Available online at www.sciencedirect.com
ScienceDirect
Journal of Mathematical Psychology 52 (2008) 21–36
Journal of Mathematical Psychology
www.elsevier.com/locate/jmp

A statistical model for discriminating between subliminal and near-liminal performance

Richard D. Morey[a,*], Jeffrey N. Rouder[a], Paul L. Speckman[b]

[a]Department of Psychological Sciences, 210 McAlester Hall, University of Missouri-Columbia, MO 65211, USA
[b]Department of Statistics, 210 McAlester Hall, University of Missouri-Columbia, MO 65211, USA

Received 21 May 2007; received in revised form 13 September 2007
Available online 5 November 2007

Abstract

The concept of a psychophysical threshold is foundational in perceptual psychology. In practice, thresholds are operationalized as stimulus values that lead to a fairly high level of performance such as .75 or .707 in two-choice tasks. These operationalizations are not useful for assessing subliminality—the state in which a stimulus is so weak that performance is at chance. We present a hierarchical Bayesian model of performance that incorporates a threshold that divides subliminal from near-liminal performance. The model provides a convenient means to measure at-chance thresholds and therefore is useful for testing theories of subliminal priming. The hierarchical nature of the model is critical for efficient analysis as strength is pooled across people and stimulus values. A comparison to Rasch psychometric models is provided.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Subliminal; Priming; Psychophysics; Psychometrics; Rasch; Bayesian statistics; Hierarchical models; MAC; Mass at chance; Chance performance

0. Introduction

The concept of a threshold has been a mainstay of psychophysics. With the advent of the method of constant stimuli, psychophysicists were able to define thresholds objectively. Stimuli below threshold are unable to be detected or identified; those above it may be detected at some level above a chance baseline. A classic source of evidence for the existence of thresholds comes from the detection of brief light pulses against dark backgrounds. For a given background, there exist light pulses that are so faint that they are unable to be detected above chance (see Rudd, 1996, for a review). In psychophysical research with forced choice designs, however, the threshold is not operationalized as the maximum intensity value for which performance is at chance. Instead, thresholds are operationalized as a stimulus value that leads to some predetermined level of performance. Popular performance levels are .75, which is halfway between floor and ceiling in 2AFC designs, and .707, which results from an adaptive staircase procedure (e.g., Taylor & Creelman, 1967). The reason for these choices of above-chance levels is their statistical convenience. In this paper, we provide a principled model for measuring an at-chance threshold.

One domain in which it is critical to measure an at-chance threshold is subliminal priming. According to many researchers, primes may be presented so quickly as to be below threshold yet 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 Fig. 1; single digits serve as both primes and targets. In the first stage of the experiment, participants

*Corresponding author.
E-mail address: moreyr@missouri.edu (R.D. Morey).

0022-2496/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2007.09.007

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 identification is faster for congruent primes than for incongruent ones. In a separate stage of the experiment, participants identify primes by their less-than-five status. Dehaene et al., among others, claim that priming occurs even when prime identification is at chance. To support the claim that prime identification is at chance, a method for determining at-chance thresholds is needed. The more conventional thresholds corresponding to accuracies of .75 and .707 are not useful.

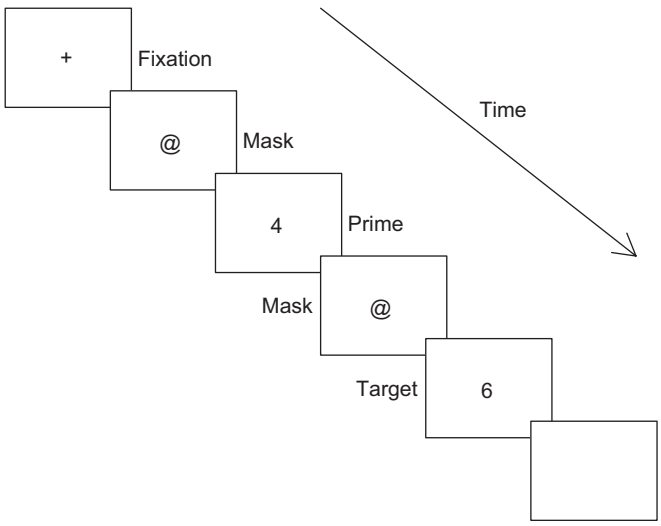

Fig. 1. The structure of a trial in the number priming paradigm. The displayed prime is incongruent with the target.

Accurate measurement of the objective at-chance threshold would greatly aid in testing theories of priming. Fig. 2A shows four different theoretical relationships between prime identification 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 objective identification threshold (e.g., Cheesman & Merikle, 1984). The dotted-dashed, 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.

