshold in a continuous-processing model. B. An example of a lack of threshold in a discrete-processing model.

from baseline, or chance performance. This definition has the following properties:

1. Thresholds are defined without recourse to underlying mental architecture, and, in particular, are compatible with graded perception.

2. The baseline level of performance associated with below-threshold stimulation is not arbitrary. For example, if participants observe either Stimulus A or Stimulus B on any trial and these two are presented equally often, then the baseline on accuracy is necessarily .5. This property contrasts with the arbitrariness of stipulating thresholds at a conventional level of performance, such as .75 or .707.

3. It is an open question whether thresholds exist. If thresholds do not exist, then the implication is that all stimulation, no matter how small, is registered and affects the response on some nonzero proportion of the trials. If thresholds do exist, then there is loss of information somewhere in processing. The answer to whether thresholds exist probably depends on the domain in question. Perhaps the clearest evidence for a threshold comes

from Loftus, Duncan, & Gehrig (1992) who manipulated the duration of briefly flashed and subsequently masked digits. Figure 1.2 shows their results for two participants. They found that performance fell to baseline when the duration was less than a critical amount. Although this amount varied across participants, it was around 60 ms and was clearly above 0 ms. Busey and Loftus (1994), Shibuya and Bundesen (1988) and Townsend (1981) provide models that explicitly account for a threshold—stimuli that are not sufficiently intense have no effect on processing.

4. The concept of threshold is associated with specific performance measures. Performance to any one stimulus may be assessed with multiple measures, each with its own threshold. For instance, in subliminal priming, performance to the prime is measured directly by accuracy and indirectly by its subsequent priming effect. The threshold on these two measures may differ. The phenomenon of subliminal priming corresponds to a higher threshold for the accuracy measure than for the priming measure.

Although the above definition of threshold is advantageous from a theoretical perspective, it entails methodological complexities. The main problem is one of precision—it is difficult to determine if near baseline-levels of observed performance reflect baseline or slightly-above-baseline true levels. For instance, consider the case in which a participant identifies targets as either Stimulus A or Stimulus B and these two stimuli are presented equally often such that .5 is the appropriate at-chance baseline. Figure 1.3 shows the expected 95% CIs on accuracy as a function of sample size for two levels of true accuracy: .5 and .52. The CIs from one level of accuracy exclude the other level after about 2,400 trials. Restated, the power of excluding chance with a true accuracy of .52 with 2,400 trials is only 1/2. Unfortunately, it is rarely practical to run several thousand trials per participant per condition.

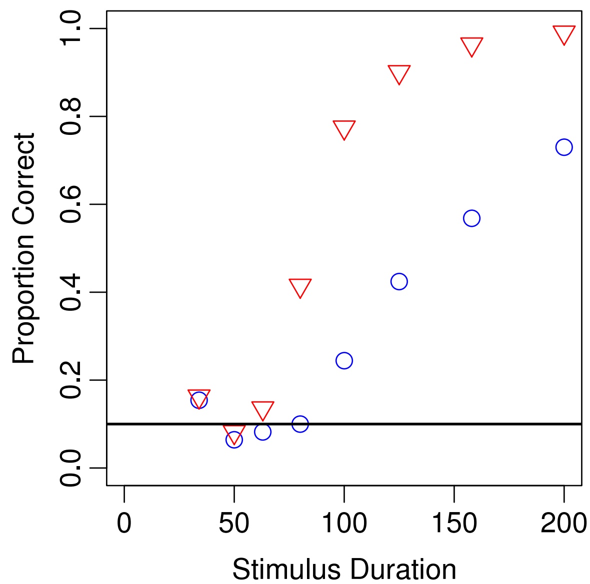

Figure 1.2: Performance of two participants in a digit identification task with a chance baseline of .10. Performance is around chance baseline for very brief durations, then increases as duration is increased. Adapted from Loftus, Duncan, and Gehrig (1992).

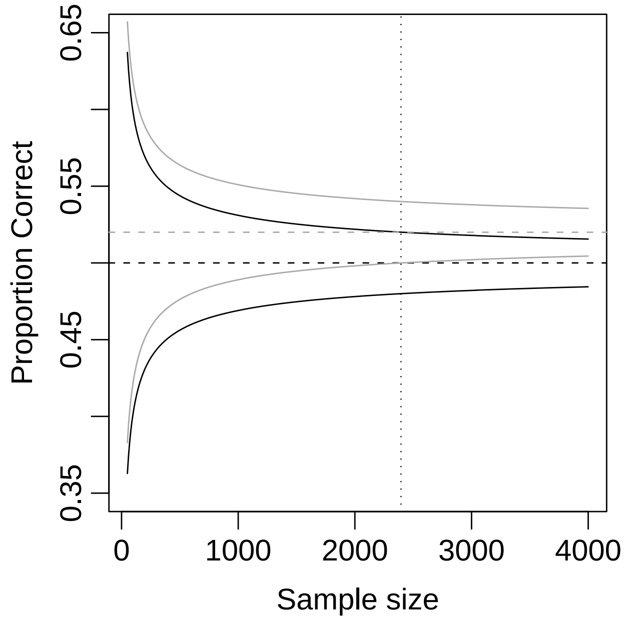

Figure 1.3: Expected 95% CIs as a function of sample size for true accuracies of .5 (dark lines) and .52. (light lines) . The vertical line at a sample size of 2395 shows the point at which the one level is outside the CI of the other.

In the psychological literature, various methods have been used to show that participants, or groups of participants are at baseline, each of which have serious drawbacks. I outline these methods briefly here and discuss the problems with their use.

- **Confidence intervals:** As discussed above, one way of testing that a specific participant is at baseline to see whether the 95% confidence interval around the observed performance score includes chance baseline. Figure 1.4, top-left panel, shows the probability of concluding baseline performance for various above-baseline levels of performance. The results are discouraging. The precision problems shown in Figure 1.3 cause very high error rates for reasonable sample sizes. For example, if the sample size is 50 trials (line A) and the participant has a true accuracy of .6, the probability of concluding the stimulus is below-threshold is .65 (marked with an "+"). Even at larger samples, this error rate remains high. For example, at 400 trials (third line from the left), true accuracy scores of .55 will be misdiagnosed as at baseline with a rate over 0.45 (marked with a "O").

- **Criteria on performance:** Alternatively, researchers have claimed baseline performance by decreasing duration, in fine increments, until a threshold criterion was reached. For example, Dagenbach, Carr, and Wilhelmsen (1989) decreased duration in increments of 5 ms until participants responded correctly on no more than 23 out of 40 trials. Figure 1.4, top-center panel, shows why this approach is problematic. For Dagenbach et al.'s criterion, there is still a high probability of concluding baseline performance when true performance is substantially above chance. Another alternative is provided by Greenwald, Klinger, and Liu (1989), who selected participants whose observed performance was below chance (i.e. accuracy was less than .5). The approach, however, is still susceptible to the same critique. Figure 1.4, top-right panel, shows the probability of accepting baseline per-

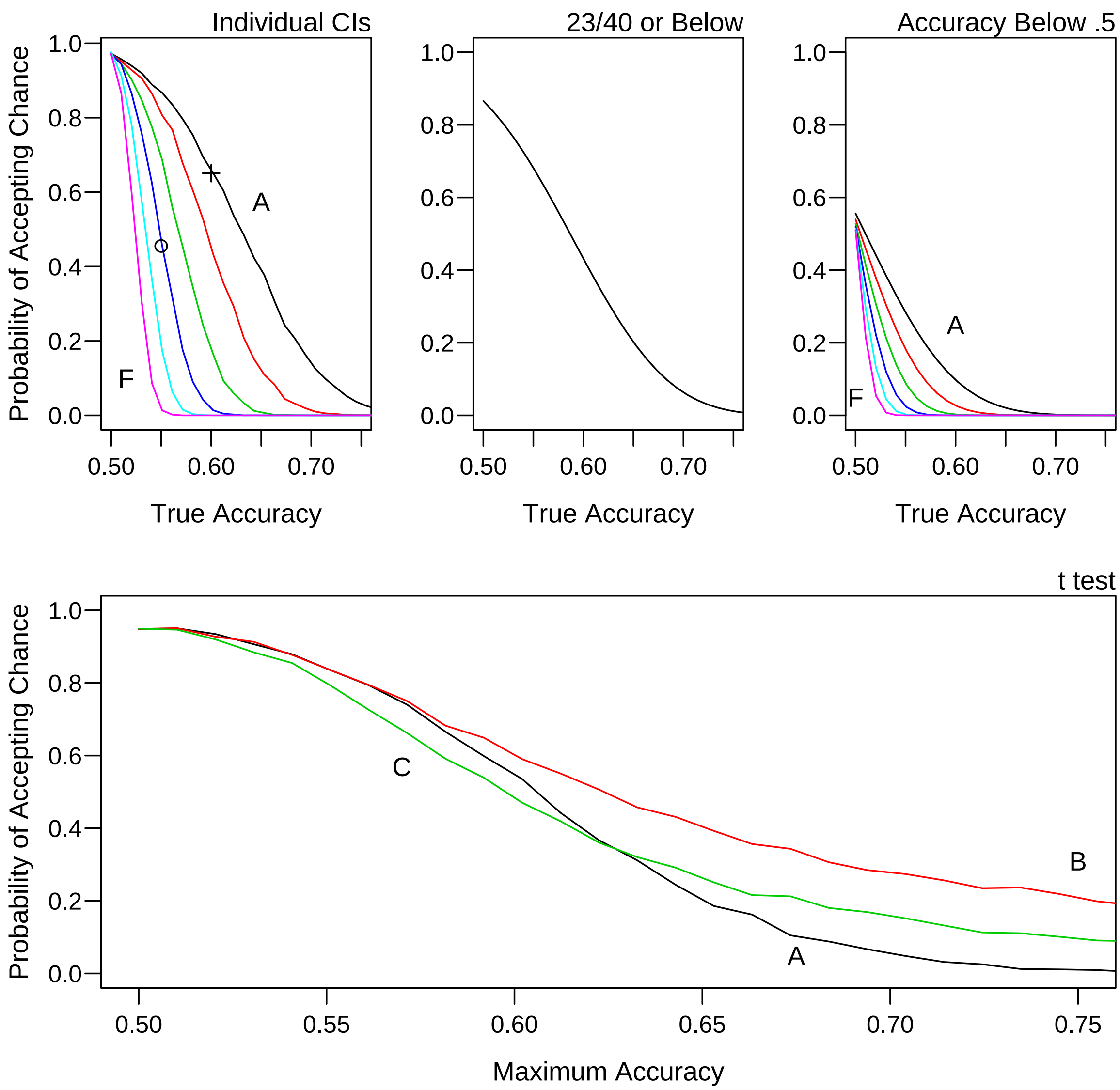

Figure 1.4: Probability of accepting chance performance as a function of true accuracy. Top-left: 95% confidence intervals method. Lines A through F denote sample sizes 50, 100, 200, 400, 800, and 1,600 respectively. Top-middle: Dagenbach et al.'s criterion of 23 of 40 correct responses.. Top-right: Greenwald et al.'s selection of below-chance performance. Lines A through F denote sample sizes 50, 100, 200, 400, 800, and 1,600 respectively. Bottom: t test method. Lines A through C represent the sample sizes from Murphy and Zajonc (1993), Exp. 1, Dehaene et al. (1998), Tasks 1 and 2, respectively.

formance as a function of sample size. Once again, if true values are near .5, the method is susceptible to mistakenly concluding baseline performance.

- t **test:** Researchers have also made claims of baseline performance for groups of participants. Typically, the collection of individuals' accuracy scores are submitted to a t test. The null hypothesis is that the true group average is at baseline and a failure to reject this null is taken as evidence that performance is at baseline for all participants. I highlight two influential papers: Dehaene et al.'s (1998) investigation of the neural correlates of subliminal priming and Murphy and Zajonc's (1993) demonstration of affective subliminal priming from emotive faces. Dehaene et al. asked six participants to discriminate numbers from blank fields (96 trials) and seven participants to discriminate numbers from letter strings (112 trials). For stimulus durations of 33 ms or less, they report that mean accuracies in both tasks were not significantly different from chance. Murphy and Zajonc (1993) asked 32 participants to identify whether a briefly-presented face was happy or scowling on twelve trials. They too found that mean accuracy did not differ significantly from baseline. I assessed the probability of mistakenly accepting baseline performance with a t test through simulations. Each participant was assumed to have some degree of above-baseline performance. True accuracies were uniformly distributed between .5 and an upper bound. Figure 1.4, bottom, shows the probability of accepting the null as a function of the upper bound for the sample sizes in these experiments. Even for high values of this upper bound, there is a substantial probability of mistakenly concluding that performance for all participants was at baseline.

Figure 1.4 presents a troubling picture; it shows the difficulty of determining that stimuli are below threshold. Conventional null hypothesis tests are ill-suited here because they do not offer appropriate safeguards from mistakenly concluding baseline performance.

One way to gain the requisite precision is to pool data across participants. To do so, item response theory (IRT; Lord & Novick, 1968) may be adapted to measure thresholds. Consider a two-choice identification experiment in which a participant decides if a target is Stimulus A or Stimulus B and the two are presented equally often. To find thresholds, the experimenter presents each of these stimuli at a number of intensities. The question is what is the maximum intensity such that performance is at the .5 baseline.

Consider the following one-parameter logistic (1PL) IRT model of performance in a psychophysical task. I refer to each intensity level of the stimulus as an item. Let y_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, denote the number of correct responses in N_{ij} attempts for the ith participant observing the jth item:

$$y_{ij} \overset{\text{indep.}}{\sim} \text{Binomial}(N_{ij}, p_{ij}),$$

where p_{ij} is the true probability of correct response. In 1PL,

$$p_{ij} = \frac{1}{2} + \frac{1}{2(1 + e^{-(\alpha_i - \beta_j)})}, \tag{1.1}$$

where α_i and β_j denote the ability and difficulty of the ith participant and jth item, respectively. Participant abilities are zero-centered normal random variables with common variance. This model may be analyzed with a variety of methods (i.e., Andersen, 1973; Rasch, 1960; Swaminathan & Gifford, 1982).

One of the problems with 1PL is that it cannot measure a threshold. An individual's threshold in this case is the minimum value of β_j necessary for $p_{ij} =$ .5. However, for all finite values of α_i and β_j, $p_{ij} > .5$. Figure 1.5A shows this fact. The line is a person-response curve for the average person ($\alpha_i = 0$). As difficulty increases, accuracy approaches the baseline but never attains it. This incompatibility with thresholds holds for all links with support on $(-\infty, \infty)$ such as the logit, probit, and t links.

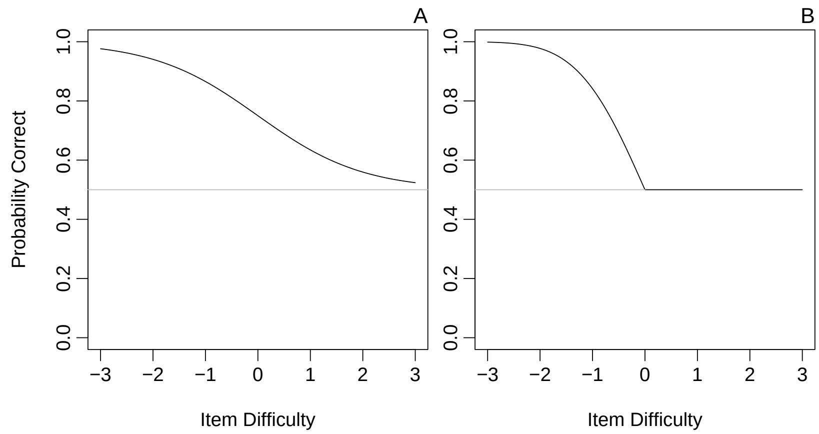

Figure 1.5: A. Person-response function for the average person in 1PL. B. The same for the MAC model.

Rouder, Morey, Speckman and Pratte (2007) proposed a truncated-probit link to model the transition in performance from baseline to above-baseline levels:

$$p(\alpha, \beta) = \begin{cases} \Phi(\alpha_i - \beta_j), & \alpha_i > \beta_j, \\ .5 & \beta_j \geq \alpha_i, \end{cases} \tag{1.2}$$

where Φ is the CDF of the standard normal distribution. If the person's ability is greater than the item's difficulty, then performance is above baseline; conversely, if the item's difficulty is greater than the person's ability, then performance is at baseline. A person-response curve for an average person ($\alpha_i = 0$) is shown in Figure 1.5B. For all positive difficulties, performance is at baseline. For all negative difficulties, performance follows a probit. Because there is nonzero probability of baseline performance, Rouder et al. called this model *Mass at Chance* (MAC).

In the following chapters, I develop MAC models for determining whether participants are at baseline. In Chapter 2, I develop a MAC model that allows for multiple items, similar to a one-parameter logistic model. In Chapter 3, I generalize the model to the two-parameter case. The final chapter illustrates the use of the two-parameter model in two subliminal priming experiments.

Chapter 2

The One Parameter Mass at Chance Model

In previous work (Rouder et al., 2007), my colleagues and I proposed a Bayesian hierarchical model to select participants whose true prime identification performance is at chance. We termed this model the *mass at chance* model (MAC) to emphasize that a threshold is explicitly assumed and modeled.

Unfortunately, Rouder et al.'s MAC model has a pronounced limitation. Although it controls the error rate in which above-chance true performance is considered subliminal, it has a tendency to mistakenly classify true at-chance performance as above chance. The reason for this is that paradigms with single conditions provide very little information for the MAC model to use in classification. To provide for more efficient estimation of thresholds as well as more efficient discrimination of at-baseline and above-baseline performance, in this chapter I develop a one parameter MAC model. In the first section, I briefly present the Rouder et al. MAC model and highlight its limitations. Following that, I present the expanded paradigm and an expanded MAC model. I show how this expanded model provides for highly accurate measurement of true performance and, consequently, provides a methods for assessing priming theories and estimating at-chance thresholds. Finally, I show how this model is similar to the Rasch model (Lord & Novick, 1968) in Item Response Theory.

2.1 Rouder et al.'s MAC Model

The Rouder et al. MAC model was designed to account for performance in two-alternative forced-choice tasks including a threshold. At the first level of Rouder et al.'s MAC model, the number of correct prime identifications for the ith participant, $i = 1, \ldots, I$, denoted y_i, is

$$y_i \overset{\text{indep.}}{\sim} \text{Binomial}(p_i, N_i). \tag{2.1}$$

Parameter p_i is the probability of a correct response and N_i is the number of trials observed by the ith participant. For two-choice paradigms, primes are below threshold if $p_i = .5$.

Each participant is assumed to have a true score, denoted x_i. True probabilities (p_i) are related to latent abilities:

$$p_i = \begin{cases} .5, & x_i \leq 0, \\ \Phi(x_i), & x_i > 0, \end{cases} \tag{2.2}$$

where Φ is the standard normal cumulative distribution function. The point $x_i = 0$ serves as the threshold. If a participant has positive true score, then performance is above chance; conversely, if true score is negative, then performance is at chance. Negative true scores are interpretable: more stimulus energy is needed to bring the participant to threshold.

True scores are assumed to be random effects and distributed as normal in the population:

$$x_i \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2).$$

True scores depend on item factors. For example, the same participant will have higher true score to primes presented for 33 ms than to those presented for 22 ms. This dependence is implemented by making μ dependent on stimulus intensity.

Rouder et al. (2007) developed MAC for a single parameter item parameter μ. Priors are needed for μ and σ^2. Based on extensive simulation, Rouder et al. chose somewhat informative priors $\mu \sim$ Normal(0,1) and $\sigma \sim \text{Uniform}(0, 1)$.

Participants are considered at-chance if much of the posterior distribution of x_i is below zero. Therefore, the following posterior probability serves as a decision statistic:

$$\omega_i = \Pr(x_i < 0 \mid \text{data}). \tag{2.3}$$

The decision rule is

$$\begin{array}{lcccl} \omega_i & \geq & .95 & \rightarrow & \text{Conclude the participant is at chance} \\ \omega_i & < & .95 & \rightarrow & \text{Conclude the participant is above chance} \end{array} \tag{2.4}$$

The main benefit of this approach is that it does not suffer from high rates of mistakenly claiming baseline performance, whereas the methods discussed in Chapter 1 do. In the MAC model, small sample sizes result in highly variable posterior estimates of x_i. Consequently, there tends to be more mass above $x_i > 0$. Lowering sample sizes makes it more difficult to claim that items are below-threshold rather than less difficult. Therefore, the MAC model provides an appropriate safeguard against high error rates in claiming baseline performance.

Although the MAC model provides a principled means for determining whether performance is at baseline, it has two major limitations. First, as mentioned previously, the model does not select people as at chance often. Figure 2.1 shows ω_i, the posterior probability that each participant is at chance, as a function of observed accuracy. The three closed circles represent participants selected as at chance. Although most of the 27 participants perform near chance, the model selects only three. The paradigm does not offer the model enough information to select more than three participants as at-chance, making it very difficult to use for researchers interested in below-threshold phenomena. Second, the MAC model is quite dependent to specification of the prior. In particular, the posterior of σ^2 was heavily influenced by the choice of prior[1]. This dependence is

[1] In typical paradigms, primes are designed to be below threshold. Consequently, much of the

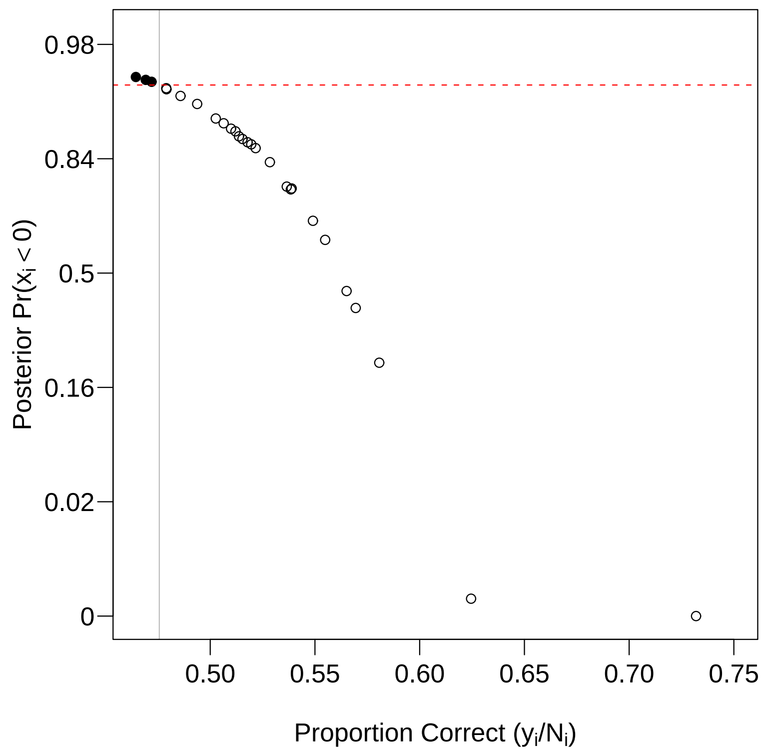

Figure 2.1: Posterior probability that an individual participant is at chance as a function of observed accuracy. Participants whose posterior probability is above .95 (closed circles) are selected as at chance. (Figure adapted from Rouder et al., 2007)

undesirable.