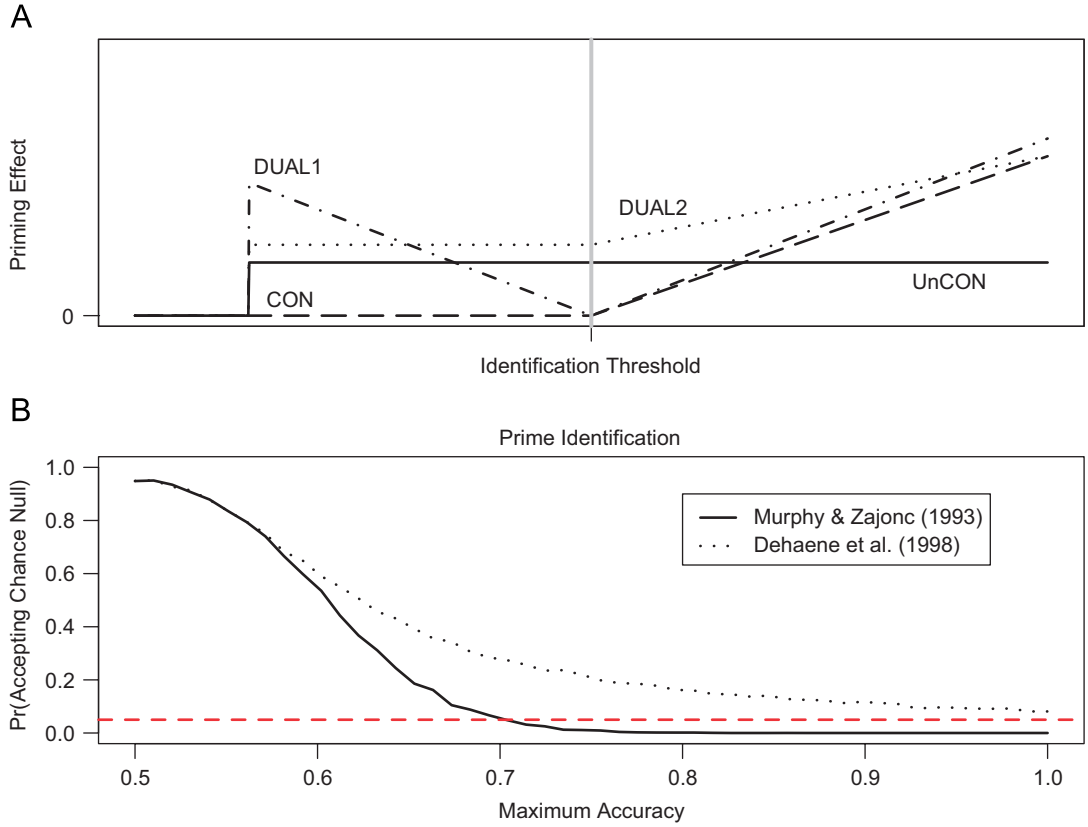

Fig. 2. (A) Predictions of four substantive priming theories. *UnCON*: priming reflects unconscious processes; *CON*: priming reflects conscious processes (Cheesman and Merikle, 1988); *DUAL1*: dual-process model of Snodgrass et al. (2004); *DUAL2* dual process model based on Greenwald et al. (1996). The vertical line denotes the objective identification threshold (Figure adapted from Snodgrass et al., 2004.) (B) Probability of incorrectly accepting chance performance for the sample sizes in Dehaene et al. (1998), Task 1, and Murphy and Zajonc (1993), Experiment 1. (Fig. adapted from Rouder et al., 2007.)

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. The statistical model developed in this paper serves as a tool in testing these four substantive theories.

One vexing problem in this line of research is determining whether a prime is above or below threshold. A standard approach to this problem is to perform a *t*-test on prime-identification performance. The null hypothesis is that performance is at chance. A failure to reject this null is taken as evidence for subliminality of the primes. Reingold and Merikle (1988) offered the following seminal critique of this logic: it is difficult to know whether a null result indicates that true performance is at chance or that the test failed to detect weak but above-chance true performance. To assess the importance of this critique, Rouder, Morey, Speckman, and Pratte (2007) performed simulations to examine the probability that researchers mistakenly accept the null. In these simulations, true accuracies for each participant were drawn from a uniform distribution between .5 and a specified maximum value greater than .5. Thus, all participants have true accuracy above chance. From these true accuracies, data were simulated. The probability of falsely accepting the null decreases with increasing sample size (both participants and number of trials per participant). In their simulations, Rouder et al. chose sample sizes from two well-cited studies: Task 1 of Dehaene et al. (1998) and Experiment 1 of Murphy and Zajonc (1993). Fig. 2B, adapted from Rouder et al., shows the results of these simulations. Even though every hypothetical participant identified the primes above chance, the *t*-test method accepts the null at alarmingly high rates. These results indicate that the studies of Dehaene et al. (1998) and Murphy and Zajonc (1993) are underpowered to detect weak but superliminal primes. This critique applies to other methods of assessing the subliminality of primes including thresholding (Dagenbach, Carr, & Wilhelmsen, 1989) and selecting participants whose observed performance is at or below chance (Greenwald, Klinger, & Liu, 1989).

In our previous work (Rouder et al., 2007), we proposed a Bayesian hierarchical model to select participants whose true prime identification performance is at chance. We termed this model the *mass* at chance (MAC) model to emphasize that a threshold is explicitly assumed and modeled. Importantly, MAC is immune to Reingold and Merikle's critique; that is, researchers cannot simply accept a null by underpowering their experiments. The model is used to identify participants for whom the primes are below threshold. If these identified participants display priming effects, then the researcher may conclude with a controlled error rate that these priming effects are indeed subliminal.

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-chance and above-chance performances, we propose an expansion of the subliminal priming paradigm and a corresponding expansion of the MAC model. In the first section, we briefly present the previous MAC model and highlight its limitations. Following that, we present the expanded paradigm and an expanded MAC model. We show how this expanded model provides for highly accurate measurement of true performance and, consequently, provides a method for assessing priming theories (see Fig. 2A) and estimating at chance thresholds.

1. Rouder et al.'s MAC model

At the first level of Rouder et al.'s MAC model, the number of correct prime identifications for the *i*th participant, $i = 1, \ldots, I$, denoted by y_i, is

$$y_i \overset{\text{indep.}}{\sim} \text{Binomial}(p_i, N_i). \tag{1}$$

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

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

$$p_i = \begin{cases} .5, & x_i \leqslant 0, \\ \Phi(x_i), & x_i > 0, \end{cases} \tag{2}$$

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 normally in the population:

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

True scores depend on stimulus 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 duration. Whereas most priming experiments are

performed with a single duration, we previously developed MAC for a single parameter μ. Priors are needed for μ and σ^2. Based on extensive simulation, we chose somewhat informative priors $\mu \sim \text{Normal}(0,1)$ 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:

$$\omega_i = \Pr(x_i < 0 \mid \text{data}). \tag{3}$$

The decision rule is

$$\begin{aligned} &\omega_i \geqslant .95 \rightarrow \text{Conclude the participant is at chance,} \\ &\omega_i < .95 \rightarrow \text{Conclude the participant is above chance.} \end{aligned} \tag{4}$$

The main benefit of the MAC model of this approach is that it mitigates Reingold and Merikle's null-sensitivity critique. 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 subliminality rather than less difficult. Therefore, the MAC model provides an appropriate safeguard against underpowered experiments.

Although the MAC model provides a principled means for testing subliminal priming, it has three limitations. First, as mentioned previously, the model does not select people as at chance often. Fig. 3 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. There is not enough information to select more than three participants as at chance, making it very difficult to test for a priming effect. Second, the MAC model is quite dependent on specification of the prior. In particular, the posterior of σ^2 was heavily influenced by the choice of prior.[1] This dependence is undesirable. Finally, because the MAC model is applicable to only a single prime duration, it cannot be used to investigate the relationship between prime identification and priming (Fig. 2A).

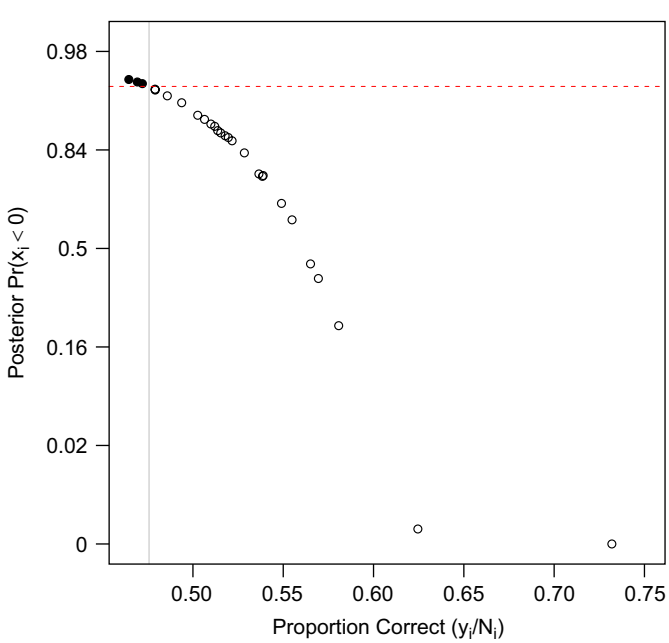

Fig. 3. Posterior probability that an individual participant is at chance as a function of observed accuracy. Participants whose posterior probability is above .95 (closed circles) are selected as at chance. (Fig. adapted from Rouder et al., 2007.)

All three of these limitations may be mitigated by allowing prime duration to vary across a range of values. We 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; and (3) the model may be used to measure a range of performance including those below threshold. Consequently, the extended model is ideal for testing the theories in Fig. 2A and estimating at-chance thresholds.

2. The extended MAC model

It is straightforward to specify the extended MAC model for multiple prime durations. 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 in the jth stimulus duration condition:

$$y_{ij} \overset{\text{indep.}}{\sim} \text{Binomial}(p_{ij}, N_{ij}). \tag{5}$$

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

$$p_{ij} = \begin{cases} .5, & x_{ij} \leqslant 0, \\ \Phi(x_{ij}), & x_{ij} > 0. \end{cases} \tag{6}$$

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

$$x_{ij} = \alpha_i + \mu_j. \tag{7}$$

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

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

The additivity assumption provides a convenient means of pooling information about latent ability (α_i) across duration conditions and about stimulus 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. Consequently, the model provides a means of

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

testing the relationships between priming and performance in Fig. 2A. The threshold for participant i is the duration 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.

2.1. The MAC link function

In both Rouder et al.'s MAC model and the extended MAC model, true scores x_{ij} are mapped to probabilities using the truncated-probit link in Eqs. (2) and (6). This link function is shown as the solid line in Fig. 4A. 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 dashed 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.

Fig. 4B, 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. (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, 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 Fig. 4C, the additivity of stimulus ease and latent ability results in specific relations between individual-level psychophysical functions. For smooth convex mappings, such as those in Fig. 4B, participants with smaller thresholds have smaller rates of gain in their predicted psychometric functions, as seen in Fig. 4C. Second, the discontinuity in the first derivative of the MAC link ensures that there is an at-chance threshold in the psychophysical function. We discuss methods of assessing whether these constraints hold in data when analyzing the experiment.

Fig. 4C 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 invariant upto translation. 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. We 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).

3. Prior distributions

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

$$\mu_j \overset{\text{iid}}{\sim} \text{Normal}(\mu_0, \sigma_0^2), \tag{8}$$

$$\sigma^2 \sim \text{Inverse Gamma}(a_0, b_0). \tag{9}$$

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

$$f(\sigma^2 \mid a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)} (\sigma^2)^{-a_0-1} \exp\left\{-\frac{b_0}{\sigma^2}\right\}, \quad \sigma^2, a_0, b_0 > 0. \tag{10}$$

The parameters in the priors (μ_0, σ_0^2, a_0, b_0) are chosen before analysis. We place a noninformative Jeffreys prior on σ^2 by setting $a_0 = b_0 = 0$ (Jeffreys, 1982).[2] We 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

[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 $I > 2$.

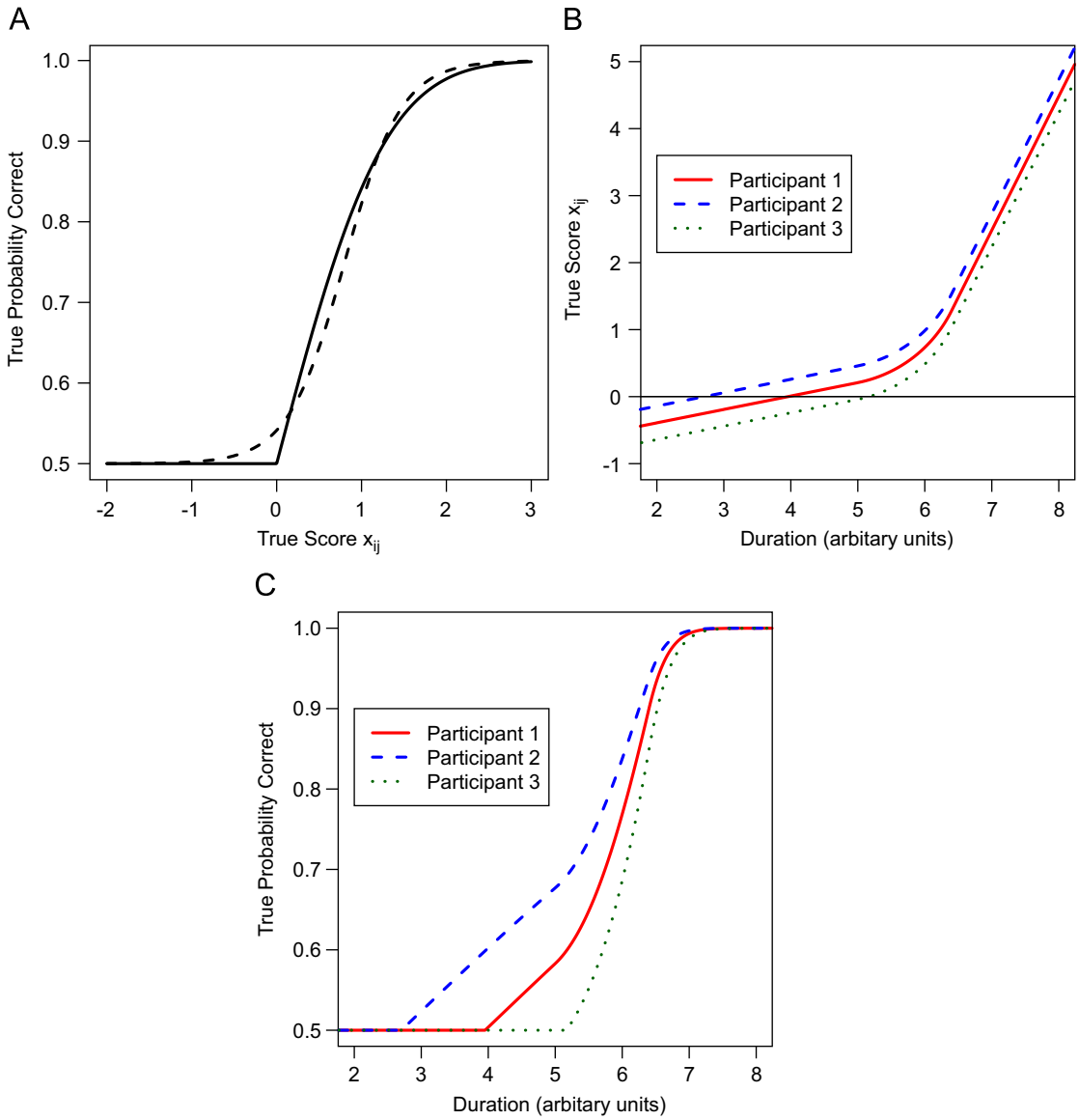

Fig. 4. (A) The MAC truncated-probit link function (solid line) and a logistic link function (dashed line). (B) A hypothetical mapping from duration to true score for three participants. (C) A hypothetical mapping from duration to probability correct for the same three participants.