These limitations may be mitigated by generalizing the model to multiple items. I extend the model for this case and find that: (1) the model allows for exceedingly accurate classification of performance as at-chance or above chance; (2) the model is less dependent on the prior for variance.

mass of the distribution of x_i is below 0. Because participants at chance are indistinguishable from one another, estimation of the variance of the distribution of latent abilities can only be based on the prior and the performance of those participants above chance. When only a few participants are above chance, the prior has a large degree of influence.

2.2 The One Parameter MAC Model

It is straightforward to specify the extended MAC model for multiple items. We call the extended model 1P-MAC, according to Item Response Theory naming conventions, because of its similarity to the 1PL model described in Chapter 1. Let y_{ij}, p_{ij}, and N_{ij} denote the number of correct responses, true probability of correct response, and number of trials, respectively, for the ith participant to the jth item:

$$y_{ij} \sim \text{Binomial}(p_{ij}, N_{ij}). \tag{2.5}$$

At the next level, it is assumed that for each participant-by-item combination, there is a true score, denoted x_{ij}. True scores determine probabilities on performance:

$$p_{ij} = \begin{cases} .5, & x_{ij} \leq 0, \\ \Phi(x_{ij}), & x_{ij} > 0. \end{cases} \tag{2.6}$$

True scores reflects both the participant's latent ability, denoted α_i, and the ease of the item, denoted μ_j in an additive manner:

$$x_{ij} = \alpha_i + \mu_j. \tag{2.7}$$

This additive model is borrowed from item response theory. Latent abilities are modeled as random effects:

$$\alpha_i \overset{iid}{\sim} \text{Normal}(0, \sigma^2).$$

The additivity assumption provides a convenient means of pooling information about latent ability (α_i) across items and about item ease (μ_j) across participants. This pooling allows for the possibility of estimating true score x_{ij}, even when this true score is below threshold. The threshold for participant i is the intensity for which the true score is $x_{ij} = 0$. The additivity assumption is made with reference to the MAC link. It is impossible to assess whether additivity holds without referencing a link. For example, if additivity holds for the MAC link, it certainly will not for a logistic one, and vice-versa.

2.2.1 The MAC link function

In both Rouder et al.'s MAC model and the 1P-MAC model, true scores x_{ij} are mapped to probabilities using the half-probit link in Eqs. (2.2) and (2.6). This link function is shown as the solid line in Figure 2.2A. The MAC link differs from typical psychophysical functions. The former starts from a point and rises quickly without any inflection point. Psychophysical functions, however, tend to be characterized by a sigmoid-shaped function that has an inflection point. Examples of the latter include the logit, probit, and cumulative distribution function of a Weibull with shapes around 2. The dotted line shows a logit link and the inflection point is clearly visible.

Given the success of previous inflected psychophysical functions in fitting extant data, it would appear that the MAC link is unwarranted. This appearance, however, is deceiving. The MAC link is not a psychophysical function. It is a mapping from true scores into performance. True scores are latent rather than observed. Psychophysical functions, on the other hand, map physical characteristics of stimuli (e.g., duration) into performance. These characteristics are observed and are not latent. In the MAC model, the researcher estimates the relationship between stimulus characteristics and μ_j, the latent ease of the stimulus. Because this relationship is estimated rather than assumed, the MAC model makes no commitments to the form of the psychophysical function.

Figure 2.2B-C provides a demonstration of how the MAC accounts for psychophysical functions with inflection points. Panel B shows hypothetical mappings from duration to true scores for three participants. The additivity assumption in Eq. (2.7) implies that these mappings are parallel. Panel C shows the resulting mappings from duration to probability correct for the same three participants. The psychophysical functions in Panel C have inflection points.

If the MAC model is flexible enough to account for any psychophysical function, it is reasonable to wonder if the model is concordant with all data and,

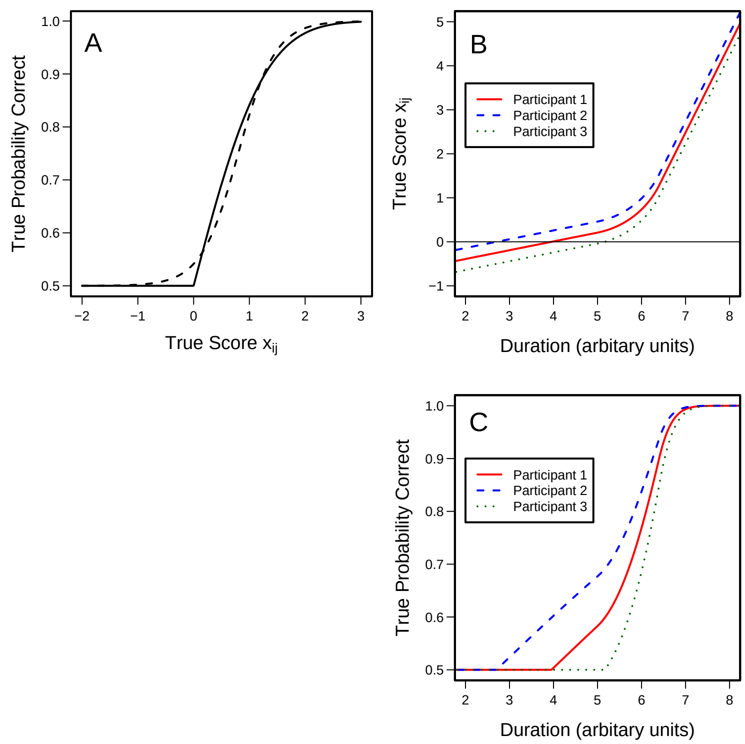

Figure 2.2: A: The MAC half-probit link function (solid line) and a logistic link function (dashed line). B: A hypothetical mapping from duration to true score for three participants. C: A hypothetical mapping from duration to probability correct for the same three participants.

thus, is vacuous. It is not. The model places constraints in two important ways. First, the model describes how psychophysical links are distributed across the population. As seen in Figure 2.2C, the additivity of item ease and latent ability results in specific relations between individual-level psychophysical functions. For smooth convex mappings, such as those in Figure 2.2B, participants with smaller thresholds have smaller rates of gain in their predicted psychometric functions, as seen in Figure 2.2C. Second, the discontinuity in the first derivative of the MAC link ensures that there is an at-chance threshold in the psychophysical function. I discuss methods of assessing whether these constraints hold in data when analyzing the experiment.

Figure 2.2C shows that the MAC model, in general, does not predict parallel psychometric functions across participants. This fact seems incongruous with themes in the field. Watson and Pelli (1983), for example, assume that psychometric functions across different conditions are parallel, that is, they are translation invariant. This assumption is common in many theoretical models including those of Green and Luce (1975), Roufs (1974), and Watson (1979); the empirical evidence for the invariance, however, is controversial (see Nachmias, 1981). The translation invariance considered is across conditions of similar stimulation; for example, across pulse trains of light flashes with varying numbers of pulses. Translation invariance across conditions, however, is not the same as translation invariance across participants. I do not know of any research focused on the latter. Nachmias' (1981) results, however, may be used to empirically assess translation invariance across participants in a brightness detection task. The estimates of the shapes of psychometric functions for some of conditions varied widely across participants (Table 1, p. 220, Bipartite Field).

2.3 Prior Distributions

Analysis of the model is performed by Bayesian methods. Consequently, priors are needed for each μ_j and σ^2:

$$\begin{aligned}\mu_j &\stackrel{iid}{\sim} \text{Normal}(\mu_0, \sigma_0^2), &(2.8)\\ \sigma^2 &\sim \text{Inverse Gamma}(a_0, b_0). &(2.9)\end{aligned}$$

The inverse gamma distribution is commonly used as priors for variance in Bayesian analysis. The density function is

$$f(\sigma^2 \mid a_0, b_0) = \frac{(b_0)^{a_0}}{\Gamma(a_0)}(\sigma^2)^{-a_0-1} \exp\left\{-\frac{b_0}{\sigma^2}\right\} \quad \sigma^2, a_0, b_0 > 0 \qquad (2.10)$$

The parameters in the priors (μ_0, σ_0^2, a_0, b_0) are chosen before analysis. A non-informative Jeffreys prior may be placed on σ^2 by setting $a_0 = b_0 = 0$ (Jeffreys, 1961)[2]. I place a reasonable but informative prior on μ_j by setting $\mu_0 = 0$ and $\sigma_0^2 = 1$. To see why this prior is reasonable, consider an average participant with $\alpha_i = 0$. The Normal$(0, 1)$ prior on μ_j implies that the marginal prior on p_{ij} has half its mass at chance and is flat over (.5,1). If σ_0^2 is increased beyond 1.0, the marginal prior distribution on p_{ij} becomes increasingly bimodal with modes at .5 and 1. The prior parameters $\mu_0 = 0$ and $\sigma_0^2 = 1$ were chosen as the most diffuse prior without bimodalities in the marginal prior on p for a participant with $\alpha_i = 0$. These choices convey information. To show that this information is not critical in analysis, I tried other values as high as $\sigma_0^2 = 16$. The effect is inconsequential.

2.4 Model Analysis

Although the model is straightforward to specify, analysis is complicated by the non-linearity in the MAC link. Closed-form expressions for the marginal

[2]With parameters $a_0 = b_0 = 0$, the prior on σ^2 becomes $\sigma^2 \propto (\sigma^2)^{-1}$. Although this prior is improper, the posterior will be proper if $I > 2$.

posterior distributions are not available. Instead, posterior quantities are estimated by Markov chain Monte Carlo techniques, specifically via Gibbs sampling (Gelfand & Smith, 1990). In this section, I provide the full conditional posterior distributions for parameters and provide efficient algorithms for sampling these distributions.

MCMC estimation of posterior distributions is greatly simplified by the data-augmentation method of Albert and Chib (1993; see Rouder & Lu, 2005, for a tutorial review). In the data-augmentation method, it is convenient to model the data as a collection of Bernoulli outcomes. Let y_{ijk} be the outcome of the kth trial, $k = 1, \ldots, N_{ij}$, for the ith participant to the jth item, where $y_{ijk} = 1$ indicates a correct response and $y_{ijk} = 0$ indicates an error. Accordingly,

$$Pr(y_{ijk} = 1) = \Phi(x_{ij} \vee 0),$$

where $a \vee b$ is $\max(a, b)$.

Let w_{ijk} denote a normally-distributed latent random variable whose sign determines the correctness of the response:

$$y_{ijk} = \begin{cases} 0 & w_{ijk} \leq 0 \\ 1 & w_{ijk} > 0 \end{cases}$$

The w_{ijk} are distributed as

$$w_{ijk} \overset{\text{indep}}{\sim} \text{Normal}(x_{ij} \vee 0, 1).$$

Then, $Pr(y_{ijk} = 1) = Pr(w_{ijk} > 0)$. It is simpler to sample from the full conditional distributions of the parameters conditioned on the latent parameters w_{ijk} than on data y_{ijk}.

With the model specified, it is straightforward to define the joint posterior distribution. Let $\mathbf{y}$ be the collection of data, $\mathbf{y} = \left\langle \left\langle \langle y_{ijk} \rangle_{k=1}^{k=N_{ij}} \right\rangle_{j=1}^{j=J} \right\rangle_{i=1}^{i=I}$, and $\mathbf{w}$ be the collection of latent parameters, $\mathbf{w} = \left\langle \left\langle \langle w_{ij} \rangle_{k=1}^{k=N_{ij}} \right\rangle_{j=1}^{j=J} \right\rangle_{i=1}^{i=I}$. The joint posterior of all parameters is

$$
\begin{aligned}
[\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\mu}, \sigma^2 \mid \mathbf{y}] \propto & \prod_i \prod_j \left\{ \prod_{\substack{k=1 \\ (y_{ijk}=0)}}^{N_{ij}} \exp\left\{ -\frac{(w_{ijk} - (\alpha_i + \mu_j) \vee 0)^2}{2} \right\} I_{(w_{ijk}<0)} \right. \\
& \left. \times \prod_{\substack{k=1 \\ (y_{ijk}=1)}}^{N_{ij}} \exp\left\{ -\frac{(w_{ijk} - (\alpha_i + \mu_j) \vee 0)^2}{2} \right\} I_{(w_{ijk}>0)} \right\} \\
& \times \prod_i (\sigma^2)^{-\frac{1}{2}} \exp\left\{ -\frac{\alpha_i^2}{2\sigma^2} \right\} \\
& \times \prod_j \exp\left\{ -\frac{(\mu_j - \mu_0)^2}{2\sigma_0^2} \right\} \\
& \times (\sigma^2)^{-(a_0+1)} \exp\left\{ -\frac{b_0}{\sigma^2} \right\}. \qquad (2.11)
\end{aligned}
$$

2.4.1 Full conditional distributions

I present the full conditional distributions of each of the parameters. Proofs are provided after each of the full conditional distributions. In this document, bold-face typeset is reserved for vectors, matrices, and multidimensional arrays. Using this notation, $\mathbf{y}$ and $\mathbf{w}$ are the collection of all data y_{ijk} and all latent variables w_{ijk}; $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ are the vectors of all α_i and all μ_j.

Fact 1. The full conditional posterior for $\sigma^2 \mid \boldsymbol{\alpha}$ is

$$
\sigma^2 \mid \boldsymbol{\alpha} \sim \text{InverseGamma}(a_0 + \frac{I}{2}, \sum_i \alpha_i^2/2 + b_0) \qquad (2.12)
$$

Proof of Fact 1: By inspection of (2.11), the full conditional posterior $[\sigma^2 \mid \boldsymbol{\alpha}]$ is

$$
\begin{aligned}
[\sigma^2 \mid \boldsymbol{\alpha}] \propto & \prod_i (\sigma^2)^{-\frac{1}{2}} \exp\left\{ -\frac{\alpha_i^2}{2\sigma^2} \right\} \\
& \times (\sigma^2)^{-(a_0+1)} \exp\left\{ -\frac{b_0}{\sigma^2} \right\}.
\end{aligned}
$$

Collecting like terms yields

$$[\sigma^2 \mid \boldsymbol{\alpha}] \;\; \propto \;\; (\sigma^2)^{-(a_0+I/2+1)} \exp\left\{-\frac{\frac{1}{2}\sum_i \alpha_i^2 + b_0}{\sigma^2}\right\}. \tag{2.13}$$

The right hand side of (2.13) is proportional to the density of an inverse gamma distribution with parameters $a = a_0 + I/2$ and $b = \sum_i \alpha_i^2/2 + b_0$. $\square$

Fact 2. The full conditional posterior for $w_{ijk} \mid x_{ij}, y_{ijk}$

$$\begin{aligned} w_{ijk} \mid x_{ij}; (y_{ijk} = 1) &\sim \mathrm{TN}_{(0,+\infty)}(x_{ij} \vee 0, 1) \\ w_{ijk} \mid x_{ij}; (y_{ijk} = 0) &\sim \mathrm{TN}_{(-\infty,0)}(x_{ij} \vee 0, 1). \end{aligned} \tag{2.14}$$

where $\mathrm{TN}_{(l,u)}(\mu, \sigma^2)$ denotes a normal distribution with parameters μ and σ^2 truncated below at l and above at u.

Proof of Fact 2: The full conditional distribution $w_{ijk} \mid y_{ijk}, \boldsymbol{\alpha}, \boldsymbol{\mu}$ follows directly from the fact that the truncated normal is the distribution of a normal conditioned on a positive (or, alternatively, negative) outcome.

Fact 3. The full conditional posteriors of α_i are independent for given $\boldsymbol{\mu}$, σ^2, and $\mathbf{w}$. It is convenient to define the following sets. Let $A_0 = (-\mu_{(1)}, \infty)$, $A_1 = (-\mu_{(2)}, -\mu_{(1)})$, ..., $A_J = (-\infty, -\mu_{(J)})$, where $\mu_{(1)} < \ldots < \mu_{(J)}$ are the order statistics of the μ_j. Let

$$\begin{aligned} J_0 &= \{1, \ldots, J\} \\ J_\ell &= \left\{j : \mu_j > \mu_{(\ell)}\right\} \end{aligned}$$

and define

$$s_{\alpha_i \ell} = \begin{cases} \left(\frac{1}{\sigma^2} + \sum_{j \in J_\ell} N_{ij}\right)^{-1} & 0 \le \ell < J \\ \sigma^2 & \ell = J \end{cases} \tag{2.15}$$

$$m_{\alpha_i \ell} = \begin{cases} s_{\alpha_i \ell} \sum_{j \in J_\ell} \left(\sum_{k=1}^{N_{ij}} w_{ijk} - N_{ij}\mu_j\right) & 0 \le \ell < J \\ 0 & \ell = J \end{cases} \tag{2.16}$$

$$h_{\alpha_i \ell} \;=\; \begin{cases} \sum_{j \in J_\ell} \left(N_{ij} \mu_j^2 - 2\mu_j \sum_{k=1}^{N_{ij}} w_{ijk} \right) & 0 \le \ell < J \\ 0 & \ell = J \end{cases} \qquad (2.17)$$

Then

$$[\alpha_i \mid \boldsymbol{\mu}, \sigma^2, \mathbf{w}] \;\propto\; \sum_{\ell=0}^{J} \exp\left\{ -\frac{1}{2} \left(h_{\alpha_i \ell} - \frac{(m_{\alpha_i \ell})^2}{s_{\alpha_i \ell}} \right) \right\} \times \exp\left\{ -\frac{(\alpha_i - m_{\alpha_i \ell})^2}{2 s_{\alpha_i \ell}} \right\} I_{(\alpha_i \in A_\ell)} \qquad (2.18)$$

Following Bayesian convention, $[\alpha_i \mid \boldsymbol{\mu}, \sigma^2, \mathbf{w}]$ denotes the conditional density of α_i given $\boldsymbol{\mu}$, σ^2, and $\mathbf{w}$. The right-hand side of (2.18) is proportional to the density of a mixture of truncated normals. Set $\mu_{J+1} = \infty$ and $\mu_0 = -\infty$. Let

$$q_{\alpha_i \ell} \;=\; \exp\left\{ -\frac{1}{2} \left(h_{\alpha_i \ell} - \frac{(m_{\alpha_i \ell})^2}{s_{\alpha_i \ell}} \right) \right\} \sqrt{2\pi s_{\alpha_i \ell}} \times \left[\Phi\left(\frac{-\mu_{(\ell)} - m_{\alpha_i \ell}}{\sqrt{s_{\alpha_i \ell}}} \right) - \Phi\left(\frac{-\mu_{(\ell+1)} - m_{\alpha_i \ell}}{\sqrt{s_{\alpha_i \ell}}} \right) \right], \qquad (2.19)$$

for $\ell = 0, \ldots, J$ and let $p_{\alpha_i \ell} = q_{\alpha_i \ell} / \sum_{\ell=0}^{J} q_{\alpha_i \ell}$, $\ell = 0, \ldots, J$. Then the full conditional distribution $[\alpha_i \mid \boldsymbol{\mu}, \sigma^2, \mathbf{w}]$ is the mixture of truncated normal distributions

$$\alpha_i \mid \boldsymbol{\mu}, \sigma^2, \mathbf{w} \;\sim\; \sum_{\ell=0}^{J} p_{\alpha_i \ell} \mathrm{TN}_{A_\ell}(m_{\alpha_i \ell}, s_{\alpha_i \ell}) \qquad (2.20)$$

Proof of Fact 3: Inspection of (2.11) reveals that the full conditional posterior of α_i satisfies

$$[\alpha_i \mid \boldsymbol{\mu}, \mathbf{w}, \sigma^2] \;\propto\; \exp\left\{ -\frac{\alpha_i^2}{2\sigma^2} \right\} \prod_{j=1}^{J} \prod_{k=1}^{N_{ij}} \exp\left\{ -\frac{1}{2} \left[(w_{ijk} - \alpha_i - \mu_j)^2 I_{(\alpha_i > -\mu_j)} + w_{ijk}^2 I_{(\alpha_i \le -\mu_j)} \right] \right\}.$$

Noting that $I_{(\alpha_i \leq \mu_j)} = 1 - I_{(\alpha_i > \mu_j)}$, the right hand side may be rewritten as

$$\begin{aligned}
[\alpha_i \mid \boldsymbol{\mu}, \mathbf{w}, \sigma^2] \quad \propto \quad & \exp\left\{-\frac{\alpha_i^2}{2\sigma^2}\right\} \\
& \times \exp\left\{-\frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{N_{ij}}\left[(w_{ijk}-\alpha_i-\mu_j)^2 I_{(\alpha_i>-\mu_j)}\right.\right. \\
& \left.\left. +\ w_{ijk}^2\left(1-I_{(\alpha_i>-\mu_j)}\right)\right] \right\}.
\end{aligned}$$

Expanding the square and simplifying yields

$$\begin{aligned}
[\alpha_i \mid \boldsymbol{\mu}, \mathbf{w}, \sigma^2] \quad \propto \quad & \exp\left\{-\frac{1}{2}\left(\frac{\alpha_i^2}{\sigma^2} + \sum_{j=1}^{J}\sum_{k=1}^{N_{ij}}\left(w_{ijk}^2 + \alpha_i^2 I_{(\alpha_i>-\mu_j)} + \mu_j^2 I_{(\alpha_i>-\mu_j)}\right.\right.\right. \\
& \left.\left.\left. +\ 2\alpha_i\mu_j I_{(\alpha_i>-\mu_j)} - 2\alpha_i w_{ijk} I_{(\alpha_i>-\mu_j)} - 2\mu_j w_{ijk} I_{(\alpha_i>-\mu_j)}\right)\right)\right\} \\
\propto \quad & \exp\left\{-\frac{1}{2}\left(\frac{\alpha_i^2}{\sigma^2} + \alpha_i^2\sum_{j=1}^{J} N_{ij} I_{(\alpha_i>-\mu_j)} + \sum_{j=1}^{J} N_{ij}\mu_j^2 I_{(\alpha_i>-\mu_j)}\right.\right. \\
& +2\alpha_i \sum_{j=1}^{J} N_{ij}\mu_j I_{(\alpha_i>-\mu_j)} - 2\alpha_i \sum_{j=1}^{J} I_{(\alpha_i>-\mu_j)} \sum_{k=1}^{N_{ij}} w_{ijk} \\
& \left.\left. -\ 2\sum_{j} \mu_j I_{(\alpha_i>-\mu_j)} \sum_{k=1}^{N_{ij}} w_{ijk}\right)\right\}.
\end{aligned}$$

Collecting terms in α_i^2 and α_i yields

$$\begin{aligned}
[\alpha_i \mid \boldsymbol{\mu}, \mathbf{w}, \sigma^2] \quad \propto \quad & \exp\left\{-\frac{1}{2}\left[\sum_{j=1}^{J} N_{ij}\mu_j^2 I_{(\alpha_i>-\mu_j)}\right.\right. \\
& \left.\left. -\ 2\sum_{j=1}^{J} \mu_j I_{(\alpha_i>-\mu_j)} \sum_{k=1}^{N_{ij}} w_{ijk}\right]\right\} \\
& \times \exp\left\{-\frac{1}{2}\left[\alpha_i^2\left(\frac{1}{\sigma^2} + \sum_{j=1}^{J} N_{ij} I_{(\alpha_i>-\mu_j)}\right)\right.\right. \\
& \left.\left. -\ 2\alpha_i\left(\sum_{j=1}^{J} I_{(\alpha_i>-\mu_j)} \sum_{k=1}^{N_{ij}} w_{ijk} - \sum_{j=1}^{J} N_{ij}\mu_j I_{(\alpha_i>-\mu_j)}\right)\right]\right\}.
\end{aligned} \tag{2.21}$$

Let

$$
\begin{aligned}
s_i(\alpha) &= \left(\frac{1}{\sigma^2} + \sum_{j=1}^{J} N_{ij} I_{(\alpha > -\mu_j)}\right)^{-1}, \\
m_i(\alpha) &= s_i(\alpha)\left(\sum_{j=1}^{J} I_{(\alpha > -\mu_j)} \sum_{k=1}^{N_{ij}} w_{ijk} - \sum_{j=1}^{J} N_{ij}\mu_j I_{(\alpha > -\mu_j)}\right), \text{ and} \\
h_i(\alpha) &= \sum_{j=1}^{J} N_{ij}\mu_j^2 I_{(\alpha > -\mu_j)} - 2\sum_{j=1}^{J} \mu_j I_{(\alpha > -\mu_j)} \sum_{k=1}^{N_{ij}} w_{ijk}.
\end{aligned}
$$

Substitution into (2.21) yields

$$
\begin{aligned}
[\alpha_i \mid \boldsymbol{\mu}, \mathbf{w}, \sigma^2] \;\propto\; & \exp\left\{-\frac{h_i(\alpha_i)}{2}\right\} \exp\left\{-\frac{1}{2s_i(\alpha_i)}\left(\alpha_i^2\right.\right. \\
& \left.\left. -\ 2\alpha_i m_{\alpha_i}(\alpha_i) + m_i(\alpha_i)^2 - m_i(\alpha_i)^2\right)\right\}.
\end{aligned}
$$

Note that each of these functions s_i, m_i, and h_i are constant within regions A_ℓ defined in Fact 3. Using the constants defined in (2.15-2.17), completing the square and simplifying yields (2.20). $\square$

Fact 4. The full conditional posterior distributions of μ_j are independent for given $\boldsymbol{\alpha}$ and $\mathbf{w}$. To define them, let $M_0 = (-\alpha_{(1)}, \infty)$, $M_1 = (-\alpha_{(2)}, -\alpha_{(1)})$, ..., $M_I = (-\infty, -\alpha_{(I)})$, where $\alpha_{(1)}, \ldots, \alpha_{(I)}$ are the order statistics of the α_i. Let

$$
\begin{aligned}
I_0 &= \{1, \ldots, I\} \\
I_\ell &= \left\{i : \mu_i > \alpha_{(\ell)}\right\}
\end{aligned}
$$

and define

$$
\begin{aligned}
s_{\mu_j \ell} &= \begin{cases} \left(\frac{1}{\sigma_0^2} + \sum_{i \in I_\ell} N_{ij}\right)^{-1} & 0 \le \ell < I \\ \sigma_0^2 & \ell = I \end{cases} \\
m_{\mu_j \ell} &= \begin{cases} s_{\mu_j \ell}\left(\frac{\mu_0}{\sigma_0^2} + \sum_{i \in I_\ell}\left(\sum_{k=1}^{N_{ij}} w_{ijk} - N_{ij}\alpha_i\right)\right) & 0 \le \ell < I \\ 0 & \ell = I \end{cases}
\end{aligned}
$$

$$h_{\mu_j \ell} = \begin{cases} \sum_{i \in I_\ell} \left(N_{ij} \alpha_i^2 - 2\alpha_i \sum_{k=1}^{N_{ij}} w_{ijk} \right) & 0 \le \ell < I \\ 0 & \ell = I \end{cases}$$

Then

$$\begin{aligned} [\mu_j \mid \boldsymbol{\alpha}, \mathbf{w}] \;\; \propto \;\; & \sum_{\ell=0}^{I} \exp\left\{ -\frac{1}{2} \left(h_{\mu_j \ell} - \frac{(m_{\mu_j \ell})^2}{s_{\mu_j \ell}} \right) \right\} \\ & \times \exp\left\{ -\frac{(\mu_j - m_{\mu_j \ell})^2}{2 s_{\mu_j \ell}} \right\} I_{(\mu_j \in M_\ell)} \end{aligned}$$

The right-hand side is proportional to the density of a mixture of truncated normal distributions. Set $\alpha_{I+1} = \infty$ and $\alpha_0 = -\infty$. Let

$$\begin{aligned} q_{\mu_j \ell} \;\; = \;\; & \exp\left\{ -\frac{1}{2} \left(h_{\mu_j \ell} - \frac{(m_{\mu_j \ell})^2}{s_{\mu_j \ell}} \right) \right\} \\ & \times \sqrt{2\pi s_{\mu_j \ell}} \left[\Phi\left(\frac{-\alpha_{(\ell)} - m_{\mu_j \ell}}{\sqrt{s_{\mu_j \ell}}} \right) - \Phi\left(\frac{-\alpha_{(\ell+1)} - m_{\mu_j \ell}}{\sqrt{s_{\mu_j \ell}}} \right) \right], \end{aligned}$$

for $\ell = 0, \ldots, I$ and let $p_{\mu_j \ell} = q_{\mu_j \ell} / \sum_{\ell=0}^{I} q_{\mu_j \ell}$, $\ell = 0, \ldots, I$. Then the full conditional distribution $[\mu_j \mid \boldsymbol{\alpha}, \mathbf{w}]$ is the mixture of truncated normal distributions

$$\mu_j \mid \boldsymbol{\alpha}, \mathbf{w} \;\; \sim \;\; \sum_{\ell=0}^{I} p_{\mu_j \ell} \mathrm{TN}_{M_\ell}(m_{\mu_j \ell}, s_{\mu_j \ell}) \tag{2.22}$$

Proof of Fact 4: The proof of Fact 4 follows analogous steps as the proof of Fact 3 and is omitted for brevity.

2.4.2 Sampling full conditional distributions

Implementing the Gibbs sampler requires samples from the full conditional distributions (2.12), (2.14), (2.20), and (2.22). I provide efficient algorithms below.

Sampling from $[\sigma^2 \mid \boldsymbol{\alpha}]$

Samples from the inverse gamma distribution may be obtained by taking the reciprocal of samples from the gamma distribution (c.f., Ahrens & Dieter, 1974).

Sampling from $[w_{ijk} \mid x_{ij}; y_{ijk}]$

Samples from the truncated normal distribution may be obtained by the inverse CDF method (Devroye, 1986). Let l be the lower bound of the truncated normal and u be the upper bound. Let $U \sim \text{Uniform}(\Phi((l-\mu)/\sigma), \Phi((u-\mu)/\sigma))$. Then,

$$\sigma\Phi^{-1}(U) + \mu \quad \sim \quad \text{TN}_{(l,u)}(\mu, \sigma^2)$$

Sampling from $[\alpha_i \mid \boldsymbol{\mu}, \mathbf{w}, \sigma^2]$ **and** $[\mu_j \mid \boldsymbol{\alpha}, \mathbf{w}]$

Sampling from the full conditionals of α_i and μ_j is more complicated. The full conditional $\alpha_i \mid \boldsymbol{\mu}, \sigma^2$ in (2.20) is a discrete mixture of truncated normals. The first step in sampling α_i is to determine each $p_{\alpha_i \ell}$. One issue is that these evaluations require a high degree of numeric precision. I discuss below approximations to ensure sufficient numeric precision.

From inspection (2.19), it can be seen that integrating α_i over the regions $A_0, \ldots, A_J$ yields terms of the form $\exp\{a\}[\Phi(b) - \Phi(c)]$. Due to numerical imprecision, computing the $q_{\alpha_i \ell}$ terms can be difficult; it is helpful to consider the natural logarithm instead, $a + \log[\Phi(b) - \Phi(c)]$. Because values of b and c may be extreme, standard approximations of Φ may fail. I use Abramowitz and Stegun's (1965) Approximation 26.2.17 of $\Phi(x)$ for $|x| > 5$. Also, because the $q_{\alpha_i \ell}$ terms can be very large, it is helpful to subtract the same constant from each term.

Once $p_{\alpha_i \ell}$ are known, generate a single discrete random variable V such that $\Pr(V = \ell) = p_{\alpha_i \ell}$, $\ell = 0, \ldots, J$, and then sample α_i from the $\text{TN}_{A_v}(m_{\alpha_i \ell}, s_{\alpha_i \ell})$ distribution. Sampling from the full conditional posterior for μ_j is done in the same manner.

Autocorrelation

MCMC techniques, including Gibbs and Metropolis-Hastings sampling, produce a sequence of *correlated* samples, say $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(M)}$, from the posterior distribution. With enough of these samples, it is theoretically possible to estimate any desired quantity from the posterior. For example, the Bayes estimate of θ_i, the posterior mean, can be estimated by the average of the MCMC draws,

$$\hat{\theta}_i = \frac{1}{M} \sum_{m=1}^{M} \theta_i^{(m)}.$$

There is considerable general theory about convergence of MCMC methods (Tierney, 1994). In practice, however, if successive MCMC samples are too highly correlated, it may take extremely long MCMC chains to produce satisfactory estimates of parameters. Even with high speed computers, computation time may take days or weeks. Indeed, the above full-conditional posteriors, when sampled with Gibbs sampling, produce samples of the joint posterior that exhibit a high degree of autocorrelation. An example is provided in Figure 2.3A, which shows the chain for a specific parameter. This chain is typical of almost all of the parameters with Gibbs sampling. The chain was run for 10,000 iterations. To make the autocorrelation clear, I show all samples between iterations 5,000 and 6,000. As can be seen, there is a fair degree of autocorrelations (Figure 2.3B provides a plot for the autocorrelation function). First, I discuss the source of this autocorrelation and then outline additional computational steps that mitigate it.

The source of autocorrelation comes about from the additive nature of the model. The true score x_{ij} is invariant to adding constant values to α_i and μ_j:

$$\begin{aligned} x_{ij} &= \alpha_i + \mu_j \\ &= (\alpha_i + z) + (\mu_j - z) \end{aligned}$$

For any value of z, the likelihood of the model is the same. The consequence is that if the estimate of α_i on a MCMC iteration is perturbed by some z, the estimate of μ_j is expected to be perturbed by $-z$. On the next iteration, the perturbation in μ_j induces a perturbation in α_i by z resulting in positive autocorrelation in parameters across iterations.

To mitigate autocorrelation in the chains, I follow Graves, Speckman and Sun (submitted) who recommend adding additional Metropolis-Hastings steps known as *decorrelating steps* (see also Liu & Sabatti, 2000). The steps proceed as follows:

Step 1. Sample $z \sim \text{Normal}(0, \sigma_z^2)$. The value of σ_z^2 is specified before analysis and determines the acceptance rate of the Metropolis-Hastings step. Appropriate acceptance rates are roughly between .25 and .5.

Step 2. Let $\boldsymbol{\alpha}^{(t)}$ and $\boldsymbol{\mu}^{(t)}$ be the samples of $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ on iteration t of the Gibbs sampler. Let $\alpha_i^{(t,z)} = \alpha_i + z$ and $\mu_j^{(t,z)} = \mu_j - z$. These are candidates for acceptance.

Step 3. Evaluate u, the ratio of the densities of the full posterior distributions for $\boldsymbol{\alpha}^{(t,z)}, \boldsymbol{\mu}^{(t,z)}$ and $\boldsymbol{\alpha}^{(t)}, \boldsymbol{\mu}^{(t)}$:

$$\begin{aligned} u &= \frac{p(\boldsymbol{\alpha}^{(t,z)}, \boldsymbol{\mu}^{(t,z)}, \mathbf{w}, \sigma^2 \mid \mathbf{y})}{p(\boldsymbol{\alpha}^{(t)}, \boldsymbol{\mu}^{(t)}, \mathbf{w}, \sigma^2 \mid \mathbf{y})} \\ &= \exp\left\{ -\frac{1}{2} \left(\sum_{i=1}^{I} \frac{1}{\sigma^2} \left[(\alpha_i^{(t)} + z)^2 - (\alpha_i^{(t)})^2 \right] \right.\right. \\ &\qquad \left.\left. + \sum_{j=1}^{J} \frac{1}{\sigma_0^2} \left[(\mu_j^{(t)} - \mu_0 - z)^2 - (\mu_j^{(t)} - \mu_0)^2 \right] \right) \right\} \end{aligned}$$

Step 4. Accept the new values $\boldsymbol{\alpha}^{(t,z)}, \boldsymbol{\mu}^{(t,z)}$ as our samples $\boldsymbol{\alpha}^{(t)}$ and $\boldsymbol{\mu}^{(t)}$ with probability $\min(u, 1)$.

The steps work because there is greater acceptance of $\boldsymbol{\alpha}^{(t,z)}$ when the $\sum_i \alpha_i^{(t,z)}$ is closer to 0, which is consistent with the prior. Figure 2.3C and 2.3D show that the decorrelating steps do indeed substantially reduce the autocorrelations in the chains. These steps greatly speed convergence with little computational cost and are recommended for additive models in general (Graves et al., submitted).

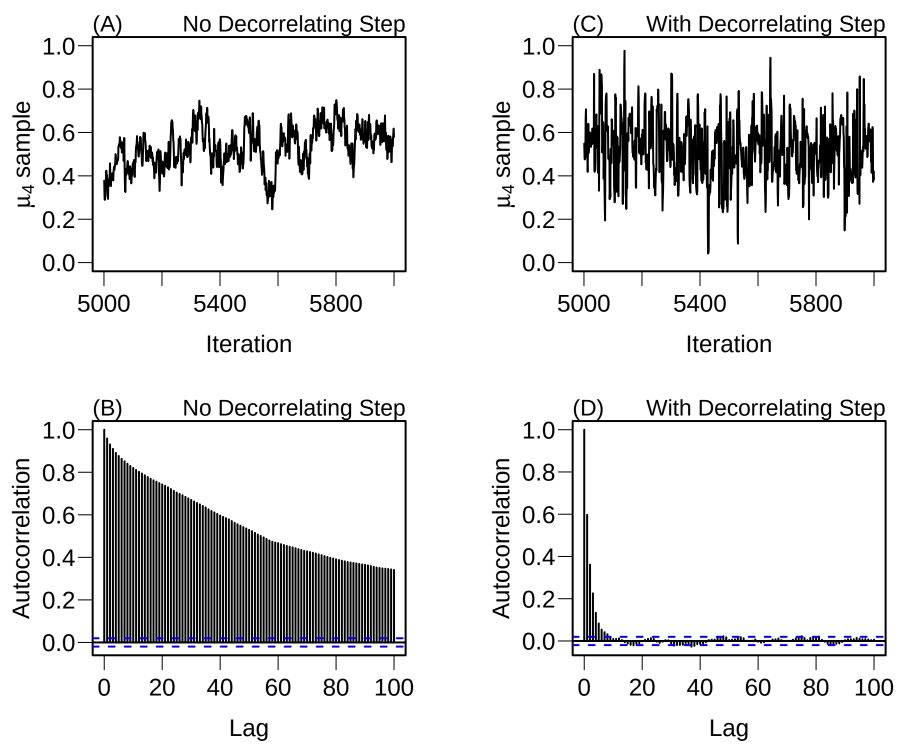

Figure 2.3: Autocorrelation in MCMC Chains. A: Segment from the MCMC chain for parameter μ_4 without a decorrelating step. B: autocorrelation function of parameter μ_4. C & D: Segment from the MCMC chain and autocorrelation with an additional decorrelating step.

2.5 Simulation

To assess the accuracy of parameter estimation and the ability to select person-by-item combination that exhibited performance at chance, I performed a small simulation study. In the simulation, 22 hypothetical participants observed 540 trials. These trials consisted of 90 replicates of 6 hypothetical item eases. These choices of sample sizes reflect the experiment reported subsequently. The "true values" of item ease for the six stimulus intensities were $\boldsymbol{\mu}$=(-1.30, -0.46, -0.13, 0.53, 0.92, 1.18). True values are also needed for latent abilities. True values were sampled from a normal with variance of .31 (this value of variance reflects the results of the subsequent experiment). With these true values, true scores x_{ij} were computed by Eq. (2.7) and true probabilities p_{ij} were computed by Eq. (2.6). For each participant-by-item combination, data were sampled from a binomial with probability p_{ij}. The data were then analyzed with the 1P-MAC model. MCMC chains were run for 10,000 iterations with the first 1,000 iterations serving as a burn-in period. This process of sampling true values, generating data, and estimating parameters was repeated 1000 times.

Figure 2.4A shows estimated latent ability (posterior means) as a function of true latent ability for the $22 \times 1000 = 22,000$ simulated participants. The model is able to recover the latent ability in almost all cases. The exception is for true latent abilities less than -1.18. These latent abilities are for simulated participants who are at chance for all intensities; that is, $\alpha_i < -\mu_6$, where $\mu_6 = 1.18$ is the easiest item. The only information available is that the latent ability is less than $-\mu_6$. The fact that the posterior mean are drawn up toward the value $-\mu_6$ reflects the influence of the hierarchical prior in the absence of information. Most importantly, the model does assess these participants as being at chance for all items.

Estimates of μ (posterior means) are shown in Figure 2.4B. Each density is a smoothed histogram of the 1000 posterior means μ_j to the jth item. The vertical

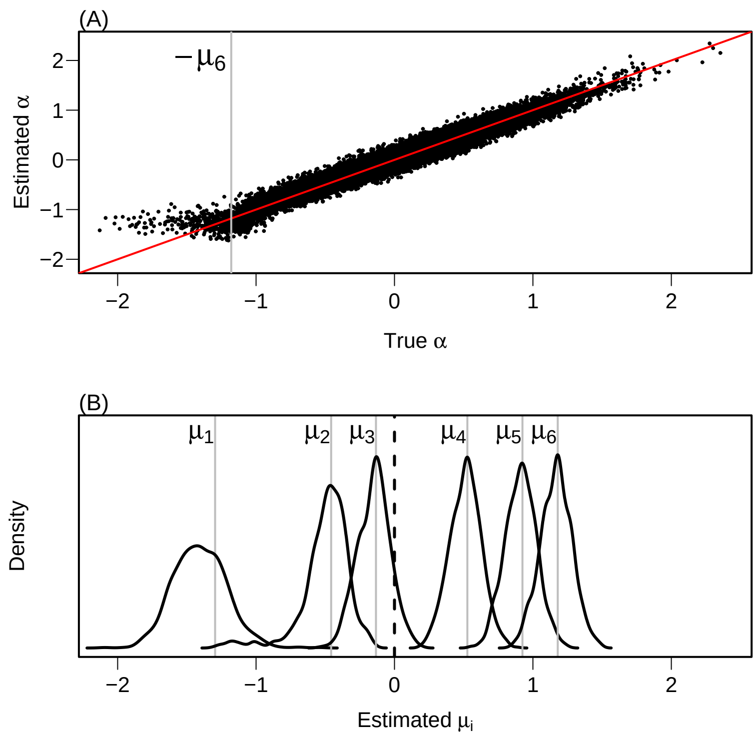

Figure 2.4: Results of simulation. A: Estimates of latent ability α as a function of true value. True latent abilities less than $-\mu_6$ (vertical line) are from simulated participants whose performance is at chance for all of the items. B: Distribution of point estimates (posterior means) of μ_j across replicates (smoothed histograms). The vertical lines represent the true values of each μ.

lines show the true values of μ_j. For all but the two most difficult conditions, the model provides accurate estimates of condition means. For $j = 2$, the condition is so difficult that 80% of the population is below threshold. Estimation is accurate if a small minority of the 22 hypothetical participants has true performance above chance. The long left tail reflects the influence of the prior in replicate experiments in which none of the 22 hypothetical participants was truly above chance. For $j = 1$, the condition is sufficiently difficult that 98% of the population is below chance. The estimate here largely reflects the prior. The simulations show that condition means may be accurately estimated unless all participants are below chance.

The simulations are useful for determining the decision characteristics in classifying participant-by-item combination as subliminal or superliminal. The decision statistic ω_{ij} for the 1P-MAC model is

$$\omega_{ij} = \Pr(x_{ij} \leq 0 \mid \text{data}).$$

As before, the decision rule is to accept participant-by-item combinations as at chance only if $\omega_{ij} \geq .95$. The *level* of this rule may be defined as the probability of classifying a participant-by-item combination as at chance given that the combination is not at chance; the *power* is likewise defined as the the probability of classifying a participant-by-item combination as at chance given that it is. The level and power for this simulation are .001 and .903, respectively. The model is Bayesian and more appropriate decision characteristics are the probability that a participant-by-item combination is truly above-chance chance given that it is classified at chance (a Bayesian equivalent to level) and the probability that a participant-by-item combination is truly above chance given that it is classified above chance (a Bayesian equivalent to power). These decision characteristics are .001 and .930 for the Bayesian equivalents to level and power, respectively.

The decision rule is based on a criterion of .95. One may therefore expect a Bayesian level at .05 rather than at the observed value of .001. The fact that the observed value is so much lower than .05 appears to be a conflict. Indeed, the expectation would hold if we used decision rule "Accept as at-chance if $\omega_{ij} = .95$." The conflict is resolved by noting that all combinations with $\omega_{ij} \geq .95$ are accepted. Therefore, many combinations for which ω_{ij} is greater than .95 will be accepted, yielding a Bayesian level of less than .05.

The simulations on selecting participants results above must be accompanied by a caveat: the results only apply to this set of μ_j and σ^2 and for the simulated sample sizes. In the simulations, there were three items of the six for which more than half of the participants performed above chance. The simulation results may be worse if there are no conditions with a substantial number of above-chance performers.

2.6 Experiment

The following experiment illustrates the application of the MAC model and provides an informal assessment of the additivity assumption in (2.7). The experiment was a conventional number-priming paradigm in which participants alternated between prime identification and target identification tasks. For both prime and target identification, the participant indicated whether the respective digit was greater-than or less-than five. Prime duration was manipulated through several levels from 16.7 ms to 100.0 ms. The prime identification task is the target of inquiry in the MAC model, and an analysis of it is presented here. Unfortunately, in this experiment, there were no clear patterns of priming on target identification and an analysis of this task is omitted. One of the drawbacks of the number-priming paradigm is that it is sensitive to a number of seemingly innocuous factors (see Pratte and Rouder, submitted).

2.6.1 Method

Participants

Twenty-two undergraduate students from the University of Missouri-Columbia introductory psychology pool served as participants. They received class credit for their participation.