$\mu_0 = 0$ 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, we tried other values as high as $\sigma_0^2 = 16$. The effect is inconsequential.

4. Model analysis

Although the model is straightforward to specify, analysis is complicated by the nonlinearity in the MAC link. Closed-form expressions for the marginal 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, we 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 in the jth duration condition, 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:

$$y_{ijk} = \begin{cases} 0, & w_{ijk} \leqslant 0, \\ 1, & w_{ijk} > 0. \end{cases}$$

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}.

4.1. Full conditional distributions

We present the full conditional distributions of each of the parameters. Derivations are provided in Appendix A. 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

$$\sigma^2 \mid \boldsymbol{\alpha} \sim \text{InverseGamma}\left(a_0 + \frac{I}{2}, \sum_i \alpha_i^2/2 + b_0\right). \tag{11}$$

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

$$\begin{aligned} &w_{ijk} \mid x_{ij}; (y_{ijk} = 1) \sim \text{TN}_{(0,+\infty)}(x_{ij} \vee 0, 1), \\ &w_{ijk} \mid x_{ij}; (y_{ijk} = 0) \sim \text{TN}_{(-\infty,0)}(x_{ij} \vee 0, 1), \end{aligned} \tag{12}$$

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

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 $A_0 = (-\mu_{(1)}, \infty)$, $A_1 = (-\mu_{(2)}, -\mu_{(1)}), \ldots, A_J = (-\infty, -\mu_{(J)})$, where $\mu_{(1)} < \cdots < \mu_{(J)}$ are the order statistics of the μ_j. Let

$$J_0 = \{1, \ldots, J\},$$
$$J_\ell = \{j : \mu_j > \mu_{(\ell)}\},$$

and define

$$s_{\alpha_i \ell} = \begin{cases} \left(\dfrac{1}{\sigma^2} + \sum_{j \in J_\ell} N_{ij}\right)^{-1}, & 0 \leqslant \ell < J, \\ \sigma^2, & \ell = J, \end{cases} \tag{13}$$

$$m_{\alpha_i \ell} = \begin{cases} s_{\alpha_i \ell} \sum_{j \in J_\ell} \left(\sum_{k=1}^{N_{ij}} w_{ijk} - N_{ij}\mu_j\right), & 0 \leqslant \ell < J, \\ 0, & \ell = J, \end{cases} \tag{14}$$

$$h_{\alpha_i \ell} = \begin{cases} \sum_{j \in J_\ell} \left(N_{ij}\mu_j^2 - 2\mu_j \sum_{k=1}^{N_{ij}} w_{ijk}\right), & 0 \leqslant \ell < J, \\ 0, & \ell = J. \end{cases} \tag{15}$$

Then

$$[\alpha_i \mid \boldsymbol{\mu}, \sigma^2, \mathbf{w}] \propto \sum_{\ell=0}^{J} \exp\left\{-\frac{1}{2}\left(h_{\alpha_i \ell} - \frac{(m_{\alpha_i \ell})^2}{s_{\alpha_i \ell}}\right)\right\} \times \exp\left\{-\frac{(\alpha_i - m_{\alpha_i \ell})^2}{2 s_{\alpha_i \ell}}\right\} I_{(\alpha_i \in A_\ell)}. \tag{16}$$

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 (16) is proportional to the density of a mixture of truncated normals. Set $\mu_{J+1} = \infty$ and $\mu_0 = -\infty$. Let

$$q_{\alpha_i \ell} = \exp\left\{-\frac{1}{2}\left(h_{\alpha_i \ell} - \frac{(m_{\alpha_i \ell})^2}{s_{\alpha_i \ell}}\right)\right\} \times \sqrt{2\pi s_{\alpha_i \ell}} \left[\Phi\left(\frac{-\mu_{(\ell)} - m_{\alpha_i \ell}}{\sqrt{s_{\alpha_i \ell}}}\right) - \Phi\left(\frac{-\mu_{(\ell+1)} - m_{\alpha_i \ell}}{\sqrt{s_{\alpha_i \ell}}}\right)\right], \tag{17}$$

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

$$\alpha_i \mid \boldsymbol{\mu}, \sigma^2, \mathbf{w} \sim \sum_{\ell=0}^{J} p_{\alpha_i \ell} \text{TN}_{A_\ell}(m_{\alpha_i \ell}, s_{\alpha_i \ell}). \tag{18}$$

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

$$I_0 = \{1, \ldots, I\},$$
$$I_\ell = \{i : \mu_i > \alpha_{(\ell)}\},$$

and define

$$s_{\mu_j \ell} = \begin{cases} \left(\dfrac{1}{\sigma_0^2} + \sum_{i \in I_\ell} N_{ij}\right)^{-1}, & 0 \leqslant \ell < I, \\ \sigma_0^2, & \ell = I, \end{cases}$$

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

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

Then

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

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

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

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

$$\mu_j \mid \boldsymbol{\alpha}, \mathbf{w} \sim \sum_{\ell=0}^{I} p_{\mu_j\ell} \mathrm{TN}_{M_\ell}(m_{\mu_j\ell}, s_{\mu_j\ell}). \tag{19}$$

4.1.1. Sampling full conditional distributions

Implementing the Gibbs sampler requires samples from the full conditional distributions (11), (12), (18), and (19). We provide efficient algorithms below.

4.1.1.1. 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 (cf., Ahrens & Dieter, 1974).

4.1.1.2. 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 $U \sim \mathrm{Uniform}(\Phi((l-\mu)/\sigma), \Phi((u-\mu)/\sigma))$. Then,

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

4.1.1.3. 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 (18) 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. We discuss below approximations to ensure sufficient numeric precision.

From inspection (17), 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. We 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 $\mathrm{TN}_{A_v}(m_{\alpha_i\ell}, s_{\alpha_i\ell})$ distribution. Sampling from the full conditional posterior for μ_j is done in the same manner.

4.1.1.4. 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 Fig. 5A, 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, we show all samples between iterations 5000 and 6000. As can be seen, there is a fair degree of autocorrelation (Fig. 5B provides a plot of the autocorrelation function). We first 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:

$$x_{ij} = \alpha_i + \mu_j = (\alpha_i + z) + (\mu_j - z).$$

For any value of z, the likelihood of the model is the same. The consequence is that if the estimate of α_i on an 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.