Stimuli and Apparatus

Primes and targets were the digits 2, 3, 4, 6, 7, and 8. The symbol “@” was used as a mask. Primes, targets, and masks were displayed in white against a black background. The experiment was displayed on 17” Dell CRTs programmed with a resolution of 640 $\times$ 480 pixels and a refresh rate of 120 Hz and controlled by PC-compatible Pentium IV computers.

Design

The main independent variable was prime duration which was manipulated through the following six levels: 16.7 ms, 25.0 ms, 41.7 ms, 58.3 ms, 75.0 ms, and 100.0 ms. Other independent variables were the prime and targets digits. All combinations of prime, target, and target duration were presented in a within-list design with the exception that primes and targets were always different digits. Each combination occurred equally often in a random order. Prime identification performance served as the main dependent variable.

Procedure

The structure of a trial is shown in Figure 4.2A. Primes were degraded by the inclusion of forward and backwards masks. Trials began with the display of a fixation cross for 458.3 ms. Immediately following, a forward mask was displayed for 58.3 ms. Then, a prime was displayed for one of the six durations. Primes were backward masked for 58.3 ms. Following the backwards mask, a blank-display interval was inserted. The duration of this interval was set such that the SOA between prime and target onsets was 275.0 ms for all prime durations. After the blank interval, the target was displayed until response. Participants pressed the 'z' or '/' keys to indicate that the prime was less-than five or greater-than five, respectively. No feedback was provided. On response, the next trial began with the presentation of the fixation cross. Participants performed 6 blocks of 90 trials, for a total of 540 trials. The session lasted approximately 20 minutes.

2.6.2 Results

Mean and individual accuracies are plotted as a function of duration in Figure 2.5A. Performance is near chance for primes at the shortest duration and increases in an orderly manner.

The data were analyzed with the MAC model. I performed 10,000 MCMC iterations with the first 1,000 serving as burn in. I ran the analysis several times with different starting values to ensure the chains had converged. The resulting chains exhibited low autocorrelation and good convergence. Figure 2.6 shows a representative segment of the MCMC chain (between iterations 5,000 and 6,000) for a few select parameters .

Estimates of μ_j (posterior means) for each duration are shown as the large points in Figure 2.5B. The ease (μ_j) of the first three durations is below 0, indicating that the majority of participants are at chance for these durations.

The relationship between stimulus duration and stimulus ease may be seen by examining how μ_i changes as a function of duration. Figure 2.5B. The function is monotonic and changes slowly. Therefore, a quadratic regression model between stimulus duration and ease seems appropriate and the regression line is shown as the solid black line. The same quadratic regression for each individual may be obtained by adding the latent ability estimate to the intercept-value of the curve; these are shown as gray lines. Finding the lowest positive root of the

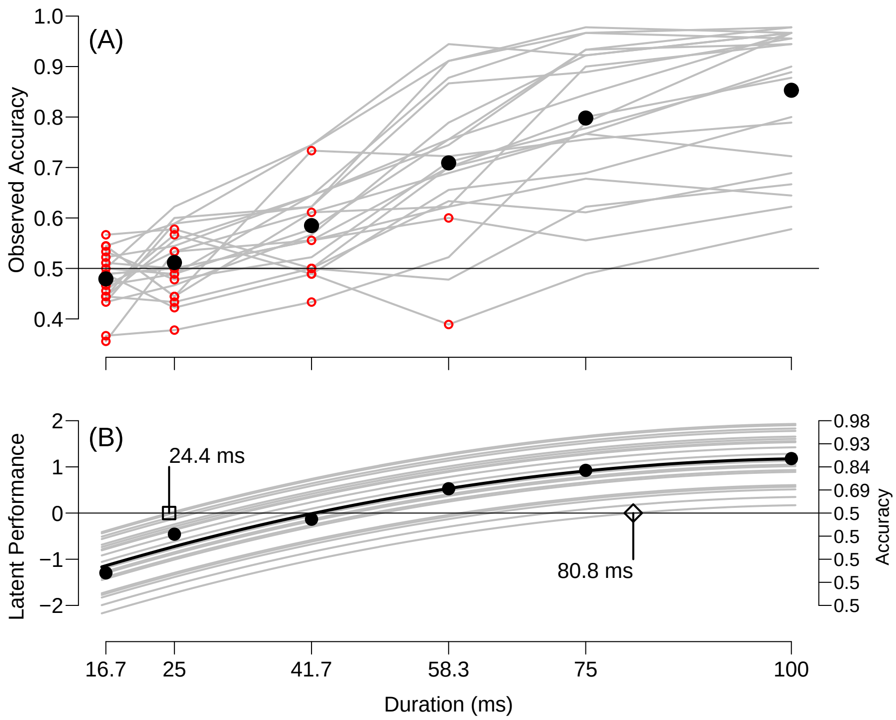

Figure 2.5: A: Mean and individual accuracy are denoted by dark points and light lines, respectively. Open circles denote participant-by-duration combinations that are classified by the MAC model as below threshold. B: Points are estimates of duration condition ease μ_j. Each line is a quadratic regression of true score on prime duration. The square and diamond symbols denote the smallest and largest thresholds estimated across the 22 participants.

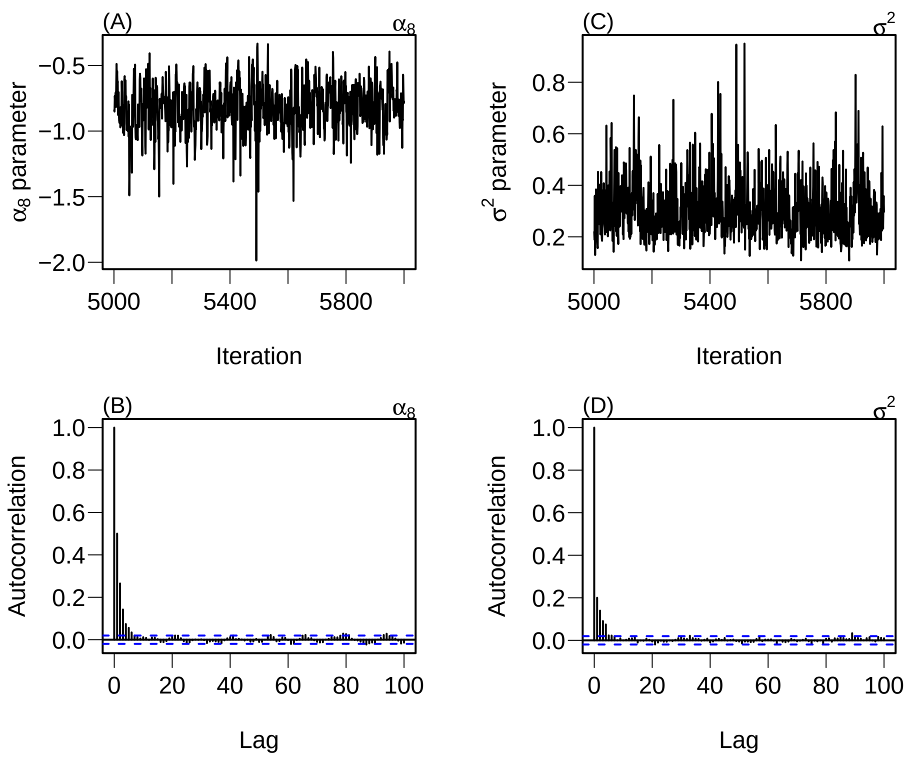

Figure 2.6: Segments from the MCMC chains and autocorrelation functions for selected parameters.

quadratic regression yields an estimate of a participant's threshold. The smallest and largest individual thresholds are indicated in Figure 2.5B as a square and diamond, respectively. Of note is that I did not assume a quadratic form until it was suggested by the data.

To assess the fit of the model, I plotted observed accuracy as a function of predicted accuracy (Figure 2.7A). If the model fits well, the points for true accuracies above chance should fall near the diagonal line. Points for true accuracies at chance should be between the two horizontal dashed lines that denote the 95% confidence interval on true at-chance performance. Overall, the model does well, with the exception of the one troubling point, denoted by "a". This participant attained an accuracy of 73% for the duration of 41.7 ms, yet the model classified this high level of performance as at chance. Because such a level of performance is highly unlikely given a true probability at chance, the model is misspecified to some degree for this participant.

I explored possible misspecification further by computing standardized residuals, r_{ij}, for each participant-by-duration combination:

$$r_{ij} = \frac{y_{ij} - N_{ij}\hat{p}_{ij}}{\sqrt{N_{ij}\hat{p}_{ij}(1 - \hat{p}_{ij})}} \tag{2.23}$$

where $\hat{p}_{ij}$ is the predicted accuracy for the participant and duration. The lines in Figure 2.7B shows these residuals for participants when $\hat{p}_{ij} > .5$. The grey lines denote participants for which the model seems well specified and additivity appears to hold. Of the 22 participants, 18 fall into this group. The dark lines, in contrast, show misspecification. The line denoted by "a" has decreasing residuals indicating less gain in true score per unit increase in duration for this participant than for other participants. The line denoted with "b"shows the opposite pattern. This participant gains more in true score per unit increase in duration than other participants. These differences in the rate of gain are violations of additivity. Fortunately, these violations occur for a small minority of participants. Unfortunately, these violations negatively affect the estimation

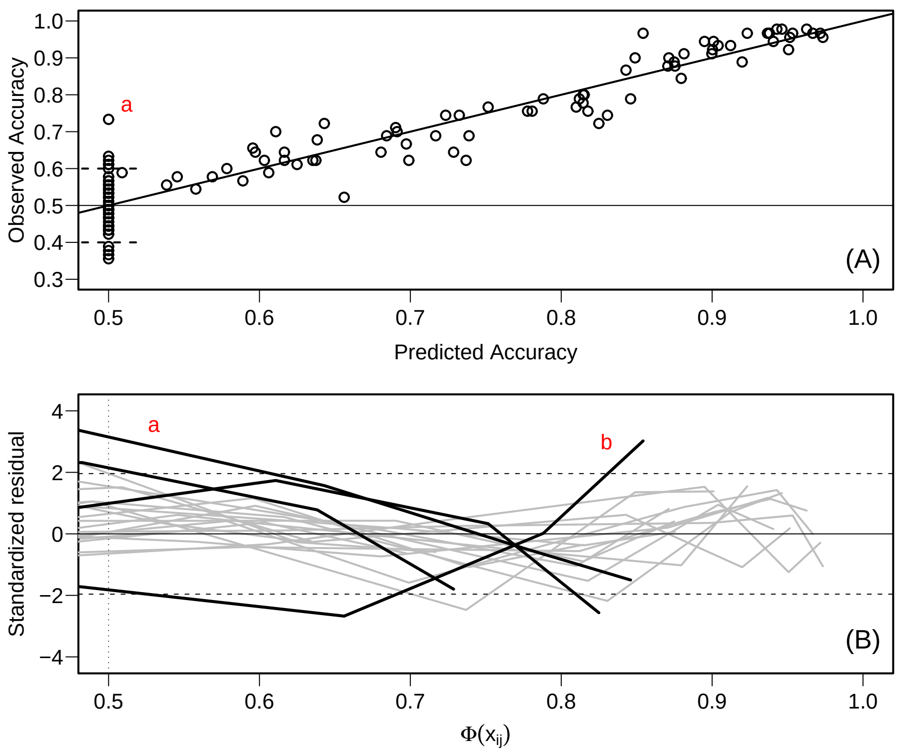

Figure 2.7: A: Observed accuracy as a function of predicted accuracy. Each point denotes a participant-by-duration combination. Horizontal dashed lines represent the .025 and .975 quantiles of accuracy for a participant at chance. B: Standardized residuals from the MAC model. Light and dark lines denote participants for which the model is well-specified and misspecified, respectively. Horizontal dashed lines at -1.96 and 1.96 represent approximate 95% bounds on the residuals.

of thresholds. The extreme point denoted “a” in Figure 2.7A arises from the participant denoted with an “a” in Figure 2.7B as a consequence of misspecification. Researchers may need to analyze residuals to identify participants for whom the model is misspecified. In Chapter 3, I develop generalize the 1P-MAC model to include individualized slope parameters to account for this obvious misspecification.

2.7 General Discussion

The MAC model for multiple prime duration conditions is a valuable psychometric model for psychophysical applications. It provides for accurate measurement of latent abilities and true scores above and below threshold. Consequently, it may be used to accurately measure thresholds, assess which participant-by-item combinations are truly below threshold, and test theories of priming. The model is a vast improvement on current methods of determining whether stimuli are above or below threshold.

In the remainder of this discussion, I discuss the relationship of the MAC model to item response theory. In many senses, the MAC model is more similar to IRT than it is to psychophysics. It shares with the former a commitment about how latent scores relate to performance rather than how physical characteristics relate to performance. The major strength and drawback to the current model is the additivity assumption: $x_{ij} = \alpha_i + \mu_j$. Much like in item response theory, this assumption is leveraged to provide better estimation of true scores. I remain concerned, however, about a lack of robustness to misspecification. Although additivity holds for the majority of our 22 participants, there are noteworthy exceptions (Figure 2.7B). If the rate of gain is different for each participant, additivity will be violated. Fortunately, additivity can be assessed informally by studying residuals, and researchers should do so.

The additivity assumption developed here is analogous to that in the Rasch model (Lord & Novick, 1968). The Rasch model for this paradigm is given by

$$\begin{aligned} p_{ij} &= \frac{1}{2} + \frac{1}{2}F(\theta[\gamma_i - \eta_j]), \\ \gamma_i &\overset{\text{iid}}{\sim} \text{Normal}(0,1) \end{aligned}$$

where γ_i is the ith participant's ability; η_j is the difficulty of primes in the j duration condition, and θ is general discriminability parameter. The function F is the link and it is often taken as a logistic function:

$$F(x) = \frac{1}{1+\exp(-x)}.$$

The MAC model may be reparameterized in terms of (θ, γ, η), where

$$\begin{aligned} \theta &= \sigma, \\ \gamma_i &= \alpha_i/\sigma, \\ \eta_j &= -\mu/\sigma. \end{aligned}$$

With this reparameterization, the MAC model may be expressed as

$$\begin{aligned} p_{ij} &= \frac{1}{2} + \frac{1}{2}G(\theta[\gamma_i - \eta_j]), \\ \gamma_i &\overset{\text{iid}}{\sim} \text{Normal}(0,1) \\ G(x) &= \begin{cases} 2[\Phi(x) - .5] & x \geq 0 \\ 0 & x < 0 \end{cases} \end{aligned}$$

As can be seen, the sole difference between the MAC model and the Rasch model is the link function. The MAC model has a half-probit link; the Rasch model has a logistic link. These two links are shown in Figure 2.2. The MAC link is concordant with the concept of a threshold and posits that some measurable fraction of the population will have true performance at chance for each item. The logistic link is not. Chance performance only occurs for true scores of $-\infty$. In other words, the logistic link posits that all participants are above chance for each item. The MAC link is certainly more appropriate than the logistic link for measuring at-chance thresholds. It may be appropriate in many other measurement contexts as well, discussed in Chapter 5. In the next Chapter, I discuss a generalization of the model presented above, which relaxes the additivity assumption and allows each participant to improve at a different rate as item difficulty decreased.

Chapter 3

The two-parameter Mass at Chance model

In Chapter 2, I developed the 1P-MAC model for multiple items and used it to estimate individuals' thresholds. In this Chapter, I develop a 2P-MAC model. The motivation for this extension is a noticeable misspecification of the one-parameter model in a the analysis of data from a digit identification task discussed in Chapter 2. Below, present the 2P-MAC model, benchmark its performance using simulations, and show how the two-parameter extension accounts for the data set described in Chapter 2.

3.1 The 2P-MAC Model

3.1.1 Model specification

The 2P-MAC model is different from the 1P-MAC model in the addition of idndividualized slope parameters θ_i:

$$p_{ij} = \begin{cases} \Phi(\theta_i[\alpha_i - \beta_j]), & \alpha_i > \beta_j, \\ .5 & \beta_j \geq \alpha_i, \end{cases} \tag{3.1}$$

In the following development, it is convenient to redefine the latent true score, x_{ij} for each participant-by-item combination: $x_{ij} = \theta_i(\alpha_i - \beta_j)$.

The model needs an additional constraint. In conventional two-parameter IRT models with item discriminability, this constraint is typically placed on the variance of participant effects. In the model in Eq. 3.1, in which discriminability varies by individuals rather than items, it is convenient to place the constraint on item difficulties rather than participant effects:

$$\beta_j \stackrel{iid}{\sim} \text{Normal}(0, 1). \tag{3.2}$$

This is the same as the prior constraint on β_j in Chapter 2. I show subsequently that this constraint works well in simulation and application.

The model is analyzed in the Bayesian framework with Markov chain Monte Carlo sampling (Geman & Geman, 1984, Gelfand & Smith 1990, Gelman, Carlin, Stern, & Rubin, 2004). Consequently, priors are needed for parameters. The following hierarchical priors are convenient for participant effects:

$$\begin{aligned}
\alpha_i &\stackrel{iid}{\sim} \text{Normal}(0, \sigma_\alpha^2), \\
\theta_i &\stackrel{iid}{\sim} \text{TN}_{\Re+}(0, \sigma_\theta^2), \\
\sigma_\alpha^2 &\sim \text{Inverse Gamma}(a_1, b_1) \\
\sigma_\theta^2 &\sim \text{Inverse Gamma}(a_2, b_2),
\end{aligned}$$

where TN_A denotes a normal distribution truncated to an interval A. The inverse gamma prior for variance parameters has pdf

$$f(x \mid a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp\left\{-\frac{b}{x}\right\}$$

for shape parameter a and scale parameter b. Values of a_1, b_1, a_2 and b_2 must be specified before analysis. We find $a_1 = a_2 = 2$, and $b_1 = b_2 = 1$ to work well in simulation and application.

3.1.2 Full Conditional Distributions

As in Chapter 2, derivation of the full conditional distributions is aided by the addition of the latent variables (Albert & Chib, 1993). Let y_{ijk} be the dichotomous response for the ith participant observing the jth item for the kth replicate, $k = 1, \ldots, N_{ij}$. Let w_{ijk} denote latent data such that $y_{ijk} = 1 \iff w_{ijk} > 0$. Then

$$w_{ijk} \stackrel{ind}{\sim} \text{Normal}(\theta_i(\alpha_i - \beta_j) \vee 0, 1)$$

It is more convenient to place the hierarchical model on latent data $\mathbf{w}$ than on observed data $\mathbf{y}$. As in Chapter 2, I use bold-face notation throughout to represent collections and vectors of data or parameters. The following facts describe all the full conditional distributions used in MCMC analysis, including those for the latent data $\mathbf{w}$.

Fact 1. The full conditional posteriors of w_{ijk}, σ^2_α, and σ^2_θ are

$$\begin{aligned} w_{ijk} \mid \cdot &\sim \begin{cases} \text{TN}_{\Re^+}(x_{ij} \vee 0, 1) & y_{ijk} = 1 \\ \text{TN}_{\Re^-}(x_{ij} \vee 0, 1) & y_{ijk} = 0 \end{cases} \\ \sigma^2_\alpha \mid \cdot &\sim \text{IG}\left(a_1 + \frac{I}{2}, b_2 + \frac{1}{2}\sum_{i=1}^{I} \alpha_i^2\right) \\ \sigma^2_\theta \mid \cdot &\sim \text{IG}\left(a_2 + \frac{I}{2}, b_2 + \frac{1}{2}\sum_{i=1}^{I} \theta_i^2\right), \end{aligned}$$

where $\cdot$ refers to all other parameters and data.

Fact 2. The full conditional posteriors of α_i are independent for given $i, \boldsymbol{\beta}$, θ_i, σ^2_α, and $\mathbf{w}$. It is convenient to define the following sets. Let $A_0 = (-\infty, \beta_{(1)})$, $A_1 = (\beta_{(1)}, \beta_{(2)})$, …, $A_J = (\beta_{(J)}, \infty)$, where $\beta_{(1)} < \ldots < \beta_{(J)}$ are the order statistics of the β_j. Let

$$J_\ell = \left\{j : \beta_j \leq \beta_{(\ell)}\right\} \tag{3.3}$$

and define

$$s_{\alpha_i \ell} = \begin{cases} \left(\frac{1}{\sigma^2_\alpha} + \sum_{j \in J_\ell} N_{ij}\theta_i^2\right)^{-1} & 0 < \ell \leq J, \\ \sigma^2_\alpha & \ell = 0, \end{cases} \tag{3.4}$$

$$m_{\alpha_i \ell} = \begin{cases} s_{\alpha_i \ell} \sum_{j \in J_\ell} \left(\theta_i \sum_{k=1}^{N_{ij}} w_{ijk} + N_{ij} \theta_i^2 \beta_j \right) & 0 < \ell \leq J, \\ 0 & \ell = 0, \end{cases} \tag{3.5}$$

$$h_{\alpha_i \ell} = \begin{cases} \sum_{j \in J_\ell} \left(N_{ij} \theta_i^2 \beta_j^2 + 2\theta_i \beta_j \sum_{k=1}^{N_{ij}} w_{ijk} \right) & 0 < \ell \leq J, \\ 0 & \ell = 0, \end{cases} \tag{3.6}$$

Then

$$[\alpha_i \mid \cdot] \propto \sum_{\ell=0}^{J} \exp\left\{ -\frac{1}{2} \left(h_{\alpha_i \ell} - \frac{(m_{\alpha_i \ell})^2}{s_{\alpha_i \ell}} \right) \right\} \times \exp\left\{ -\frac{(\alpha_i - m_{\alpha_i \ell})^2}{2 s_{\alpha_i \ell}} \right\} I_{(\alpha_i \in A_\ell)} \tag{3.7}$$

Following Bayesian convention, $[\alpha_i \mid \cdot]$ denotes the density of α_i given all other parameters and data. The right-hand side of (3.7) is proportional to the density of a mixture of truncated normals. Set $\beta_{(J+1)} = \infty$ and $\beta_{(0)} = -\infty$. Let

$$q_{\alpha_i \ell} = \exp\left\{ -\frac{1}{2} \left(h_{\alpha_i \ell} - \frac{(m_{\alpha_i \ell})^2}{s_{\alpha_i \ell}} \right) \right\} \times \sqrt{2\pi s_{\alpha_i \ell}} \left[\Phi\left(\frac{\beta_{(\ell+1)} - m_{\alpha_i \ell}}{\sqrt{s_{\alpha_i \ell}}} \right) - \Phi\left(\frac{\beta_{(\ell)} - m_{\alpha_i \ell}}{\sqrt{s_{\alpha_i \ell}}} \right) \right], \tag{3.8}$$

for $\ell = 0, \ldots, J$ and let $p_{\alpha_i \ell} = q_{\alpha_i \ell} / \sum_{\ell=0}^{J} q_{\alpha_i \ell}$, $\ell = 0, \ldots, J$. Then the full conditional distribution $[\alpha_i \mid \cdot]$ is the mixture of truncated normal distributions

$$\alpha_i \mid \cdot \sim \sum_{\ell=0}^{J} p_{\alpha_i \ell} \mathrm{TN}_{A_\ell}(m_{\alpha_i \ell}, s_{\alpha_i \ell}) \tag{3.9}$$