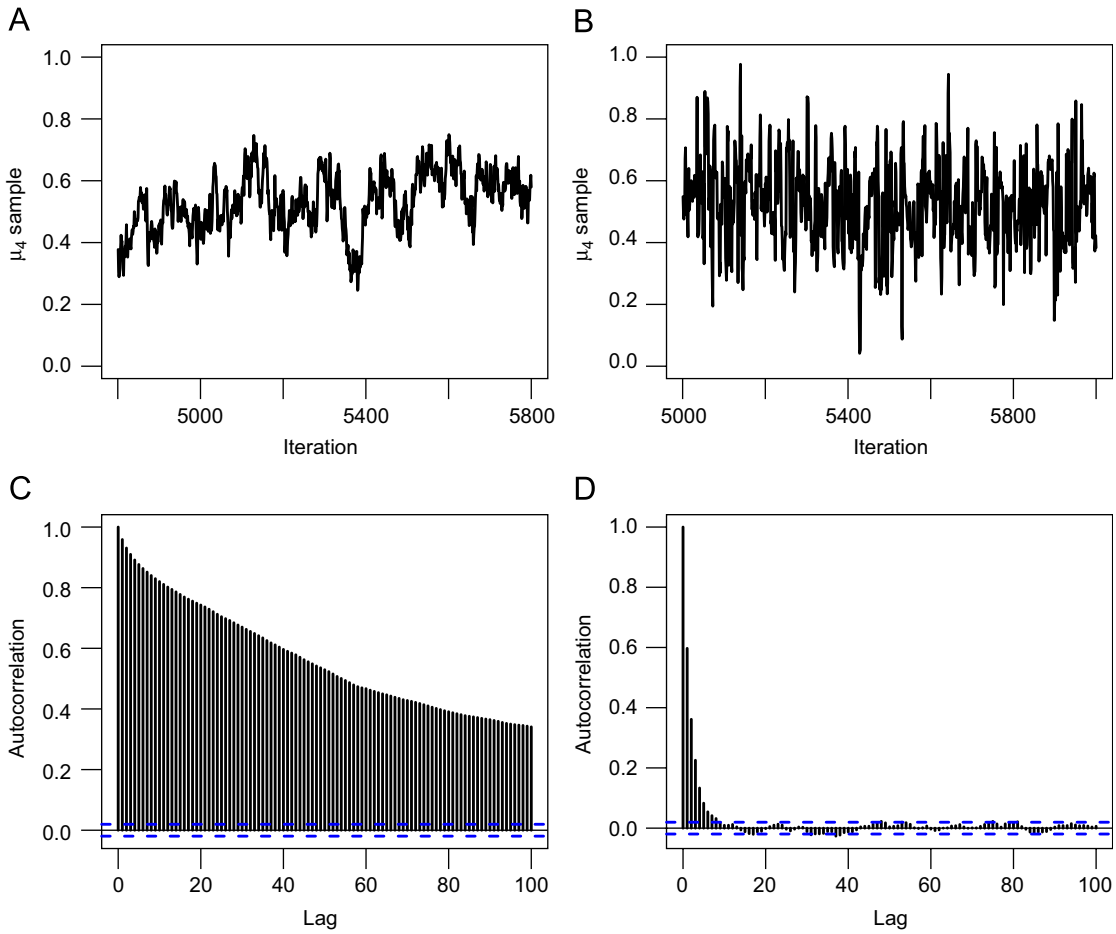

Fig. 5. Autocorrelation in MCMC Chains. (A) Segment from the MCMC chain for parameter μ_4 without a decorrelating step. (B) Autocorrelation function of parameter μ_4. (C & D) Segment from the MCMC chain and autocorrelation with an additional decorrelating step.

To mitigate autocorrelation in the chains, we 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 $\boldsymbol{\alpha}^{(t)}, \boldsymbol{\mu}^{(t)}$:

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

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. Fig. 5C and D 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).

5. Simulation

To assess the accuracy of parameter estimation and the ability to select participant-by-duration combinations that exhibited performance at chance, we performed a small simulation study. In the simulation, 22 hypothetical participants observed 540 trials. These trials consisted of 90 replicates of six hypothetical stimulus durations. These choices of sample sizes reflect the experiment we report subsequently. The "true values" of stimulus ease for the six stimulus durations were $\boldsymbol{\mu} = (-1.30, -0.46, -0.13, 0.53, 0.92, 1.18)$. True values are also needed for latent abilities. We sampled these true values 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. (7) and true probabilities p_{ij} were computed by Eq. (6). For each participant-by-item combination, data were sampled from a binomial with probability p_{ij}. The data were then analyzed with the extended MAC model. MCMC chains were run for 10,000 iterations with the first 1000 iterations serving as a burn-in period. This process of sampling true values, generating data, and estimating parameters was repeated 1000 times.

Fig. 6A 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 durations; that is, $\alpha_i < -\mu_6$, where $\mu_6 = 1.18$ is the easiest duration condition. 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 stimulus durations.

Estimates of μ (posterior means) are shown in Fig. 6B. Each density is a smoothed histogram of the 1000 posterior means μ_j in the jth duration condition. The vertical 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 extended MAC model is

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

As before, we accept participant-by-item combinations as at chance only if $\omega_{ij} \geqslant .95$. 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 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

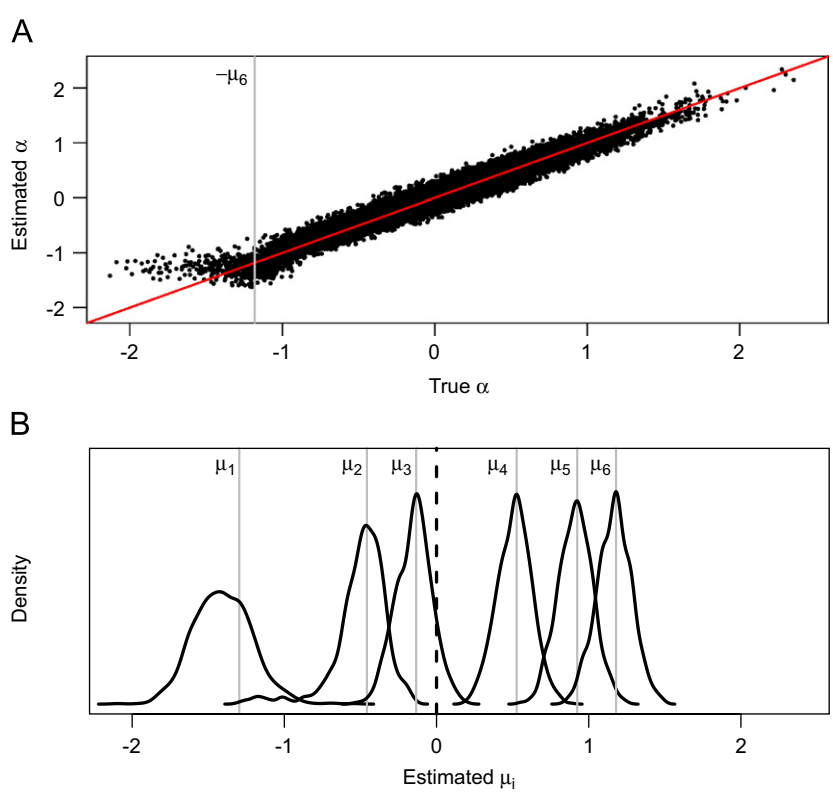

Fig. 6. Results of simulation. (A) Estimates of latent ability α as a function of true value. True latent abilities less than $-\mu_6$ (vertical line) are from simulated participants whose performance is at chance for all of the six duration conditions. (B) Distribution of point estimates (posterior means) of μ_j across replicates (smoothed histograms). The vertical lines represent the true values of each μ.

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 we accept all combinations with $\omega_{ij} \geqslant .95$. We will therefore accept many combinations for which ω_{ij} is greater than .95, 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 durations of 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.

6. Experiment

The following experiment illustrates the application of the extended MAC model and provides an informal assessment of the additivity assumption in (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 to 100.0 ms. The prime identification task is the target of inquiry in the extended 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 (e.g., Pratte, 2007).

6.1. Method

6.1.1. Participants

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

6.1.2. 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 × 480 pixels, a refresh rate of 120 Hz, and controlled by PC-compatible Pentium IV computers.

6.1.3. Design

The main independent variable was prime duration, which was manipulated through the following six levels: 16.7, 25.0, 41.7, 58.3, 75.0, and 100.0 ms. Other independent variables were the prime and target 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.

6.1.4. Procedure

The structure of a trial is shown in Fig. 1. Primes were degraded by the inclusion of forward and backward 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 six blocks of 90 trials, for a total of 540 trials. The sessions lasted approximately 20 min.

6.2. Results

Mean and individual accuracies are plotted as a function of duration in Fig. 7A. Performance is near chance for primes at the shortest duration and increases in an orderly manner.

The data were analyzed with the extended MAC model. We performed 10,000 MCMC iterations with the first 1000 serving as burn in. We ran the analysis several times with different starting values to ensure the chains had converged. The resulting chains exhibited low autocorrelation and good convergence. Fig. 8 shows a representative segment of the MCMC chain (between iterations 5000 and 6000) for a few select parameters.

Estimates of μ_j (posterior means) for each duration are shown as the large points in (Fig. 7B). 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

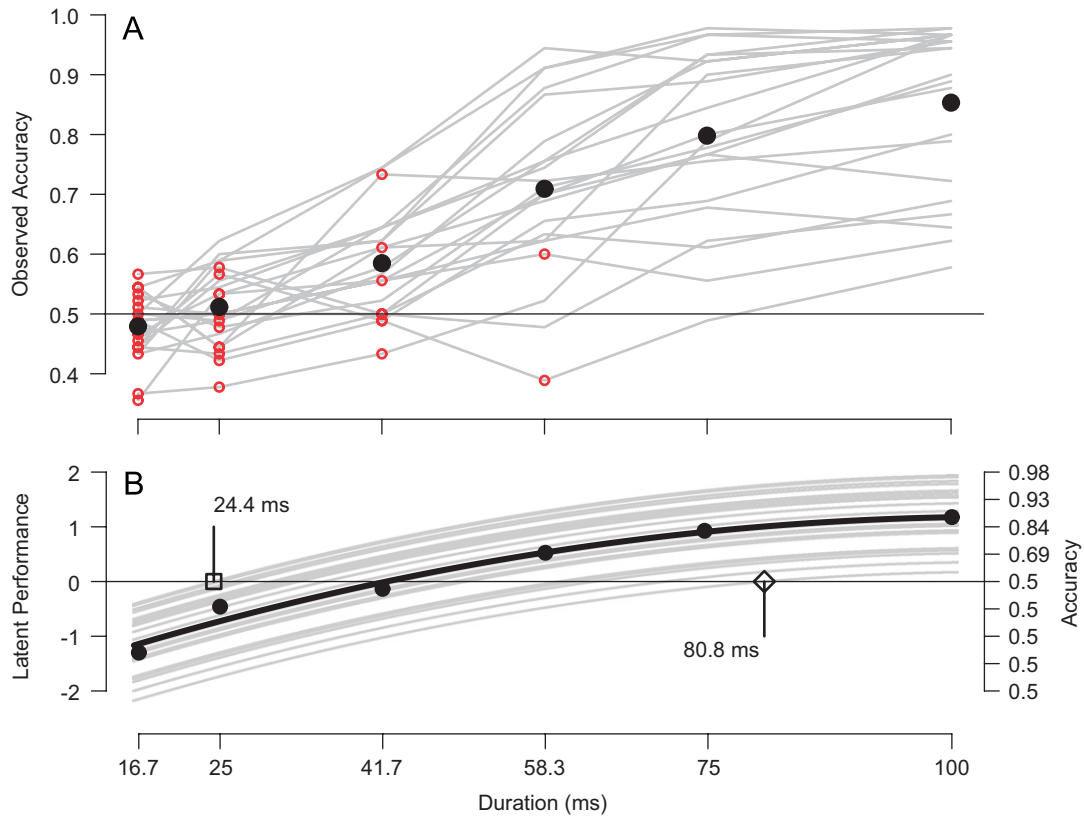

Fig. 7. (A) Mean and individual accuracy are denoted by dark points and light lines, respectively. Open circles denote participant-by-duration combinations that are classified by the extended MAC model as below threshold. (B) Points are estimates of duration condition ease μ_j. Each line is a quadratic regression of true score on prime duration. The square and diamond symbols denote the smallest and largest thresholds estimated across the 22 participants.

function of duration. (Fig. 7B). 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 quadratic regression yields an estimate of a participant's threshold. The smallest and largest individual thresholds are indicated in Fig. 7B as a square and diamond, respectively. Of note is that we did not assume a quadratic form until it was suggested by the data.

To assess the fit of the model, we plotted observed accuracy as a function of predicted accuracy (Fig. 9A). 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.