Fact 3. The full conditional posterior distributions of the β_j are independent for given $j, \boldsymbol{\alpha}, \boldsymbol{\theta}$, and $\mathbf{w}$. To define them, let $M_0 = (\alpha_{(J)}, \infty)$, $M_1 = (\alpha_{(J-1)}, \alpha_{(J)})$, …, $M_I = (-\infty, \alpha_{(1)})$, where $\alpha_{(1)}, \ldots, \alpha_{(I)}$ are the order statistics of the α_i. Let

$$I_\ell = \left\{ i : \alpha_i > \alpha_{(I-\ell)} \right\}$$

and define

$$
\begin{aligned}
s_{\beta_j\ell} &= \begin{cases} \left(\frac{1}{\sigma_\beta^2} + \sum_{i\in I_\ell} N_{ij}\theta_i^2\right)^{-1} & 0 < \ell \leq I, \\ \sigma_\beta^2 & \ell = 0, \end{cases} \\
m_{\beta_j\ell} &= \begin{cases} s_{\beta_j\ell}\left(-\sum_{i\in I_\ell}\left(\theta_i \sum_{k=1}^{N_{ij}} w_{ijk} - N_{ij}\theta_i^2\alpha_i\right)\right) & 0 < \ell \leq I, \\ 0 & \ell = 0, \end{cases} \\
h_{\beta_j\ell} &= \begin{cases} \sum_{i\in I_\ell}\left(N_{ij}\theta_i^2\alpha_i^2 - 2\theta_i\alpha_i \sum_{k=1}^{N_{ij}} w_{ijk}\right) & 0 < \ell \leq I, \\ 0 & \ell = 0, \end{cases}
\end{aligned}
$$

Then

$$
\begin{aligned}
[\beta_j \mid \cdot] \;\propto\; & \sum_{\ell=0}^{I} \exp\left\{-\frac{1}{2}\left(h_{\beta_j\ell} - \frac{(m_{\beta_j\ell})^2}{s_{\beta_j\ell}}\right)\right\} \\
& \times \exp\left\{-\frac{(\beta_j - m_{\beta_j\ell})^2}{2s_{\beta_j\ell}}\right\} I_{(\beta_j\in M_\ell)}
\end{aligned}
$$

The right-hand side is proportional to the density of a mixture of truncated normal distributions. Set $\alpha_{(I+1)} = \infty$ and $\alpha_{(0)} = -\infty$. Let

$$
\begin{aligned}
q_{\beta_j\ell} &= \exp\left\{-\frac{1}{2}\left(h_{\beta_j\ell} - \frac{(m_{\beta_j\ell})^2}{s_{\beta_j\ell}}\right)\right\} \\
& \quad \times \sqrt{2\pi s_{\beta_j\ell}}\left[\Phi\left(\frac{\alpha_{(I-\ell+1)} - m_{\beta_j\ell}}{\sqrt{s_{\beta_j\ell}}}\right) - \Phi\left(\frac{\alpha_{(I-\ell)} - m_{\beta_j\ell}}{\sqrt{s_{\beta_j\ell}}}\right)\right],
\end{aligned}
$$

for $\ell = 0,\ldots,I$ and let $p_{\beta_j\ell} = q_{\beta_j\ell}/\sum_{\ell=0}^{I} q_{\beta_j\ell}$, $\ell = 0,\ldots,I$. Then the full conditional density $[\beta_j \mid \cdot]$ is the mixture of truncated normal distributions

$$
\beta_j \mid \cdot \;\sim\; \sum_{\ell=0}^{I} p_{\beta_j\ell}\mathrm{TN}_{M_\ell}(m_{\beta_j\ell}, s_{\beta_j\ell}) \tag{3.10}
$$

Fact 4. The full conditional posterior distributions of θ_i are independent for given α_i, $\boldsymbol{\beta}$, and $\mathbf{w}$. Let

$$
\begin{aligned}
J^* &= \{j : \beta_j < \alpha_i\}. \\
s_{\theta_i} &= \left(\frac{1}{\sigma_\theta^2} + \sum_{j\in J^*} N_{ij}(\alpha_i - \beta_j)^2\right)^{-1}
\end{aligned}
$$

$$m_{\theta_i} = s_{\theta_i} \sum_{j \in J^*} (\alpha_i - \beta_j) \sum_{k=1}^{N_{ij}} w_{ijk}$$

Then

$$\theta_i \mid \cdot \sim \mathrm{TN}_{\Re^+}(m_{\theta_i}, s_{\theta_i}) \tag{3.11}$$

The proofs of these four facts are similar to the analogous proofs in Chapter 2, and are omitted for brevity.

3.1.3 Model Analysis

Like the 1P-MAC model, the Gibbs sampler for the 2P-MAC model produces highly autocorrelated chains. Figure 3.1A shows a chain for a typical item parameter, β_4, from the data set discussed in Chapter 2. This chain is highly autocorrelated, as shown by the slowly wandering nature of the chain. The source of the autocorrelation is straightforward to diagnose. As shown in Eq. (3.1), the likelihood of the model given the parameters is a function of $\theta_i(\alpha_i - \beta_j)$. The likelihood is invariant to the inclusion of additive and multiplicative constants as follows: Let $\alpha_i^* = v(\alpha_i + z)$, $\beta_j^* = v(\beta_j + z)$ and $\theta_i^* = \theta_i / v$. Then the likelihood is invariant to all values of z and v. Hence, values of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ cannot be identified from the likelihood alone; identifiability comes from the random effects assumptions expressed in the hierarchical prior. Even so, the constraint from the prior on the iteration-to-iteration values in the chains is small. The result is correlation; for example when the α parameters are estimated high on average, the β parameters will shift as well. Likewise, a large estimate of θ attenuates the estimates of α and β.

As in Chapter 2, additional Metropolis-Hastings steps may be added to decorrelate the chains. For 2P-MAC, two decorrelating Metropolis steps are needed. The needed additive step to account for the invariance of the likelihood as a

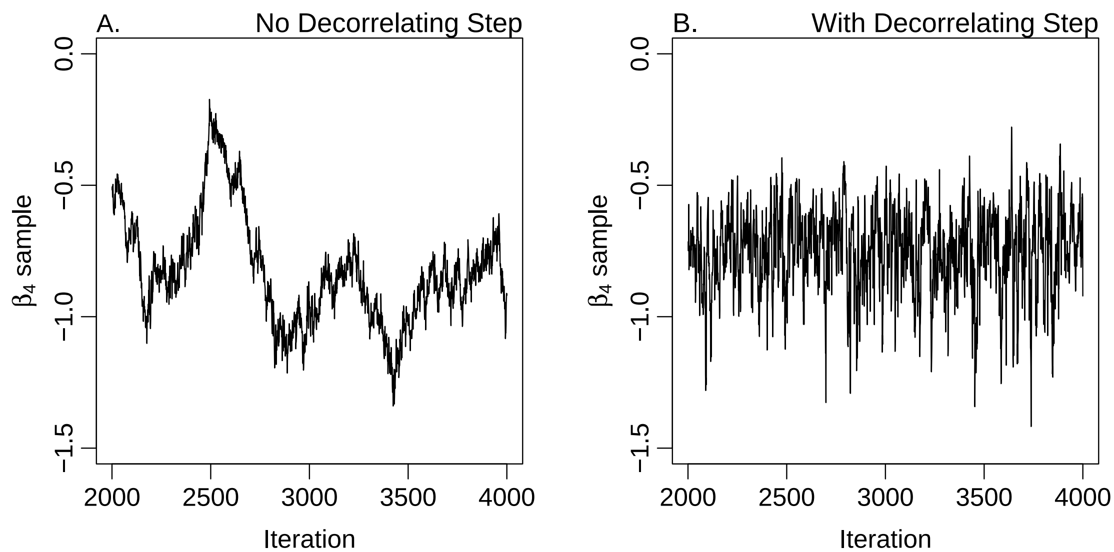

Figure 3.1: A. Autocorrelation for a select item effect (β_4) from Gibbs Sampling. B. Greatly reduced autocorrelation from the inclusion of additional decorrelating Metropolis steps. The degree of autocorrelation is typical for all participant and item parameters.

function of z noted above is described in Chapter 2. The step to account for the invariance of the likelihood as a function of v is described here.

On every iteration of the chain, the Metropolis step proceeds as follows:

Step 1. Sample v from some proposal distribution restricted to positive values. Let $\pi(v)$ denote the density function of the proposal distribution. Selection of this distribution will be discussed later.

Step 2. Let $\boldsymbol{\theta}^{(t)}$, $\boldsymbol{\alpha}^{(t)}$, and $\boldsymbol{\beta}^{(t)}$ be the samples of $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ on iteration t of the Gibbs sampler. Let $\theta_i^{(t,v)} = \theta_i^{(t)}/v$, $\alpha_i^{(t,v)} = v\alpha_i^{(t)}$, and $\beta_j^{(t,v)} = v\beta_j^{(t)}$ on iteration t. These are candidates for acceptance.

Step 3. Evaluate u, the ratio of the posterior distributions for the candidate and the current sample:

$$\begin{aligned} u &= \frac{p(\boldsymbol{\theta}^{(t,v)}, \boldsymbol{\alpha}^{(t,v)}, \boldsymbol{\beta}^{(t,v)}, \mathbf{w}, \sigma^2 \mid \mathbf{y})}{p(\boldsymbol{\theta}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t)}, \mathbf{w}, \sigma^2 \mid \mathbf{y})} \\ &= \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_\theta^2}\sum_{i=1}^{I}\left[(\theta_i^{(t)}/v)^2 - (\theta_i^{(t)})^2\right]\right.\right. \\ &\qquad + \sum_{i=1}^{I}\frac{1}{\sigma_\alpha^2}\left[(v\alpha_i^{(t)})^2 - (\alpha_i^{(t)})^2\right] \\ &\qquad \left.\left. + \sum_{j=1}^{J}\left[(v\beta_j^{(t)})^2 - (\beta_j^{(t)})^2\right]\right)\right\} \end{aligned}$$

This term is a measure of how well the candidate values fit when compared with the samples from the tth iteration of the chain.

Step 4. Compute r:

$$r = \frac{v^J \pi(v^{-1}) v^{-2}}{\pi(v)}$$

This term ensures that the steps yield samples with the proper joint distribution.

Step 5. Accept candidate values $\boldsymbol{\theta}^{(t,v)}, \boldsymbol{\alpha}^{(t,v)}, \boldsymbol{\beta}^{(t,v)}$ with probability $\min(ur, 1)$.

A candidate distribution π must be specified. It is convenient if the expected value is about 1 to insure that both $v < 1$ and $v > 1$ are likely. Gamma distributions with equal rate and shape have this property. The variance of the proposal distribution may be tuned to provide for a desired acceptance rate in Step 5. I have found acceptance rates of between .25 and .5 provide for sufficient decorrelation.

Figure 3.1B shows a chain for a typical item parameter, β_4, with the inclusion of the additive decorrelating step described in Chapter 2 and the multiplicative step described above. These steps do a good job of decorrelating the chains, leading to parameter estimates in minutes instead of hours.

3.2 Simulations

In order to assess the performance of the model, I performed simulations. The goal of these simulations is to determine whether model analysis can recover true parameters in reasonable sample sizes. In each of 500 simulations, 22 hypothetical participants judged 6 items 90 times each. These sample sizes are from the experiment discussed in Chapter 2. True values of the six β parameters were (1.17,0.35,-0.25, -0.91, -1.45, -1.89). True values for α_i and θ_i were drawn from independent Normal($0,\sigma^2_\alpha = .14$) and $\mathrm{TN}_{\Re^+}(0, \sigma^2_\theta = .62)$ distributions, respectively. These true values come from the subsequent analysis of the data set in Chapter 2.

For each of the 500 simulations, values of α_i and θ_i were drawn from their respective distributions. These values were then used to compute true latent scores x_{ij}. The true accuracies corresponding to these latent scores were used to generate binomial data for each participant-by-item combination. These simulated data were then analyzed via the 2P-MAC model. MCMC chains were run for 10,000 iterations with the first 1,000 serving as burn-in.

Figure 3.2A shows estimated participants ability parameters (posterior means) as a function of true values. In total, there are $500 \times 22 = 11000$ true participant abilities. Shown are two-dimensional kernel densities of these 11000 points. Overall, the points cluster on the diagonal showing good recovery. There is a small degree of curvilinearity due to difficulty estimating participants who have extremely low latent ability (those 2 standard deviations below the mean). These participants are at chance for many of the items. Consequently, there is a paucity of information to discriminate how little ability they have. This paucity leads to a greater influence of the prior and subsequent shrinkage of the estimate. The effect is slight and occurs only for those simulated participants who had near baseline accuracies for all items. Whereas these participants inherently provide no information other than that they perform poorly, the shrinkage is not problematic. Figure 3.2B and 3.2C show the same plots for discriminability and true score. The points lie close to the diagonal, indicating good recovery of participant discriminability and latent score estimates.

Recovery of the true values of $\beta_1, \ldots, \beta_6$ is more nuanced. Figure 3.2D-F show CDF coverage plots (Morey, Pratte, & Rouder, in press) for parameters β_1, β_2, and β_3. Each dark point denotes the point estimate for item difficulty from a single simulation (there are 500 points). These point estimates are sorted in order and presented as an empirical CDF plot. On either side of each dark point is a smaller, light point. The area between these two light points is the 95% credible interval that corresponds to the point estimate. The solid vertical line denotes the true value.

The recovery for the most difficult item (β_1, Figure 3.2D) is quite good, and this is surprising. In fact, the true value is so low that only 1 in 1000 simulated participants have true performance above baseline to this item. Hence, there is very little information in the data and recovery should be quite poor. The reason for this is that the estimate of 1.17 is due solely to the prior. The

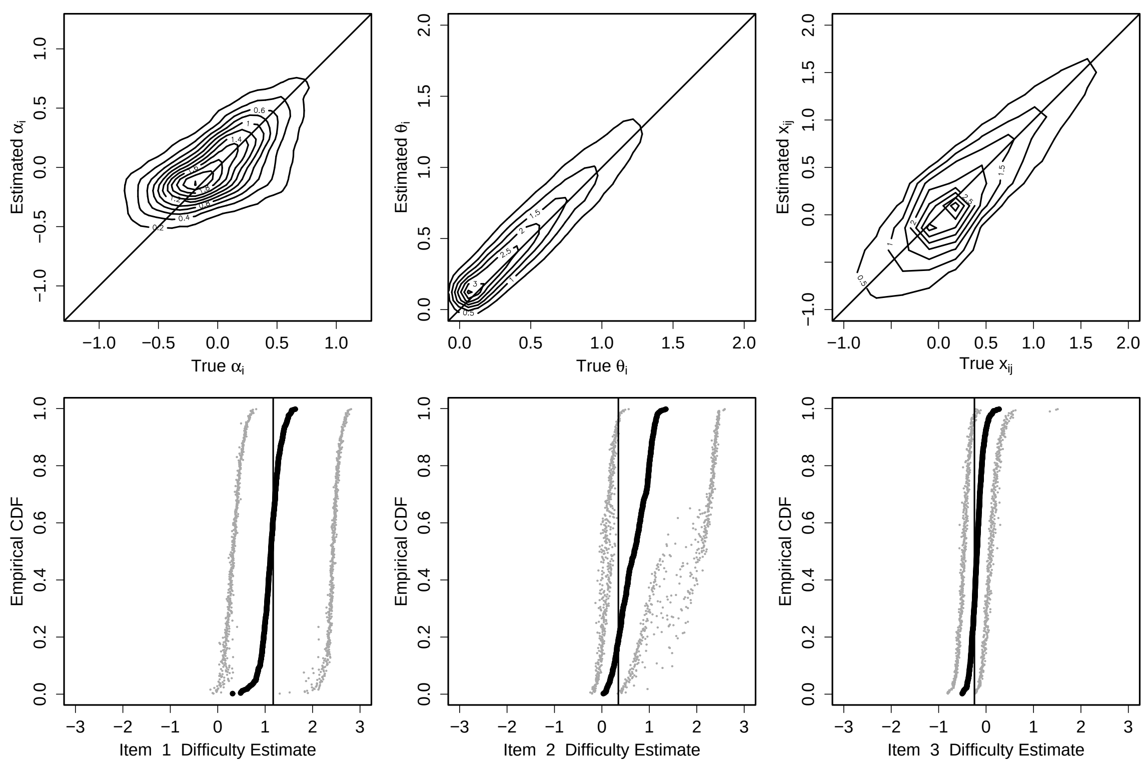

Figure 3.2: Simulation results. A-C: Two-dimensional kernel density estimates for estimated value as a function of true value for the parameters α_i, θ_i, and $x_{ij} = \theta_i(\alpha_i - \beta_j)$, respectively. D-F: CDF coverage plots for model estimates of β_1, β_2, and β_3. Dark points represent point estimates (posterior mean); light points on either side of each dark point represent corresponding 95% posterior credible intervals. The dark, solid vertical line represents the true value.

standard normal prior on difficulties limits how large this estimate may be. In this case, for this true value and sample size, the shrinkage from the prior almost leads serendipitously to near perfect estimates. The lower bound does reflect information; the true value must be greater than it. The upper bound and posterior mean, however, simply reflect the choice of the prior.

Figure 3.2E shows the CDF coverage plot for β_2, the second most difficult item. Inspection of the upper bound on difficulty reveal a mixture of two types of simulations. On some simulations, no estimated abilities were above β_2, and the estimated value of β_2 reflected the prior. On other simulations, there were estimated abilities above β_2 and the estimate of β_2 reflects these values. This mixture leads to estimates of β_2 that are upwardly biased. The credible intervals, fortunately, quantify the amount of uncertainty well: over 95% of the credible intervals contain the true value.

Figure 2.4F shows the CDF coverage plot for parameter β_3. On almost all simulations, there were estimated abilities greater than this item's difficulty. Hence, there was much information to localize the parameter and the prior had almost no effect. Credible intervals are narrow and over 95% contain the true value. The average parameter estimate is close to the true value for the parameter. The remaining item difficulty parameters are estimated as well as β_3. In sum, the model does well to estimate item difficulty if at least one participant is estimated above chance for all items. Therefore, items that are so hard such that nobody does well are not useful. This problem is not unique to the MAC model; it applies for conventional IRT models as well.

As discussed in Chapter 2 (see Eq. 2.3), participant-by-item combinations may be classified as at-chance or above chance through placing a criterion on $\omega_{ij} = \Pr(x_{ij} < 0 \mid \text{data})$. Participant-by-item combinations above the criterion are classified as at-chance, while those below the criterion are classified as above chance. Because classification with the 2P-MAC model is used in Chapter 4 to study the phenomenon of subliminal priming, it is important to examine classification rates in the reported simulations.

Using a conservative criterion of .95 on ω_{ij}, the model correctly classified participant-by-item combinations that are truly at chance as at-chance 59% of the time. The model incorrectly classified combinations above-chance as at-chance only .6% of the time. It is also desirable to determine the rates conditioned on classification instead of true chance status. Only 2% of classifications at chance were actually above chance. Given a classification of above chance, 82% of the classifications were correct.

It is also interesting to consider the effect of choosing a more liberal criterion on ω_{ij}. For instance, in Experiment 1 in Chapter 4, I will use the more liberal criterion of .5 to classify participant-by-item combinations as at-chance. With the .5 criterion on ω_{ij}, the model correctly classified participant-by-item combinations that are truly at chance as at-chance 93% of the time. The model incorrectly classified combinations above-chance as at-chance only 6% of the time. Given an at-chance classification, 11% of the classified combinations were incorrect, while 96% of the above-chance classifications were correct. By manipulating the criterion, one type of error can be decreased at the expense of increasing the opposite error. The usefulness of manipulating the criterion in this way will be discussed in Chapter 4.

3.3 Analysis of a data set

In the previous section, I noted the misspecification of 1P-MAC to the data set described in Chapter 2. In this section, I show the 2P-MAC analysis. The classification accuracy of each participant to each duration is shown in Figure 3.3A; each line represents the accuracy of a participant. Average performance is near chance for the most difficult durations, and increases for higher durations.

The data were analyzed with the 2P-MAC model. Chains were run for 10,000

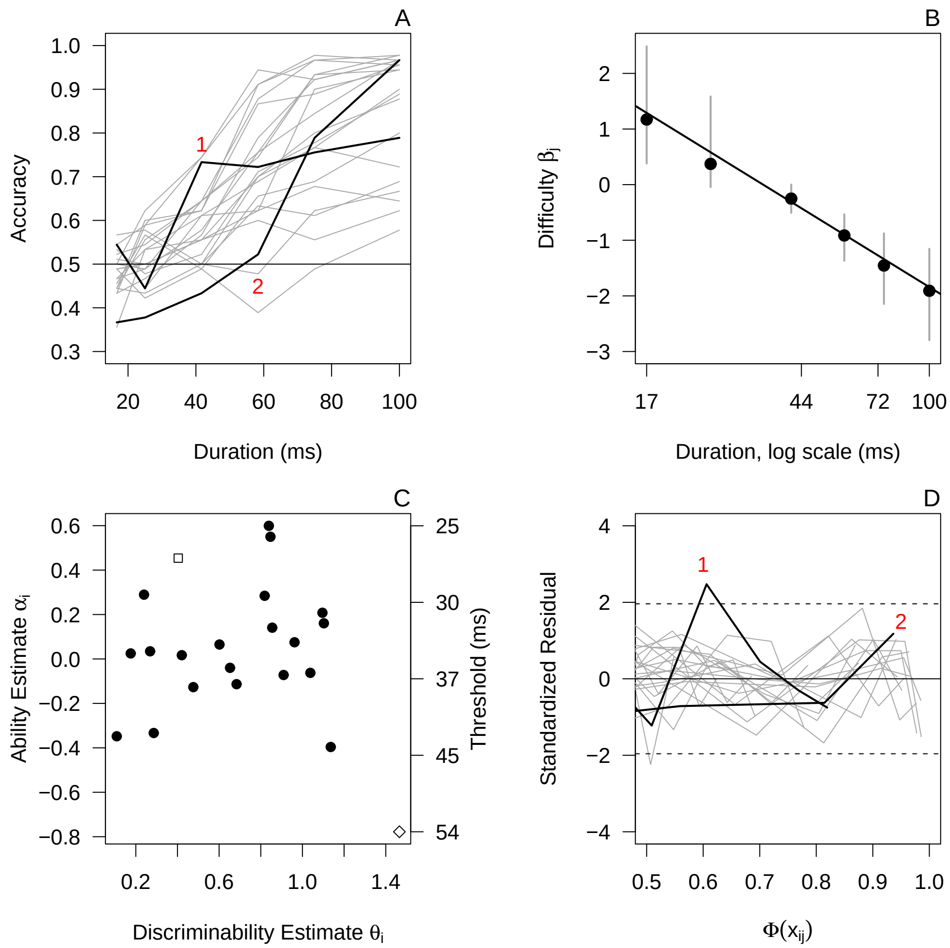

Figure 3.3: Results of the described experiment and analyses. A: Accuracy as function of stimulus duration for all 22 participants. Each line represents a participant. Participants marked 'a' and 'b' are the same participants shown in Figure 2.7B. B: 2P-MAC item difficulty (β) estimates as a function of duration. Error bars are 95% credible intervals. Fitted regression line is best linear fit with weighted least squares. C: Ability estimates (α) as a function of discriminability estimates (θ) for all 22 participants. The square and diamond are Participants 1 and 2 from Panel A, respectively. Right axis shows the threshold in ms corresponding to each ability. D: Standardized residuals from the 2P-MAC analysis. The dark lines correspond to participants whose residuals were large and patterned in the 1P-MAC analysis (see Figure 2.7B). These residuals are greatly improved.

iterations, with the first 1,000 iterations serving as burn-in. The chains exhibited low autocorrelation and good convergence when decorrelating steps are included. Figure 3.1B shows a typical chain for an item parameter (β_4).