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

$$r_{ij} = \frac{y_{ij} - N_{ij}\hat{p}_{ij}}{\sqrt{N_{ij}\hat{p}_{ij}(1-\hat{p}_{ij})}}, \tag{20}$$

where $\hat{p}_{ij}$ is the predicted accuracy for the participant and duration. The lines in Fig. 9B shows these residuals for participants when $\hat{p}_{ij} > .5$. The gray 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 misspecifications. 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 by "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 of thresholds. The extreme point denoted by

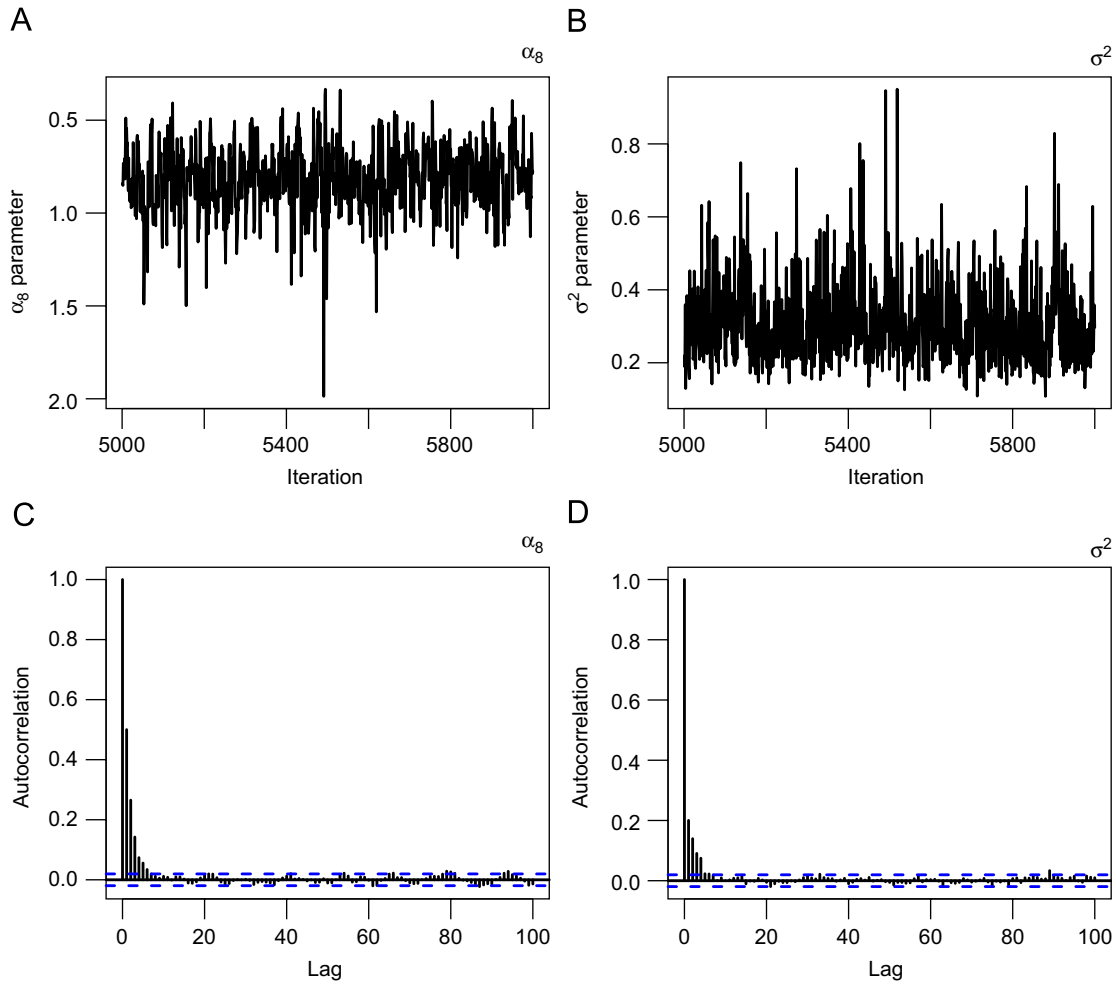

Fig. 8. Segments from the MCMC chains and autocorrelation functions for selected parameters.

"a" in Fig. 9A arises from the participant denoted by an "a" in Fig. 9B as a consequence of misspecification. Researchers may need to analyze residuals to identify participants for whom the model is misspecified. Future versions of the model may need an additional individualized slope terms, though this extension is nontrivial.

7. General discussion

The extended 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, we discuss the relationship of the extended 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: $x_{ij} = \alpha_i + \mu_j$. As in item response theory, this assumption is leveraged to provide better estimation of true scores. We remain concerned, however, about a lack of robustness to misspecification. Although additivity holds for the majority of our 22 participants, there are noteworthy exceptions (Fig. 9B). If the rate of gain is different for each participant, additivity will be violated. Fortunately, additivity can be assessed informally by studying residuals, and we recommend researchers 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

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

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

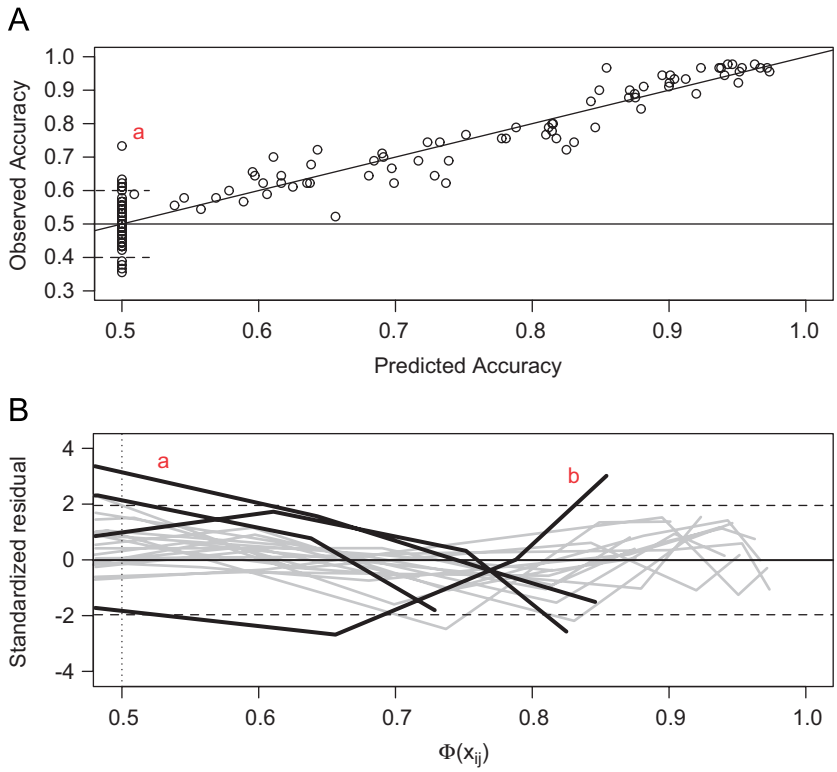

Fig. 9. (A) Observed accuracy as a function of predicted accuracy. Each point denotes a participant-by-duration combination. Horizontal dashed lines represent the .025 and .975 quantiles of accuracy for a participant at chance. (B) Standardized residuals from the MAC model. Light and dark lines denote 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.

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

$$\theta = \sigma,$$
$$\gamma_i = \alpha_i/\sigma,$$
$$\eta_j = -\mu/\sigma.$$

With this reparameterization, the MAC model may be expressed as

$$p_{ij} = \tfrac{1}{2} + \tfrac{1}{2}G(\theta[\gamma_i - \eta_j]),$$

$$\gamma_i \overset{\text{iid}}{\sim} \text{Normal}(0,1),$$

$$G(x) = \begin{cases} 2[\Phi(x) - .5], & x \geqslant 0, \\ 0, & x < 0. \end{cases}$$

As can be seen, the sole difference between the MAC model and the Rasch model is the link function. The MAC model has a truncated-probit link; the Rasch model has a logistic link. These two links are shown in Fig. 4. 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.

Acknowledgments

We thank Mike Pratte and Andrew Kent for help in running the reported experiment. This research is supported by an Adeline Hoffman Fellowship, NSF Grant SES-0351523 and NIMH Grant R01-MH071418.

Appendix A

In this appendix, we provide the joint posterior distribution and prove Facts 1–4 about the conditional posterior distributions.

A.1. Joint posterior

Let $\mathbf{y}$ be the collection of data, $\mathbf{y} = \langle\langle\langle y_{ijk}\rangle_{k=1}^{k=N_{ij}}\rangle_{j=1}^{j=J}\rangle_{i=1}^{i=I}$, and $\mathbf{w}$ be the collection of latent parameters, $\mathbf{w} = \langle\langle\langle w_{ij}\rangle_{k=1}^{k=N_{ij}}\rangle_{j=1}^{j=J}\rangle_{i=1}^{i=I}$. The joint posterior of all

parameters is

$$
\begin{aligned}
&[\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\mu}, \sigma^2 \mid \mathbf{y}] \\
&\quad \propto \prod_i \prod_j \left\{ \prod_{\substack{k=1 \\ (y_{ijk}=0)}}^{N_{ij}} \exp\left\{ -\frac{(w_{ijk} - (\alpha_i + \mu_j) \vee 0)^2}{2} \right\} I_{(w_{ijk} \leqslant 0)} \right. \\
&\qquad \left. \times \prod_{\substack{k=1 \\ (y_{ijk}=1)}}^{N_{ij}} \exp\left\{ -\frac{(w_{ijk} - (\alpha_i + \mu_j) \vee 0)^2}{2} \right\} I_{(w_{ijk} > 0)} \right\} \\
&\qquad \times \prod_i (\sigma^2)^{-1/2} \exp\left\{ -\frac{\alpha_i^2}{2\sigma^2} \right\} \\
&\qquad \times \prod_j \exp\left\{ -\frac{(\mu_j - \mu_0)^2}{2\sigma_0^2} \right\} \\
&\qquad \times (\sigma^2)^{-(a_0+1)} \exp\left\{ -\frac{b_0}{\sigma^2} \right\}. \qquad (21)
\end{aligned}
$$

A.2. Conditional posteriors

Proof of Fact 1. By inspection of (21), the full conditional posterior $[\sigma^2 \mid \boldsymbol{\alpha}]$ is

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

Collecting like terms yields

$$
[\sigma^2 \mid \boldsymbol{\alpha}] \propto (\sigma^2)^{-(a_0+I/2+1)} \exp\left\{ -\frac{\frac{1}{2}\sum_i \alpha_i^2 + b_0}{\sigma^2} \right\}. \qquad (22)
$$

The right-hand side of (22) is proportional to the density of an inverse gamma distribution with parameters $a = a_0 + I/2$ and $b = \sum_i \alpha_i^2/2 + b_0$. □

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. □

Proof of Fact 3. Inspection of (21) reveals that the full conditional posterior of α_i satisfies

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

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

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

Expanding the square and simplifying yields

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

Collecting terms in α_i^2 and α_i yields

$$
\begin{aligned}
&[\alpha_i \mid \boldsymbol{\mu}, \mathbf{w}, \sigma^2] \\
&\quad \propto \exp\left\{ -\frac{1}{2} \left[\sum_{j=1}^{J} N_{ij} \mu_j^2 I_{(\alpha_i > -\mu_j)} - 2 \sum_{j=1}^{J} \mu_j I_{(\alpha_i > -\mu_j)} \sum_{k=1}^{N_{ij}} w_{ijk} \right] \right\} \\
&\qquad \times \exp\left\{ -\frac{1}{2} \left[\alpha_i^2 \left(\frac{1}{\sigma^2} + \sum_{j=1}^{J} N_{ij} I_{(\alpha_i > -\mu_j)} \right) \right. \right. \\
&\qquad \left. \left. -2\alpha_i \left(\sum_{j=1}^{J} I_{(\alpha_i > -\mu_j)} \sum_{k=1}^{N_{ij}} w_{ijk} - \sum_{j=1}^{J} N_{ij} \mu_j I_{(\alpha_i > -\mu_j)} \right) \right] \right\}. \qquad (23)
\end{aligned}
$$

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 (23) yields

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

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 (13)–(15), completing the square and simplifying yields (18). □

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

References

Abramowitz, M., & Stegun, I. A. (1965). *Handbook of mathematical functions: With formulas, graphs, and mathematical tables*. New York: Dover.

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

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

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

Bar, M., & Biederman, I. (1998). Subliminal visual priming. *Psychological Science*, *9*, 464–469.

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

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

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

Dehaene, S., Naccache, L., Le Clech, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imaging unconscious semantic priming. *Nature*, *395*, 597–600.

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

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

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

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

Graves, T. L., Speckman, P. L., & Sun, D. (submitted). *Improved mixing in MCMC algorithms for linear models*.

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

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

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

Greenwald, A. G., Klinger, M. R., & Liu, T. J. (1989). Unconscious processing of dichoptically masked words. *Memory & Cognition*, *17*, 35–47.

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

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

Jeffreys, H. (1982). *Theory of probability*. New York: Oxford University Press.

Kunde, W., Keisel, A., & Hoffman, J. (2003). Conscious control over the content of unconscious cognition. *Cognition*, *88*, 223–242.

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

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

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

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

Murphy, S. T., & 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, J. (1981). On the psychometric function for contrast detection. *Vision Research*, *21*, 215–223.

Pratte, M.S. (2007). Subliminal Priming as a Task-Characteristic Artifact. *Unpublished Masters Thesis*. University of Missouri-Columbia.

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

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

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

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

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.

Rudd, M. E. (1996). A neural timing model of visual threshold. *Journal of Mathematical Psychology*, *40*, 1–29.

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

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

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

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

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

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