Estimates of item difficulty parameters (β) and their 95% posterior credible intervals are shown as a function of duration in Figure 3.3B. Not only do these difficulties decrease with duration, there is a striking linear relationship between the logarithm of duration and difficulty. The line in Figure 3.3B shows the weighted least squares best fit. Figure 3.3C shows participant ability and discriminability as a scatter plot. There is no evident relationship, though this may not be so surprising with only 22 participants. The linear relationship between log-duration and item difficulty aids in the estimate of a duration threshold for each participant. The threshold estimate, denoted $\hat{d}_i$, may be obtained by solving $\hat{\alpha}_i = 6.2 - 1.7 \log \hat{d}_i$. These values range from 25 ms to 54 ms as indicated on the right-hand axis of Figure 3.3C.

As in Chapter 2, standardized residuals r_{ij} may be computed as:

$$r_{ij} = \frac{y_{ij} - N_{ij}\hat{p}_{ij}}{\sqrt{N_{ij}\hat{p}_{ij}(1 - \hat{p}_{ij})}}$$

where $\hat{p}_{ij}$ is $\Phi(\hat{x}_{ij} \vee 0)$ and $\hat{x}_{ij}$ is the posterior estimate of latent true score x_{ij}. If the model fits well, these residuals will fall between -1.96 and 1.96 for about 95% of participant-by-item combinations. Standardized residuals for the 2P-MAC fit are shown in Figure 3.3D as a function of predicted performance for above-baseline combinations. Each line represents a single participant. The residuals generally fall around zero and inside the 95% bounds, denoted by horizontal broken lines. No systematic trends are obvious. These residuals may be contrasted with those for the 1P-MAC model in Figure 2.7B. The fit of the 2P-MAC model greatly improves upon the 1P-MAC model; the highlighted participants ‘a’ and ‘b’ no longer show large, sweeping residuals. In sum, the 2P-MAC model appears to fit the data well.

The 2P-MAC model fits revealed two interesting results. First, item difficulty appears linearly related to the logarithm of duration in the task described. Second, participants' abilities do not appear to be correlated to their discriminabilities. Exploring these potential psychophysical invariances is made possible in reasonably sample sizes using the 2P-MAC model.

3.4 Summary

The assumptions underlying one-parameter model developed in Chapter 2 are too strong. In particular, the assumption of additivity in the 1P-MAC model cannot account for participants who improve at different rates in psychophysical tasks. In this Chapter, I developed a two-parameter version using standard Bayesian techniques and applied the resulting model to the perceptual task described in Chapter 2. The model appears to fit the data well, and the model performs well in simulation. In the next Chapter, I show how the 2P-MAC model may be applied to a psychophysical data set to explore the phenomenon of subliminal priming.

Chapter 4

Application: Subliminal Priming

Much of human behavior is affected by unconscious information processing. There is evidence that we may be affected by the environment in ways of which we are not aware. Bargh, Chen, and Burrows (1996) present an experiment which demonstrates that participants may be unaware of the effects that stimuli have. In the experiment, participants were asked to unscramble sentences. Some participants received sentences suggesting rude behavior; others received sentences suggesting polite or neutral behavior. When they were done unscrambling the sentences, they were told to report to an experimenter. The experimenter, however, was talking to a confederate, requiring the participant to interrupt the experimenter to continue the experiment. Of interest was whether participants would interrupt the experimenter, and how this was affected by the type of sentences the participant unscrambled. Figure 4.1 shows the proportion of participants who interrupted the experimenter as a function of which sentences they had previously unscrambled. Participants who unscrambled polite-related sentences only interrupted 18% of the time, whereas those who unscrambled rude-related sentences interrupted 64% of the time. The politeness manipulation yielded an extremely large effect on behavior. When participants were asked whether they felt the unscrambling task had any effect on them, no participants reported any effect. Bargh et al. took this to indicate that the participants were not conscious

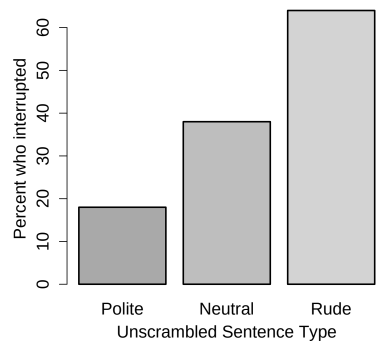

Figure 4.1: Proportion of participants who interrupted an experimenter as a function of the type of sentences unscrambled in a previous task. Figure adapted from Bargh, Chen, and Burrows (1996).

of the large effect the sentences had on behavior.

The phenomenon described by Bargh et al. may be referred to as *priming without awareness* (Cheesman & Merikle, 1984). The participants in the experiment were primed for rude or polite behavior by unscrambling rude or polite sentences, but they were unaware of the sentences' effects. Similarly, unconscious effects have been found in other domains, including memory (Process Dissociation: Jacoby, 1991), likability of abstract shapes (Kunst-Wilson & Zajonc, 1980), and learning of artificial grammar (Reber, 1967), among others. The role of unconscious or implicit processing is uncontroversial in psychology.

Not all claims of unconscious effects are equally uncontroversial, however. *Subliminal priming* refers to priming effects from stimuli presented below threshold. According to many researchers, primes may be presented so quickly as to be below threshold, yet they still influence responses to subsequent targets (Abrams, Klinger, & Greenwald, 2002; Bar & Biederman, 1998; Breitmeyer, Ogmen, & Chen, 2004; Dehaene et al., 1998; Eimer & Schlaghecken, 2002; Greenwald, Abrams, Naccache, & Dehaene, 2003; Greenwald, Draine, & Abrams, 1996; Holender & Duscherer, 2004; Jaskowski, Skalska, & Verleger, 2003; Kunde, Keisel, & Hoffman, 2003; Mattler, 2006; Merikle, Smilek, & Eastwood, 2001; Reynvoet & Ratinckx, 2004; Snodgrass, Bernat, & Shevrin, 2004; Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach, 2003). Dehaene, et al. (1998) serves as a suitable example. The structure of a trial is shown in Figure 4.2A; single digits serve as both primes and targets. In the first stage of the experiment, participants judge whether targets are greater-than or less-than five. The critical question is whether the speed of these judgments is affected by the primes. A prime is considered *congruent* with the target if both have the same less-than-five status (e.g., a prime of *6* and a target of *8*); a prime is considered incongruent if the prime and target have the opposite less-than-five status (e.g., prime of *3* and a target of *8*). Priming occurs if target categorization is faster for targets preceded by congruent primes than for those preceded by incongruent ones. In a separate stage of the experiment, participants categorize the primes themselves by their less-than-five status. Dehaene et al., among others, claim that priming my occur even when prime categorization is at chance.

The finding of subliminal priming is taken as evidence for a strong dissociation between unconscious and conscious processing (Dehaene et al., 1998). The accuracy in the prime categorization task is assumed to index conscious processing, and the priming effect in reaction time target categorization task is assumed to index unconscious processing. If subliminal priming exists, it implies that there may be unconscious processes that act on information completely unavailable to conscious processes. This stands in contrast to phenomena such as the priming without awareness effects mentioned previously. In studies like Bargh et al., information in the stimuli was available to both conscious and unconscious processes, but the processes acted on the information differently. Understanding subliminal priming is therefore vital for understanding the relationship between the conscious and the unconscious in human information processing.

The claims of subliminal priming described above, such as those of Dehaene

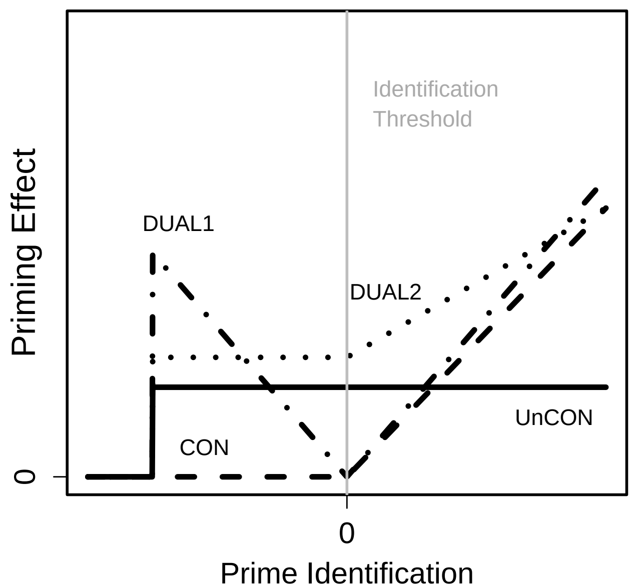

Figure 4.2: A: The structure of a trial in the number priming paradigm. The displayed prime is incongruent with the target. B: Predictions of four substantive priming theories. *UnCON*: priming reflects unconscious processes; *CON*: priming reflects conscious processes (Cheesman & Merikle, 1988); *DUAL1*: dual-process model of Snodgrass et al. (2004); *DUAL2* dual process model based on Greenwald et al. (1996). The vertical line denotes the threshold on prime identification or categorization (Figure adapted from Snodgrass et al., 2004.)

et al. have not gone unchallenged. There are a number of important critiques of subliminal priming claims. The first major criticism is that the methods discussed in Chapter 1 of measuring at-chance performance are the same ones used by subliminal priming researchers. These methods are typically based on accepting the null hypothesis, and are subject to high numbers of errors in which above baseline performance is claimed as at-baseline. If the null hypothesis is baseline performance, low power biases researchers to claim that stimuli are below threshold. Because the combination of a significant priming effect and baseline performance are considered evidence for subliminal priming, researchers will be biased toward finding evidence for subliminal priming. The critique, popularized by Reingold and Merikle (1988), is referred to as the *null sensitivity* critique.

The second major criticism of subliminal priming claims is that the two tasks used to argue for subliminal priming, prime categorization and target categorization, are not equivalent (Merikle & Reingold, 1998; Reingold & Merikle, 1988). If tasks are not equivalent, there may be information available from the prime in one task that is not available in the other. There are many ways in which the two tasks could differ that could bias researchers toward claims of subliminal priming, even when none exists. Pratte and Rouder (submitted) point out a major violation of equivalence in the subliminal priming paradigm. In subliminal priming experiments, the prime stimulus is presented very quickly and masked so that it is difficult to see. The target, on the other hand, is always easy to see. When participants are asked to classify the target, the task is easy and participants are motivated. In contrast, when participants are asked to classify the prime, the task is very hard and motivation is low. Participants may simply disengage from the prime categorization task, leading to chance performance. If participants were engaged during the target categorization task, but disengaged during the prime categorization task, then it would be expected that even if all priming is superliminal, evidence for subliminal priming would be found.

In the following section, I outline a number of proposed solutions to the null sensitivity critique. Following that, I will present two experiments which attempt to account for the two critiques described above.

4.1 Proposed solutions to the Null Sensitivity Critique

In this section I outline three proposed solutions to the null sensitivity critique. I describe the first two solutions, Greenwald, Klinger, and Schuh's (1995) regression method and Snodgrass, Bernat, and Shevrin's (2004) model-based approach, and then discuss their flaws before reviewing the third method, the Mass at Chance model.

Regression Method

Greenwald, Klinger, and Schuh (1995) suggested a method to demonstrate that a nonzero priming effect exists even when prime categorization is at chance that does not rely on accepting the null hypothesis. Greenwald et al. suggest that researchers regress the priming measure from the target categorization task onto the performance measure from the prime categorization. The intercept in this regression is interpreted as the size of the priming effect when the performance effect is at chance[1]. Figure 4.3 shows an example of the use of this logic, from Draine and Greenwald (1998). The x axis is the sensitivity of the participant in the prime categorization task. The y axis is the standardized reaction time priming effect in the target categorization task. Each point is a participant-by-duration combination. The line shows the best fit OLS regression line. The significant intercept is interpreted as evidence of a priming effect when there is no sensitivity to the prime in the prime categorization task, and therefore as

[1]In the case of proportion correct, .5 may be subtracted to make the chance baseline equal to 0.

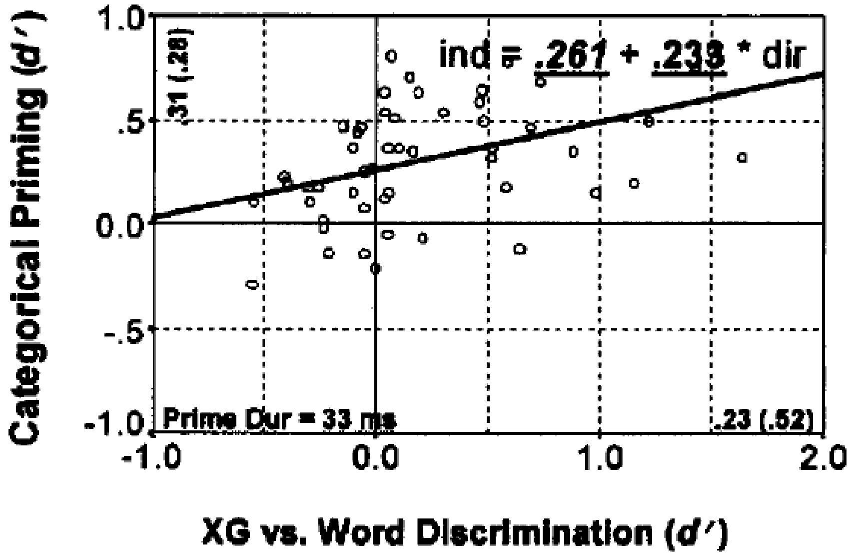

Figure 4.3: The regression method applied to data from a subliminal priming experiment, from Draine and Greenwald (1998). Priming effect in a target categorization task is shown as a function of sensitivity in a prime detection task. The significant intercept is taken as evidence for subliminal priming.

evidence of subliminal priming.

Merikle and Reingold (1998), Dosher (1998), and Miller (2000) note that there are several problems with the regression technique.

- **Errors in variables measurement problem.** In simple linear regression, the independent variable is assumed to be measured without error. If this assumption is not met, estimates of the slope and intercept will be biased. In the case of of a line with a positive slope and zero intercept, the slope will be attenuated and the intercept biased greater than zero. In subliminal priming experiments, the performance measure is typically proportion correct or signal detection d', both of which are measured with error. Dosher (1998) pointed out that this bias tends to favor finding evidence for subliminal priming by the regression method even if none truly exists.

In response to this criticism, Klauer, Draine, and Greenwald (1998) present a regression model that accounts for the measurement error in both variables. The corrected model still assumes that the relationship between the two variables is linear, but it includes two new assumptions: first, the independent variable (accuracy or d'[2]) is measured with error, and this error is the same for all levels of the independent variable; and second, the true value of the performance measure cannot be below chance. Klauer and Greenwald (1998) reanalyzed several data sets with the corrected model, most of the results were unchanged. They argue that this is further evidence for subliminal priming.

- **Linearity Assumption.** The assumption of a linear relationship between performance in the prime categorization task and the priming effect in the target categorization is a strong one. The strength of the method relies on accurate measurement of the intercept; if the true relationship is nonlinear and a linear relationship is fit, a spurious significant intercept is likely to result. Dosher (1998) and Miller (2000) used a simulation approach to show that when the relationship between priming and the performance measure is nonlinear, the regression method yields spurious significant intercepts. In the sample sizes used in typical experiments, error in both the independent and dependent variable will make nonlinear relationships difficult to detect (see, for example, Figure 1E-F in Dosher, 1998).

  In response to the linearity critique, Greenwald and Draine (1998) use two examples of nonlinear regression: polynomial regression and nonparametric regression (lowess; Cleveland, 1981). According to Greenwald and Draine,

[2]The signal detection measure d' (Green & Swets, 1966) is used in many subliminal priming studies because it corrects for bias in responding. In the priming experiments presented in this Chapter, there is no *a priori* reason to believe that there is bias in responding, and none was evident in the data. Therefore, accuracy serves as an acceptable alternative to d'. In fact, it is easily shown that in the absence of response bias, $x_{ij} = d'_{ij}/2$.

> both of these methods yielded significant intercepts. They cite this consistency across linear and nonlinear models as evidence that the subliminal priming effects found in Draine and Greenwald (1998) are not methodological artifacts.

One major problem remains with the regression method, however. Although Greenwald and colleagues have addressed both the errors-in-variables problem and the linearity problem, they have not done so jointly. In no analysis to date have they combined the two solution, and so any analysis suffers from at least one of these serious criticisms.

The reason for this is likely that it is simply too difficult. Nonlinear nonparametric errors in variables regression is an exceedingly difficult problem that has only recently begun to be solved (Carroll, Maca, & Ruppert, 1999). Convergence of estimates to the true relationship is extremely slow as sample size increases (Fan & Truong, 1993) and the techniques are difficult to implement. All of these techniques also assume smoothness in the relationship, which may or may not be warranted. In sum, the regression technique appears to have insoluble problems which render it unuseful in studying subliminal priming.

Snodgrass, Bernat, and Shevrin's model-based approach

Snodgrass, Bernat, and Shevrin suggested that instead of focusing on the priming effect at the point of null sensitivity, priming over a range of performance levels may be assessed and used to test models of priming. Figure 4.2B, adapted from Snodgrass et al., shows four different theoretical relationships between prime categorization performance and priming effects. The horizontal solid line labeled *UnCON* represents a model in which priming is due solely to automatic, unconscious effects. The priming effect is constant over a range of prime durations because the unconscious process acts when the prime is both unconscious and when it is conscious. The dashed line labeled *CON* represents a model in which priming is due solely to conscious processing. There is no priming if the primes are below the threshold on categorization performance (e.g. Cheesman & Merikle, 1984). The dotted, v-shaped line labeled *DUAL1* is from Snodgrass et al.'s (2004) dual-process theory. In this model, both unconscious and conscious processes influence priming. Conscious processes, when they are active, override unconscious processes and lead to a smaller degree of priming near threshold. The mixture of these two processes leads to a nonmonotonic relationship between priming and performance. The dotted line labeled *DUAL2* shows a schematic of the theory underlying Greenwald's regression method (Draine & Greenwald, 1998; Greenwald, et al., 1996). Unconscious priming holds below threshold whereas the sum of conscious and unconscious priming holds above it. By looking for broad trends in the relationship between an prime categorization performance and priming effects in target categorization, the null sensitivity critique is sidestepped; if, for instance, a nonmonotonic trend is found (DUAL1), then this strongly implicates two processes without needing to show that performance is at baseline for any particular participant. Figure 4.4 shows a nonmonotonic trend used by Snodgrass et al. to argue for dual processes in a subliminal priming paradigm.

Although the idea has intuitive appeal, there are several problems with Snodgrass et al's approach. First, as evidence for a dual process model, Snodgrass et al. relied on the interpretation of negative d' estimates as estimates of sensitivity. Negative d' measurements, similar to accuracies below the chance baseline, are assumed to only occur as a result of sampling error. They do not represent accurate measurements of below-threshold stimulation, and thus cannot be interpreted in favor of a dual process model. Without a way of measuring how far below threshold a participant-by-item combination is, it is difficult to see how the models in Figure 4.2B could be differentiated. The second problem is that if a monotonic trend is found, the model supported depends critically on the location of the threshold. Because Snodgrass et al. do not suggest any method

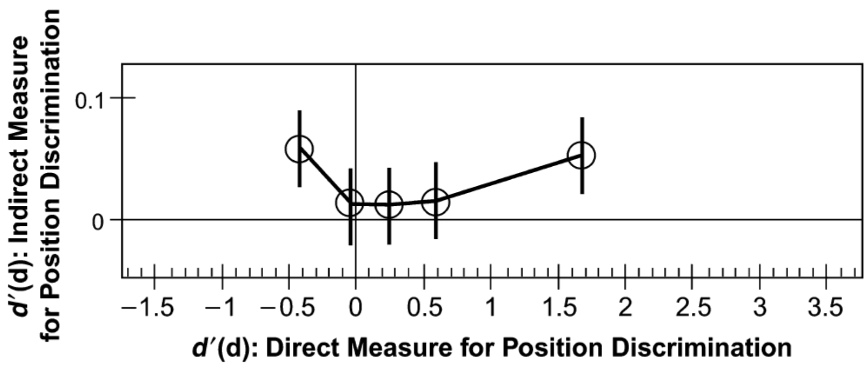

Figure 4.4: The nonparametric regression of priming effect in a target categorization task onto sensitivity in a prime categorization task. Vertical lines are standard errors of the mean. The nonmonotonicity is taken as evidence for a dual process model of Snodgrass et al. Figure adapted from Snodgrass, Bernat, and Shevrin, (2004); data from Greenwald, Klinger, and Schuh (1995).

for measuring thresholds, the model based approach is unlikely to be fruitful in studying subliminal priming without a method of measuring thresholds.

The Mass at Chance Model

The primary question underlying subliminal priming research is whether it is possible for stimuli that are evoke chance performance in a prime categorization task to induce a priming effect. In Chapters 2 and 3, I presented the Mass at Chance (MAC) model, which can be used to classify participant-by-duration conditions as at chance. For all combinations classified as at chance, the priming effect may be examined for evidence of subliminal priming.

The MAC model approach has several distinct advantages over the regression method. First, although the MAC model is still affected by errors in variables through errors in the classification as at-chance or above-chance, the error rates may be controlled by manipulating the chance-classification criterion on ω_{ij}, discussed in Chapters 2 and 3. The MAC model therefore does not suffer from the errors in variables problem in the same manner as the regression method. The effect of errors in variables on MAC is discussed further in the General Discussion. The second advantage the MAC model has over the regression method is that the MAC model avoids the nonlinearity critique, because no relationship is assumed between the prime categorization measure and the priming measure from target categorization. The MAC model can dichotomize performance on the prime categorization task into below-threshold and above-threshold performance, making the assessment of priming effects into a simple t test.

The MAC model is not limited to dichotomization, however. The MAC model yields estimates of prime categorization task ability, x_{ij}, even for participant-by-duration conditions that are below threshold. Two at-chance participants may differ in how far below threshold they are to a given prime duration. The models of priming shown in Figure 4.2B have different predictions for how the priming effect changes as a function of ability below the threshold. Both the regression method and Snodgrass et al.'s model-based approach are limited in that they cannot differentiate far-below-threshold participants from just-below-threshold performance, whereas MAC can. A related advantage is MAC's ability to accurately estimate thresholds. Because thresholds can be localized naturally by the MAC model, the monotonic models in Figure 4.2B can be differentiated from one another by where the threshold lies with respect to the priming effects.

A further advantage of the MAC model over the regression approach is the MAC model's hierarchical nature. In many subliminal priming experiments, participants judge primes at multiple prime durations. As discussed in Chapter 2 and 3, the MAC model uses Item Response Theory techniques to estimate an ability and rate-of-improvement parameter for each participant and a difficulty parameter for each prime duration. Estimation in a hierarchical framework is more efficient than if the data from each participant-by-duration combination is considered separately, as in the regression method.

4.2 Experiment 1

Dehaene et al. (1998) report several experiments in which they find a significant subliminal number priming effect. As noted in Chapter 1, Dehaene et al. used the t test method to argue that participants were below threshold in the prime categorization task. The t test method, however, is prone to high error rates. These high error rates bias researchers in favor of claiming that primes are below threshold. Experiment 1 is a number priming experiment similar to Dehaene et al.'s number priming experiments, and is designed to enable the testing of subliminal number priming using the MAC model.

Experiment 1 is composed of two tasks. In the target categorization task, participants judge a single-digit target to be less-than or-greater than 5. This target may be preceded by a prime, which can be on the same side of 5 as the target (a *congruent* prime) or on the opposite side of 5 (an *incongruent* trial). The priming effect typically observed is that congruent primes lead to faster responses to the target on average.

In the prime categorization task, participants judge the prime as greater-than or less-than 5. The dependent variable of interest in the prime categorization task is accuracy. Subliminal priming is said to occur at a particular prime duration if true accuracy of a participant in the prime categorization task is at chance, yet there is an effect of the prime at the same prime duration in the target categorization task.

Experiment 1 is designed to mitigate both the null sensitivity critique of Reingold and Merikle (1988) and the motivation critique of Pratte and Rouder (submitted). In order to mitigate the null sensitivity critique, the MAC model will be used to determine whether participants are below threshold in the prime categorization task. In order to mitigate the motivation critique, participants will be presented with a large proportion of easy-to-see, long-duration prime stimuli in the prime categorization task. Pratte and Rouder showed that mixing easy primes with harder ones raises performance to the harder primes, presumably by making the task less frustrating to the participant and increasing motivation.

4.2.1 Method

Participants

Twenty-three undergraduate students from the University of Missouri-Columbia introductory psychology pool served as participants. They were given class credit in exchange for their participation.

Stimuli and Apparatus

Primes and targets were the digits 2,3,4,6,7, and 8, and masks were the number sign ("#"). All characters were white on a black background. Stimuli were presented on a 17" CRT display at a resolution of 640x480 running at 120 Hz, controlled by PC-compatible computers.

Design

The experiment consisted of two tasks: first, target categorization, and second, prime categorization. The independent variables in the target categorization task were prime-target congruency (congruent/incongruent) and prime duration. All combinations of prime and target were presented in a within-list design with the exception that primes and targets were always different digits. Each combination occurred equally often in a random order. For each participant, prime duration was manipulated through 3 levels, selected from the durations 8.3ms, 16.7ms, 25ms, 33.3ms, 41.7ms, and 50ms. In order to insure that the average difficulty was roughly the same for each participant, one duration was selected from the fastest two durations, one from the middle two, and one from the slowest two. The dependent variable in the prime categorization task was response time.

In the prime categorization task, the independent variable was prime duration. Prime duration was manipulated through 4 levels: the 3 used for that participant in the target categorization task, and 100ms. The dependent variable in the prime categorization task was proportion correct.

Procedure

The structure of a trial is shown in Figure 4.2. On each trial, a fixation was shown for 500ms, followed by the mask character for 58ms and then the prime. After the prime, the mask character was shown for another 58ms, followed by the target for 4s. Following the target, the screen was blank until a response was made. Participants were allowed to respond any time after the target began.

Participants were first given instructions for the target categorization task. They were instructed to ignore everything on a trial except the target, and to press 'z' if the target was less than 5 and '/' if the target was greater than 5. Each participant performed 180 trials in each of three selected prime duration conditions, yielding a total of 540 trials. These trials were divided into blocks of 90 trials each. Feedback was provided.

Following the target categorization task, participants were given instructions for the prime categorization task. They were told to ignore everything except the prime, and to use the same responses to categorize the primes as they used to categorize the targets. Each participant performed 80 trials in each of the three prime duration used in the target categorization task, and 160 trials in the 100ms prime condition. Because 40% of the trials were relatively easy at 100ms, participants are assumed to have higher motivation than if we had not included the easy prime trials (Pratte and Rouder, submitted). These 400 trials were divided into four blocks of 100 trials. Feedback was not provided.

4.2.2 Results

Priming effects in the target categorization task were computed by subtracting the mean congruent reaction time in a condition from the mean incongruent reaction time in a condition. Figure 4.5A shows the mean priming effects in each of the to each of the prime durations for the target categorization task. The priming effects are small or nonexistent in the short duration conditions, and increase for longer durations, indicating that incongruent combinations slowed down responding as expected. Figure 4.5B shows the mean accuracies in each of the prime duration condition in the prime categorization condition. Accuracies are low to the briefest primes and increase as duration is increased.

The 2P-MAC model discussed in Chapter 3 was fit to the prime categorization data. For each participant-by-prime-duration combination, two quantities are of interest: first, the posterior probability that the performance in that combination is at baseline, or $\omega_{ij} =$Pr$(x_{ij} < 0 \mid \text{data})$; and second, the estimated latent ability x_{ij} for each combination. Figure 4.5C shows w_{ij} for each combination. The probabilities order well, with the probabilities decreasing as the primes are of longer duration. It is important to choose a decision criterion on ω_{ij}, in order to categorize combinations as below threshold or above threshold. As discussed in Chapter 2, a suitably conservative criterion is .95; however, there are few combinations above .95, limiting the usefulness of the very conservative criterion. For this analysis, the criterion may be relaxed to .5, the most liberal criterion suitable. A .5 criterion on ω_{ij}, may be interpreted as being 50% certain or more that a combination is below threshold. If there is no evidence of subliminal priming with the liberal .5 criterion, more conservative criteria will not be needed.

Figure 4.6A shows the mean priming effect as a function of chance classification, with 95% confidence intervals. The mean priming effect for participants-by-prime-duration combinations categorized as at-chance, the mean priming effect is small or nonexistent. This contrasts with the priming effect for participants

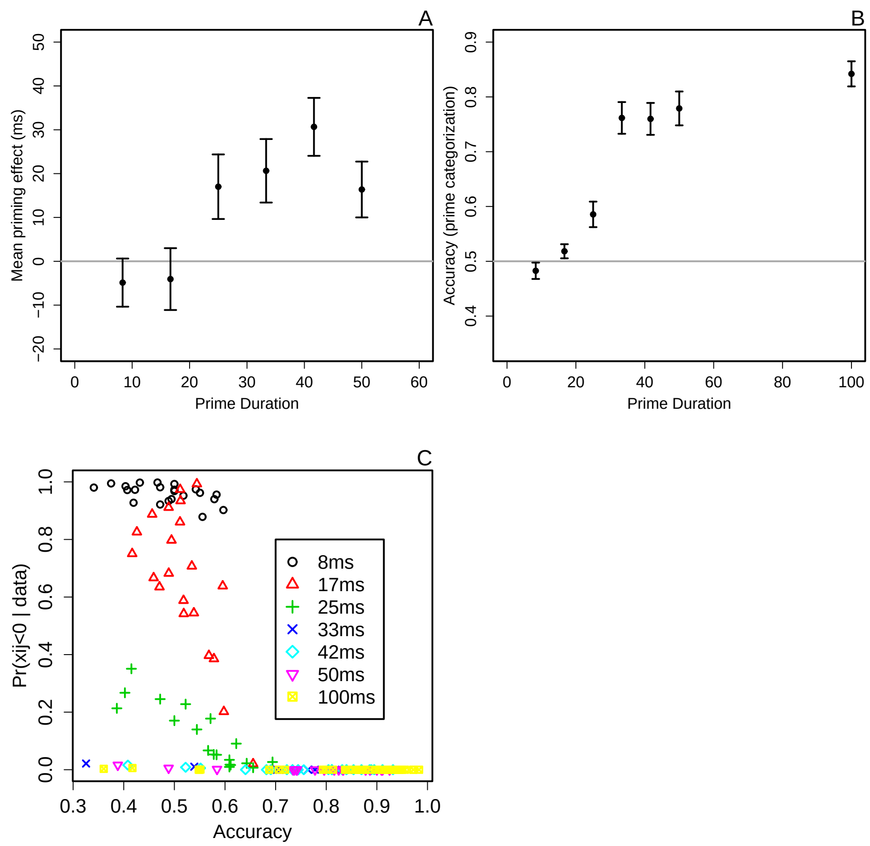

Figure 4.5: Results of Experiment 1. A: Mean priming effects in each prime duration condition in the target categorization task. B: Mean accuracies to categorize the prime in the prime categorization task. C: Posterior probability of true baseline performance for each participant-by-duration combination. Error bars in B and C are standard errors of the mean.

categorized as above chance, whose mean priming effect is substantially above 0. For the priming effects for at-chance combinations (left bar), the Bayes Factor comparing the null model (mean is 0) with the alternative model (mean is not 0) is shown. This number may be interpreted as the odds of the null model against the nonnull model, given data. For details, see Rouder, Speckman, Sun, Morey, and Iverson (submitted). Although the value of 5.8 is not exceedingly large, it must be considered in the context of the many studies which have claimed replicable subliminal priming. In this context, odds of 5.8 to 1 in favor of the null are respectable and surprising.

The relationship between latent ability x_{ij} and the mean priming effect can be seen in Figure 4.6B. For latent abilities below $x_{ij} = 0$, no systematic deviation from a null priming effect is evident. However, above $x_{ij} = 0$, most combinations have a priming effect above 0. Also notable is that there is no evidence of any nonmonotonicity, and thus no evidence for Snodgrass et al.'s dual process model.

Draine and Greenwald (1998) suggested that subliminal priming effects have a brief time course. It is possible that by using the mean reaction time, subliminal priming effects in the earliest responses are obscured. In order to test this, instead of computing the difference between mean incongruent and congruent reaction times, I computed the difference in the 10^{th} percentiles for incongruent and congruent trials. This may serve as an alternate measure of priming for the fastest 10% of responses. Figure 4.6C and D show the analyses for the fast responses. The patterns revealed by the mean reaction times remain clear: there does not appear to be any evidence for subliminal priming.

In using the regression method, it is common for researchers to standardize effects in reaction time before analyzing them (Greenwald, Klinger, and Schuh, 1995). One way to standardize reaction time is to find the difference in mean RT, and then divide by a pooled standard deviation estimate of the RT distributions. This has the advantage of putting all participants' priming effects on a common

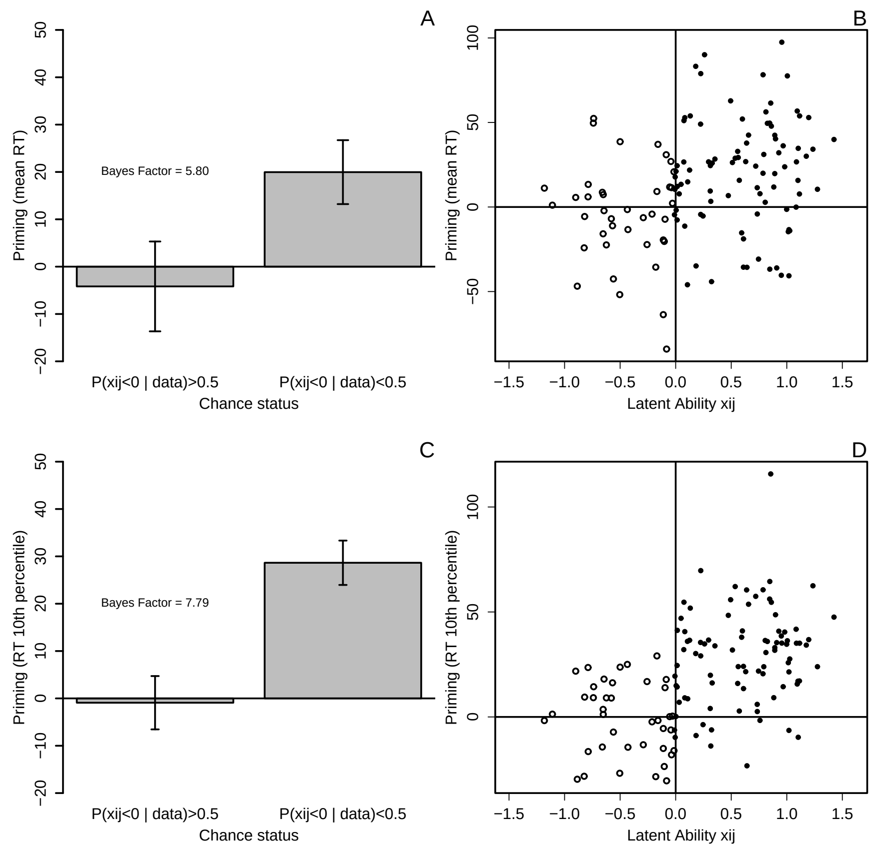

Figure 4.6: MAC model analysis of Experiment 1. A: Bar plot showing the mean priming effect as a function of chance categorization status. See text for an explanation of the Bayes Factor. B: Priming effect in mean RT as a function of latent ability x_{ij}. C and D: Same as A and B, but for the 10^{th} RT percentile instead of the mean RT. Error bars in A and C are 95% confidence intervals on the mean. Open circles in B and D represent combinations categorized as below threshold.

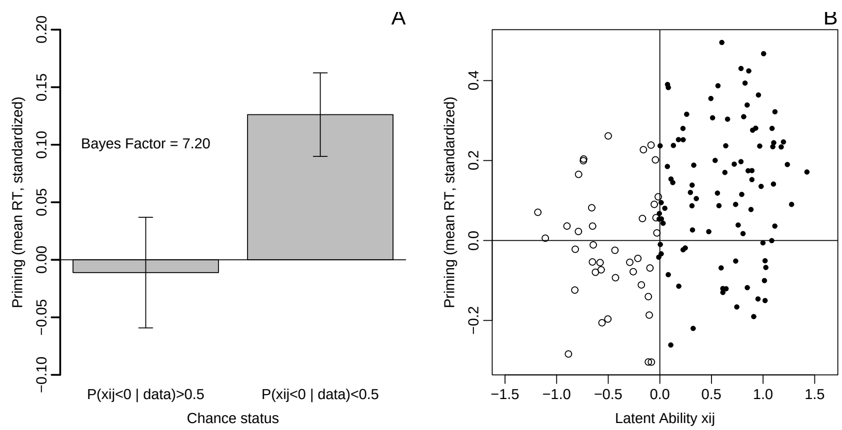

Figure 4.7: MAC model analysis of Experiment 1, with standardized mean RT. A: Bar plot showing the mean standardized priming effect as a function of chance categorization status. See text for an explanation of the Bayes Factor. Error bars are 95% confidence intervals on the mean. B: Priming effect in standardized mean RT as a function of latent ability x_{ij}.

scale, with the drawback that the absolute magnitudes of effects are no longer clear. Figure 4.7A and B show bar and scatter plots using the standardized priming effect measures. As was true of the other priming measures, they provide no evidence of subliminal priming.

4.2.3 Discussion

The results of Experiment 1 are surprisingly clear; the experiment yielded no evidence of subliminal priming. The priming effects above threshold were robust, but no sign of a priming effect was found below threshold. If we adopt Snodgrass et al.'s suggestion of using a model-based approach to testing subliminal priming, it seems clear that the model supported by the data is a single, conscious process model (CON in Figure 4.2B). Figure 4.6B shows no nonmonotonicity and no priming effect below threshold, ruling out all models except the single process theory. This pattern holds for all priming effect measures examined in Experiment 1.

4.3 Experiment 2

Number priming is only one of a number of paradigms in which subliminal priming has been claimed. A second common paradigm is word-categorization. Draine and Greenwald (1998), for instance, explored subliminal priming using pleasant (eg, *love*) or unpleasant (eg,*vomit*) words. They found when the prime is congruent with the target, reaction times are faster than when the prime and target are incongruent. Using the regression method they argued that the effect held subliminally, as well as superliminally.

Their results, however, are suspect for two reasons. First, as already discussed, the regression method has several distinct drawbacks. Second, early piloting of this experiment revealed that at the durations Draine and Greenwald used, the prime categorization task is exceedingly difficult. Their results could be explained via the motivation critique: the participants were paying attention to the display during the easier target classification task, but gave up during the prime categorization task. If most stimuli were superliminal, this would create the illusion of subliminal priming.

Experiment 2 is designed to explore subliminal priming in the word-categorization task while mitigating both criticisms. As in Experiment 1, the MAC model will be used to avoid the null sensitivity critique. Because a pilot study suggested that the prime categorization task in this experiment is harder than the number prime categorization, additional steps must be taken to address the motivation critique. In Experiment 1, easy, long-duration primes were included in the prime categorization task to increase motivation. In Experiment 2 long-duration primes were also included. In addition to the easy primes, participants were given an enforced break in the middle of the prime categorization task. The breaks were added to minimize the effects of fatigue, which would tend to decrease motivation.

4.3.1 Method

Participants

Forty-three undergraduate students from the University of Missouri-Columbia introductory psychology pool served as participants. They were given class credit in exchange for their participation.

Stimuli and Apparatus

Primes and targets were 50 words taken from Draine and Greenwald (1998). These words are listed in the Appendix A, Table A. Masks were random strings of 13 capital letters. All characters were capital and white on a black background. Stimuli were presented on a 17" CRT display at a resolution of 640x480 running at 120 Hz, controlled by PC-compatible computers.

Design

The design was similar to that of Experiment 1, with two exceptions: First, all participants were exposed to all durations. Prime durations in the target categorization task were 24.0ms, 33.3ms, and 66.7ms. For the prime categorization task, the prime duration of 150ms was also included to increase motivation. Second, not all prime-target combinations were presented. Within a block of 50 trials, prime-target combinations were selected such that each word occurred as the prime and the target exactly once, and no word could be both prime and target in the same trial.

Procedure

The procedure was similar to that of Experiment 1. Each trial began with a fixation cross which was displayed for 500ms, followed by the premask, displayed for 150ms. The prime was then displayed for the specified time, followed by the mask, which was displayed for 16.7ms. In order to be consistent with Draine and

Greenwald's procedure, a blank screen was inserted between the premask and target with a length that varied. The length of the blank was chosen to keep the time between the prime onset and target onset constant at 66.7ms. In the case of 66.7ms and 150ms primes, no blank was inserted. Participants were instructed to press 'z' for pleasant words and '/' for pleasant words.

Participants performed 9 blocks of 60 trials each in target categorization. Feedback was provided. Each target categorization block consisted of 20 trials in each prime duration. Participants then performed 8 prime categorization blocks of 60 trials. Each prime categorization block consisted of 12 24ms primes, 24 33.3ms primes, 12 66.7ms primes, and 12 150ms primes. Feedback was not provided in the prime categorization task.

After the fourth prime categorization block, the computer instructed the participant to report to the experimenter, and did not proceed with the experiment. The participants were encouraged by the experimenter to take several minutes to rest. When the participants signaled that they were ready to continue, the computer proceeded with the prime categorization task.

4.3.2 Results

The priming effects in mean reaction time for the target categorization task are shown in Figure 4.8A. For prime exposures of 24.0ms and 33.3ms, the priming effect is small or nonexistent. However, a robust priming effect of almost 30ms was found for 66.7ms primes. The accuracies in the prime categorization task for each prime duration are shown in Figure 4.8B. Mean accuracies are close to the chance baseline for the 24.0ms prime duration and increase as prime duration is increased.

MAC model estimates of ω_{ij} =Pr($x_{ij} < 0 \mid$ data) are shown in Figure 4.8C. The model gives more participant-by-prime-duration combinations higher probabilities of at chance performance than in Experiment 1. Because more participants achieve high probabilities of at-chance performance, a more conservative criterion on ω_{ij} may be used to categorize participant-by-prime-durations as at chance than in Experiment 1, where .5 was used. In the case of Experiment 2, a criterion of .9 gives a suitable balance between sample size and conservatism for assessing subliminal priming effects.

Figure 4.9A shows a bar plot of priming effects in mean RT as a function of MAC model chance categorization. As in Experiment 1, no evidence of a priming effect for participants was found. The same holds true for the analysis of the 10^{th} reaction time percentile and the standardized reaction times.

What was not seen, however, in Experiment 1 but is seen in Experiment 2 is an apparent relationship between the latent ability and priming effect for the latent abilities above 0. Because the priming effects are highly variable, it is difficult to say whether this is a genuine trend or not. If this is a genuine trend, the form of the relationship may yield insights into how priming effects arise from very weak stimuli.

4.3.3 Discussion

The results of Experiment 2 are consistent with those of Experiment 1: no evidence of a priming effect was found for participants categorized as at chance. The results are largely consistent with the single, conscious process model described in Figure 4.2B. Surprisingly, the data appear to exhibit a trend for latent abilities above 0. As latent ability increases, the priming effect increases. This is surprising both because Draine and Greenwald did not find the same trend, and because it was not found in Experiment 1. Draine and Greenwald did not have the statistical method to efficiently estimate latent ability, which could explain why they did not see the same trend. It is also possible that the methodological differences between Experiments 1 and 2, such as the constant time between prime and target onset, may explain why the trend was not seen in Experiment 2.

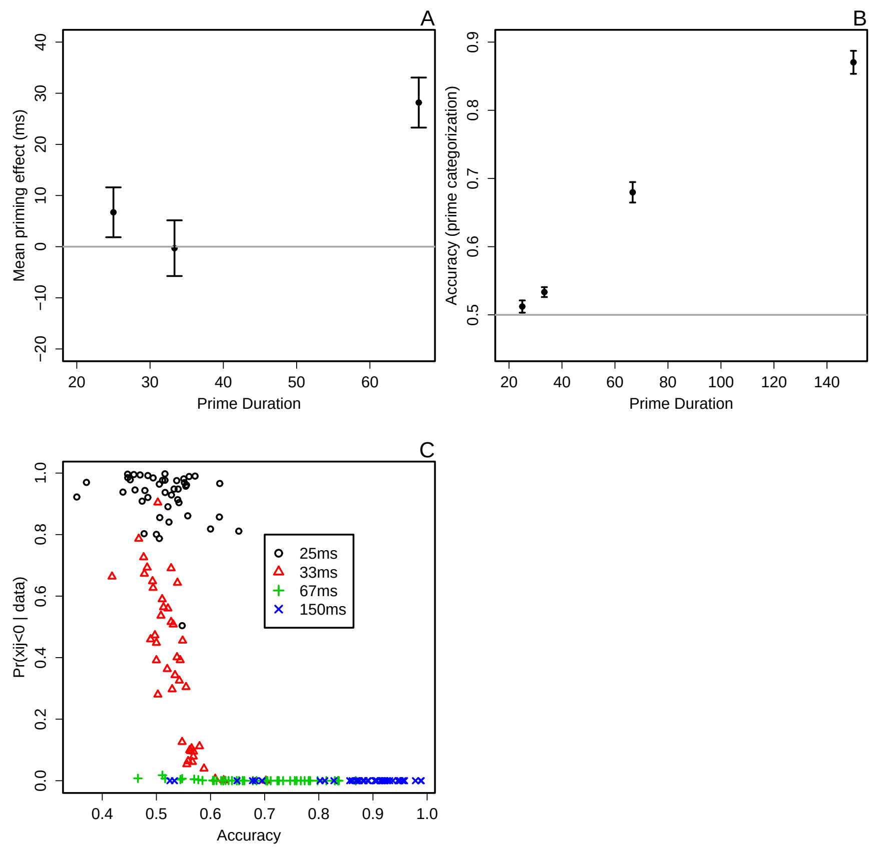

Figure 4.8: Results of Experiment 2. A: Mean priming effects in each prime duration condition in the target categorization task. B: Mean accuracies to categorize the prime in the prime categorization task. C: Posterior probability of true baseline performance for each participant-by-duration combination. Error bars in B and C are standard errors of the mean.

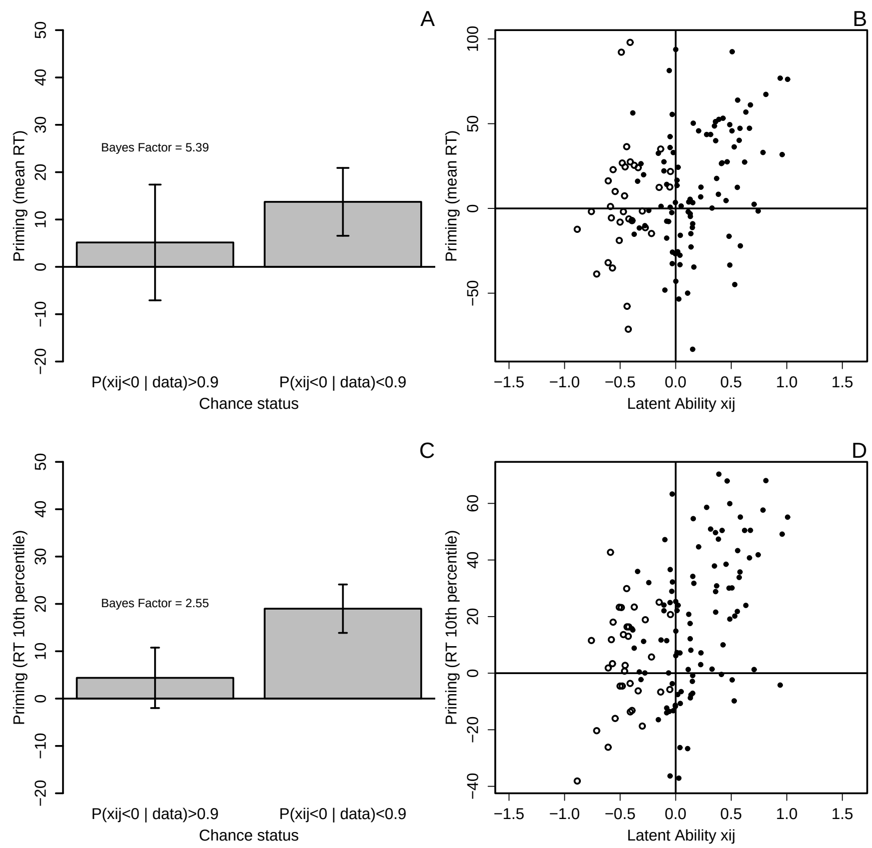

Figure 4.9: MAC model analysis of Experiment 2. A: Bar plot showing the mean priming effect as a function of chance categorization status. See text for an explanation of the Bayes Factor. B: Priming effect in mean RT as a function of latent ability x_{ij}. C and D: Same as A and B, but for the 10^{th} RT percentile instead of the mean RT. Error bars in A and C are 95% confidence intervals on the mean.

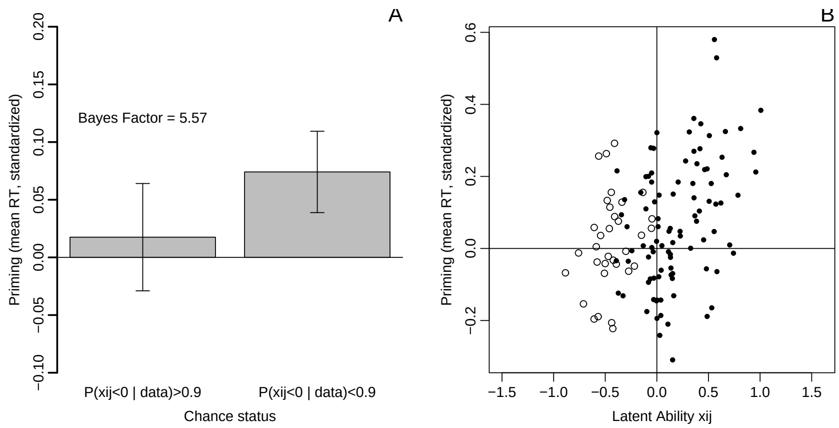

Figure 4.10: MAC model analysis of Experiment 2, with standardized mean RT. A: Bar plot showing the mean standardized priming effect as a function of chance categorization status. See text for an explanation of the Bayes Factor. Error bars are 95% confidence intervals on the mean. B: Priming effect in standardized mean RT as a function of latent ability x_{ij}.

The trend is worth exploring in future experiments.

4.4 General Discussion

In this Chapter, I presented two experiments designed to test for subliminal priming. The experiments and data analysis were designed to mitigate two major critiques of the subliminal priming paradigm: the null sensitivity critique of Reingold and Merikle and the motivation critique of Pratte and Rouder. In two experiments, no support for subliminal priming could be found. Given that previous findings of subliminal priming were based on flawed methodology, it is reasonable to believe that the phenomenon of subliminal priming is simply an artifact of bad methodology. The results of the two experiments suggest that if subliminal priming exists, the magnitudes of the priming effects are very small.

It is possible that subliminal priming exists in other domains outside of number priming and affective word priming. Although they were not tested in the reported experiments, emotional faces, affectively negative scenes, and perceptually simple stimuli such as arrows may be effective subliminal primes. The Mass at Chance model provides an principled method for determining whether stimuli are below threshold. Researchers interested in exploring subliminal priming further are encouraged to use the Mass at Chance model to avoid the methodological problems which have plagued the field of subliminal priming.

As a final note, I discuss how the conclusions drawn by researchers using the MAC model might be affected by errors in variables. In Greenwald and colleague's regression method, modeling the error in estimates of discriminability in the prime classification task is critical. If this variability is not modeled, estimates of the slope of the relationship between the priming effect and discrimination ability will the attenuated, and the intercept will also be biased. This bias in the intercept threatens inferences about subliminal priming from the regression method.

The MAC model also suffers from errors in variables: classification as at-chance or above-chance is error prone. This has an effect analogous to attenuation of the slope; the difference between the estimates of the mean priming effect of the at-chance group and the above-chance group will be attenuated. If there is no priming in the at-chance group, contamination from the above-chance group will increase the estimated priming effect, and vice versa. On advantage of the MAC model is that by the choice of criterion on ω_{ij} for classification at chance, it is possible to manipulate the effect of errors in classification. Assuming that the mean priming effect of the at-chance group is lower than the above-chance group, a liberal criterion will increase the number of above-chance combinations categorized as at chance. This will have the effect of increasing the estimate of the mean priming effect in the at-chance group, and will decrease bias estimating the mean priming effect in the above-chance group. Similarly, choosing a conservative criterion will decrease the number of above-chance classifications categorized as at-chance. This has the effect of decreasing the bias in estimating the mean priming effect of combinations at-chance. Because subliminal priming researchers are typically interested in computing the mean priming effect for at-chance combinations, a conservative criterion is warranted to guard against upward bias in estimating subliminal priming effects. At any rate, as discussed in Chapters 2 and 3, classification error rates are low overall. The ability to control the error rates through manipulating the criterion on ω_{ij} is an positive feature of the MAC model that the regression method does not share.

Chapter 5

Summary and Conclusions

Thresholds, defined as stimulus intensities where performance is at chance baseline, are difficult to measure. The formal measurement of thresholds has not been a focus of psychophysics. Instead, thresholds have been defined as stimulus intensities required to achieve a particular level of performance, such as .75 or .707. When researchers in the field of subliminal priming required the measurement of chance thresholds, no methodological tools were available. In Chapter 1 I outlined several of the methods that researchers have used to argue that performance is at chance baseline. These methods are all inadequate to the task at hand due to low efficiency, high error rates, and a reliance on acceptance of the null hypothesis.

Item Response Theory has provided psychometricians with a variety of modeling tools for measuring human abilities. The models are well-established and there are a variety of methods available for fitting the models. These models provide efficient estimation of ability, even when there is very little data to work with. However, in Chapter 1 I show that these methods are inappropriate to measuring performance at chance baselines. In fact, the models disallow chance performance by design. What is needed is an adaptation of Item Response Theory to allow performance at a true chance baseline. The MAC model, with its half-probit link function, is such an adaptation.

In Chapters 2 and 3, two Item Response models using the MAC half-probit link were developed that allow for the efficient measurement of performance with thresholds. The one-parameter model in Chapter 2 assumes that item difficulty and participant ability are simply additive. This assumption appears to be too strong in some contexts, and so the the two-parameter generalization in Chapter 3 was developed. The two-parameter model allows participants to improve at different rates as difficulty is increased. These two models will allow researchers to measure thresholds in an efficient manner, without the high sample sizes needed and high error rates inherent in the methods described in Chapter 1.

Subliminal priming research has relied on flawed methods for many decades. Defined narrowly as an effect of a unidentifiable prime on identification of a target, subliminal priming depends critically on showing that primes are not identifiable; eg, that performance is at chance. However, until recently there were no acceptable methods of assessing chance performance. Although there have been many claims of subliminal priming in the psychological literature all are suspect because all have rested on flawed methodology.

In order to test subliminal priming in a more rigorous manner, I apply the IRT MAC model developed in Chapter 3 to the data from two subliminal priming experiments in Chapter 4. By using the MAC model to classify participant-by-prime-duration combinations as either at-chance or above-chance in the prime categorization task, it is possible to give subliminal priming the most rigorous test to date. The analysis of the data in the two experiments is clear: for participants performing at chance, there is little or no priming effect. For those above chance, the priming effect is robust. I conclude, therefore, that there is no evidence for subliminal priming in the paradigms studied once certain methodological problems are accounted for.

To conclude, I speculate on the applicability of threshold links to other domains. Although I have only applied to model to a psychophysical task, true chance performance may be of interest in other domains. Consider the following two examples. The first is the applicability to content domains dominated by factual knowledge. A given person may have no knowledge whatsoever of the capital of China's Sichuan province. In these cases, the concept of a threshold is quite convenient—one may simply model this item's difficulty as greater than the participants geographic ability. The resulting predicted performance is at the appropriate baseline. A second domain in which thresholds may be applicable is describing factors that adversely affect performance. Consider, for example, performance on cognitive tasks after a variable degree of sleep. One can define baseline performance as that following a normal 8-hour night sleep. It is certainly true that many tasks exhibit declines after a very abbreviated nights sleep such as a 2-hour night's sleep. The question is whether there is some levels that are equivalent to baseline, for example a 7-hours night sleep. That is, there may be a threshold on the length of sleep needed for baseline performance. If this threshold is not met, then performance declines. In sum, while the MAC models are motivated by psychophysics, they may be broadly applicable outside psychophysics.

Appendix A

Appendix A: Words used in Experiment 2

The following table lists all the words used in Experiment 2. These words were obtained from Draine and Greenwald (1998).

Unpleasant	Pleasant
EVIL	HONOR
CANCER	LUCKY
SICKNESS	DIAMOND
DISASTER	LOYAL
POVERTY	FREEDOM
VOMIT	RAINBOW
BOMB	LOVE
ROTTEN	HONEST
ABUSE	PEACE
MURDER	HEAVEN
ASSAULT	PLEASURE
SLIME	FAMILY
DIVORCE	DIPLOMA
POISON	KISS
KILL	CHEER
DEATH	HEALTH
HATRED	FRIEND
SCUM	CARESS
ACCIDENT	SUNSET
JAIL	HAPPY
STINK	MIRACLE
TORTURE	SUNRISE
CRASH	PARADISE
FILTH	VACATION
POLLUTE	TREASURE
CURSE	TRUST
DESPAIR	LAUGH
UGLY	CALM
MISERY	BLISS
STUPID	EAGER

Table A.1: List of stimuli used in Experiment 2.

Bibliography

[Abramowitz and Stegun, 1965] Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover, New York.

[Abrams et al., 2002] Abrams, R., Klinger, M., and Greenwald, A. (2002). Subliminal words activate semantic categories (not automated motor responses). *Psychonomic Bulletin and Review*, 9:100–106.

[Ahrens and Dieter, 1974] Ahrens, J. and Dieter, U. (1974). Computer methods for sampling from gamma, beta, poisson and binomial distributions. *Computing*, 12:223–246.

[Albert and Chib, 1993] Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88:669–679.

[Andersen, 1973] Andersen, E. B. (1973). Conditional inference for multiple-choice questionnaires. *British Journal of of Mathematical and Statistical Psychology*, 26:31–44.

[Bar and Biederman, 1998] Bar, M. and Biederman, I. (1998). Subliminal visual priming. *Psychological Science*, 9:464–469.

[Bargh et al., 1996] Bargh, J., Chen, M., and Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71:230–244.

[Blackwell, 1953] Blackwell, H. R. (1953). Psychological thresholds: Experimental studies of methods of measurement. *Bulletin of the Engineering Research Institute of the University of Michigan*, No. 36.

[Breitmeyer et al., 2004] Breitmeyer, B., Ogmen, H., and Chen, J. (2004). Unconscious priming by color and form: Different processes and levels. *Consciousness and Cognition*, 13:138–157.

[Busey and Loftus, 1994] Busey, T. A. and Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review*, 101:446–469.

[Carrol et al., 1999] Carrol, R. J., Maca, J. D., and Ruppert, D. (1999). Nonparametric regression in the presences of measurement error. *Biometrika*, 86:541.

[Cheesman and Merikle, 1984] Cheesman, J. and Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, 36:387–395.

[Cleveland, 1981] Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, 35:54.

[Dagenbach et al., 1989] Dagenbach, D., Carr, T., and Wilhelmsen, A. (1989). Task-induced strategies and near-threshold priming: Conscious influences on unconscious perception. *Journal of Memory and Language*, 28:412–443.

[Dehaene et al., 1998] Dehaene, S., Naccache, L., Le Clech, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P.-F., and Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, 395:597–600.

[Devroye, 1986] Devroye, L. (1986). *Non-uniform random variate generation*. Springer-Verlag, New York.

[Dosher, 1998] Dosher, B. (1998). The response-window regression method - Some problematic assumptions: Comment on Draine and Greenwald (1998). *Journal of Experimental Psychology: General*, 127:311–317.

[Draine and Greenwald, 1998] Draine, S. C. and Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, 127:286–303.

[Eimer and Schlaghecken, 2002] Eimer, M. and Schlaghecken, F. (2002). Links between conscious awareness and response inhibition: Evidence from masked priming. *Psychonomic Bulletin and Review*, 9:514–520.

[Fan and Truong, 1993] Fan, J. and Truong, Y. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics*, 21:1900–1925.

[Fechner, 1966] Fechner, G. T. (1966). *Elements of psychophysics*. Holt, Rinehart and Winston, New York.

[Gelfand and Smith, 1990] Gelfand, A. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.

[Gelman et al., 2004] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian data analysis (2nd edition)*. Chapman and Hall, London.

[Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

[Graves et al., tted] Graves, T. L., Speckman, P. L., and Sun, D. (submitted). Improved mixing in MCMC algorithms for linear models.

[Green and Luce, 1975] Green, D. M. and Luce, R. D. (1975). Parallel psychometric functions from a set of independent detectors. *Psychological Review*, 82:483–486.

[Green and Swets, 1966] Green, D. M. and Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley. Reprinted by Krieger, Huntington, N.Y., 1974, New York.

[Greenwald et al., 1995] Greenwald, A., Klinger, M., and Schuh, E. (1995). Activation by marginally perceptible ("subliminal") stimuli: Dissociation of unconscious from conscious cognition. *Journal of Experimental Psychology: General*, 124:22–42.

[Greenwald et al., 2003] Greenwald, A. G., Abrams, R. L., Naccache, L., and Dehaene, S. (2003). Long-term semantic memory versus contextual memory in unconscious number processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 29:235–247.

[Greenwald et al., 1996] Greenwald, A. G., Draine, S. C., and Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science*, 273(5282):1699–1702.

[Greenwald et al., 1989] Greenwald, A. G., Klinger, M. R., and Liu, T. J. (1989). Unconscious processing of dichoptically masked words. *Memory and Cognition*, 17:35–47.

[Holender and Duscherer, 2004] Holender, D. and Duscherer, K. (2004). Unconscious perception: The need for a paradigm shift. *Perception & Psychophysics*, 66:872–881.

[Jacoby, 1991] Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30:513–541.

[Jaskowski et al., 2003] Jaskowski, P., Skalska, B., and Verleger, R. (2003). How the self controls its "automatic pilot" when processing subliminal information. *Journal of Cognitive Neuroscience*, 15:911–920.

[Jeffreys, 1961] Jeffreys, H. (1961). *Theory of Probability (3rd Edition)*. Oxford University Press, New York.

[Kunde et al., 2003] Kunde, W., Keisel, A., and Hoffman, J. (2003). Conscious control over the content of unconscious cognition. *Cognition*, 88:223–242.

[Kunst-Wilson and Zajonc, 1980] Kunst-Wilson, W. R. and Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, 207:557–558.

[Liu and Sabatti, 2000] Liu, S. J. and Sabatti, C. (2000). Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika*, 87:353–369.

[Loftus et al., 1992] Loftus, G. R., Duncan, J., and Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, 18:530–549.

[Lord and Novick, 1968] Lord, F. M. and Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley, Reading, MA.

[Luce, 1963a] Luce, R. D. (1963a). Detection and recognition. In Luce, R. D., Bush, R. R., and Galanter, E., editors, *Handbook of mathematical psychology (Vol. 1)*, pages 103–189. Wiley, New York.

[Luce, 1963b] Luce, R. D. (1963b). A threshold theory for simple detection experiments. *Psychological Review*, 70:61–79.

[Mattler, 2006] Mattler, U. (2006). On the locus of priming and inverse priming effects. *Perception & Psychophysics*, 68:975–991.

[Merikle et al., 2001] Merikle, P., Smilek, D., and Eastwood, J. (2001). Perception without awareness: perspectives from cognitive psychology. *Cognition*, 79:115–134.

[Merikle and Reingold, 1998] Merikle, P. M. and Reingold, E. M. (1998). On demonstrating unconscious perception: Comment on Draine and Greenwald (1998). *Journal of Experimental Psychology: General*, 127:304–310.

[Miller, 2000] Miller, J. (2000). Measurement error in subliminal perception experiments: simulation analyses of two regression methods. *Journal of Experimental Psychology: Human Perception and Performance*, 26:1461–1477.

[Morey et al., ress] Morey, R. D., Pratte, M. S., and Rouder, J. N. (in press). Problematic effects of aggregation in zROC analysis and a hierarchical modeling solution. *Journal of Mathematical Psychology*.

[Morey et al., 2008] Morey, R. D., Rouder, J. N., and Speckman, P. L. (2008). A statistical model for discriminating between subliminal and near-liminal performance. *Journal of Mathematical Psychology*, 52:21–36.

[Murphy and Zajonc, 1993] Murphy, S. T. and Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, 64:723–739.

[Nachmias, 1981] Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research*, 21:215–223.

[Pratte and Rouder, tted] Pratte, M. S. and Rouder, J. N. (submitted). A task-difficulty artifact in subliminal priming.

[Reber, 1967] Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6:855–863.

[Reingold and Merikle, 1988] Reingold, E. and Merikle, P. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*, 44:563–575.

[Reynvoet and Ratinckx, 2004] Reynvoet, B. and Ratinckx, E. (2004). Hemispheric differences between left and right number representations: Effects of conscious and unconscious priming. *Neuropsychologia*, 42:713–726.

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

[Rouder et al., 2007] Rouder, J. N., Morey, R. D., Speckman, P. L., and Pratte, M. S. (2007). Detecting chance: A solution to the null sensitivity problem in subliminal priming. *Psychonomic Bulletin and Review*, 14:597–605.

[Roufs, 1974] Roufs, J. A. J. (1974). Dynamic properties of vision—VI. Stochastic threshold fluctuations and their effect on flash-to-flicker sensitivity ratio. *Vision Research*, 14:871–888.

[Shibuya and Bundesen, 1988] Shibuya, H. and Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance*, 14:591–600.

[Snodgrass et al., 2004] Snodgrass, M., Bernat, E., and Shevrin, H. (2004). Unconscious perception: A model-based approach to method and evidence. *Perception & Psychophysics*, 66:846–867.

[Swaminathan and Gifford, 1982] Swaminathan, H. and Gifford, J. A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics*, 7:175–191.

[Swets, 1961] Swets, J. A. (1961). Is there a sensory threshold? *Science*, 134:168–177.

[Swets, 1996] Swets, J. A. (1996). *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics*. Laurence Erlbaum Associates, Mahwah, NJ.

[Taylor and Creelman, 1967] Taylor, M. M. and Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 41:782–787.

[Tierney, 1994] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1728.

[Townsend, 1981] Townsend, J. T. (1981). Some characteristics of visual whole report behavior. *Acta Psychologica*, 47:149–173.

[Vorberg et al., 2003] Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., and Schwarzbach, J. (2003). Different time course for the visual perception and action priming. *Proceedings of the National Academy of Sciences*, 100:6275–6280.

[Watson, 1979] Watson, A. B. (1979). Probability summation over time. *Vision Research*, 19:515–522.

[Watson and Pelli, 1983] Watson, A. B. and Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33:113–120.

VITA

Richard D. Morey was born on November 26, 1978 in Springfield, Massachusetts. He received a Bachelor's degree in Music with a minor in Philosophy from the Florida State University in 2001. In 2002, he joined the Department of Psychological Sciences at the University of Missouri-Columbia as a PhD student in the Perception and Cognition Lab. In 2005 Richard joined the Statistics department as a Master's student. In August of 2008, he will join the Department of Psychology at the Royal University of Groningen in the Netherlands as an Assistant Professor.