Available online at www.sciencedirect.com
ScienceDirect
Journal of Mathematical Psychology 51 (2007) 198–204
Journal of Mathematical Psychology
www.elsevier.com/locate/jmp

Analytic expressions for the BCDMEM model of recognition memory

Jay I. Myung[a,*], Maximiliano Montenegro[b], Mark A. Pitt[a]

[a]Department of Psychology, Ohio State University, 225 Psychology Building, 1835 Neil Avenue, Columbus, OH 43210-1287, USA
[b]Department of Mathematics, Science and Technology Education, Ohio State University, Columbus, OH 43210, USA

Received 9 October 2006; received in revised form 2 February 2007
Available online 30 March 2007

Abstract

We introduce a Fourier transformation technique that enables one to derive closed-form expressions of performance measures (e.g., hit and false alarm rates) of simulation-based models of recognition memory. Application of the technique is demonstrated using the bind cue decide model of episodic memory (BCDMEM; [Dennis, S., & Humphreys, M.S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452–478]). In addition to reducing the time required to test the model, which for models like BCDMEM can be excessive, asymptotic expressions of the measures reveal heretofore unknown properties of the model, such as model predictions being dependent on vector length.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Recognition memory; Cognitive modeling; Fourier transformation; Signal detection theory

1. Introduction

A number of simulation-based models of recognition memory have been proposed in recent years. Examples include bind cue decide model of episodic memory (BCDMEM) (Dennis & Humphreys, 2001), REM (Shiffrin & Steyvers, 1997), SAM (Shiffrin, Ratcliff, & Clark, 1990), MINERVA (Hintzman, 1988), and SLiM (McClelland & Chappell, 1998). The popularity of this style of modeling is due in part to the fact that computing technologies permit both flexibility in model design and precision in implementation. A model can be built that incorporates realistic assumptions about the processes underlying encoding, storage, and retrieval. With the aid of random number generators, a model's predictions are obtained as sample averages over a large number of simulation replications.

In this paper, we introduce a technique for deriving closed-form solutions of these models. It involves performing a Fourier transform (FT) on a model, which in essence requires one to rewrite the model in another form. The value of the technique is at least two-fold. Most importantly, the asymptotic expressions provide insights into properties of the model that are difficult to glimpse in the model's original formulation. In addition, the excessive computing time required to derive model predictions and estimate model parameters in simulation-based models is eliminated, thereby facilitating model evaluation. We demonstrate the application of the technique and what can be learned with it using BCDMEM.

BCDMEM is a model of episodic recognition memory that, among other things, simulates human performance in tasks in which participants have to discriminate old from new words after a study phase in which only the old words were presented. The contextual information accompanying each encounter with a word (i.e., episodes) is fundamental to the model and its behavior. Context is encoded with the word and used to discriminate old from new ones during recognition. The essence of the context-based recognition process is briefly reviewed.

Associated with each word is a context vector made of multiple feature units, each taking on the value of either 1 (active, or presence of the feature) or 0 (inactive, or absence of the feature). At study, presentation of a word activates a noisy context vector representing the current experimental context. As a result, some of the feature units in the context

*Corresponding author. Fax: +1 614 688 3984.
E-mail addresses: myung.1@osu.edu (J.I. Myung), montenegro.3@osu.edu (M. Montenegro), pitt.2@osu.edu (M.A. Pitt).

0022-2496/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2007.02.001

vector are "learned", meaning that the corresponding feature values become stabilized or permanent and are stored in memory along with the word. Context vectors can undergo spontaneous forgetting, with the value of each feature unit of the vector decaying from 1 to 0. During recognition, when a test word is presented, the memory retrieval process is activated and the context vector corresponding to the test word is retrieved from memory. Simultaneously, the current experimental context vector is reinstated in memory. If the test word is old (i.e., presented during study phase), the overlap between the retrieved context and the reinstated context should be maximal. If the test word is new (i.e., a distracter), then the retrieved context vector would randomly overlap with the reinstated context vector, which in this case would consist of all the contexts in which the test word has been encountered. On both types of trials, the two context vectors, retrieved and reinstated, are compared against each other, feature by feature, to determine whether the test word should be judged old or new.

The model has five parameters: d, p, r, s, v. The parameter d is the contextual reinstatement parameter, and represents the probability that an originally active feature unit in the experimental context vector fails to get reinstated, so becomes inactive in the recognition process. The parameter p is the context noise parameter, and denotes the probability that any feature element of a context vector is active as a result of previous learning. The learning rate parameter, r, stands for the probability that each new feature in the current context vector is copied successfully (i.e., learned) to the memory image. The sparsity parameter, s, represents the probability that each feature in the context vector is turned on spontaneously, independent of prior learning. Finally, the vector length parameter, v, represents the number of distinctive binary features that make up the context vector.

In the model, the recognition decision of 'old' or 'new' is based on the evaluation of a likelihood ratio defined as the probability of observing the data (i.e., test word) under the hypothesis that it is old to the probability of observing the data under the hypothesis that it is new:

$$\text{Decide "old" if } L = \frac{P(data|old)}{P(data|new)} > 1 \text{ and "new" otherwise.} \quad (1)$$

Given that both the retrieved context vector $\boldsymbol{m} = (m_1, \ldots, m_v)$ and the reinstated context vector $\boldsymbol{c} = (c_1, \ldots, c_v)$ are binary (i.e., $m_i = \{0, 1\}, c_i = \{0, 1\}$), there are four types of match that arise in the comparison between the two vectors. They are denoted by the symbols '00', '01', '10', and '11' where, for example, the '01' match represents the observation of $m_i = 0$ and $c_i = 1$ for the ith feature unit. The likelihood ratio, L, in terms of match types and model parameters is given by (Dennis & Humphreys, 2001, Eq. 6)

$$\begin{aligned} L(\boldsymbol{n} = (n_{00}, n_{01}, n_{10}, n_{11})|\ \theta = (d, p, r, s), v) \\ = \left[\frac{1 - s + ds(1 - r)}{1 - s + ds}\right]^{n_{00}} \left[\frac{p(1 - s) + ds(r + p - rp)}{p(1 - s) + dsp}\right]^{n_{01}} \\ \times (1 - r)^{n_{10}} \left[\frac{r + p - rp}{p}\right]^{n_{11}}, \end{aligned} \quad (2)$$

where n_{00} denotes the total number of 00 matches, and so on. By definition, the sum of the four match counts must be equal to the vector length parameter v, i.e., $n_{00} + n_{01} + n_{10} + n_{11} = v$. Note that $\boldsymbol{n} = (n_{00}, n_{01}, n_{10}, n_{11})$ is a random variable, and so is the likelihood ratio $L(\boldsymbol{n}|\ \theta, v)$.

2. Statistical formulation of BCDMEM

In this section, we provide a statistical treatment of the context noise process assumed in BCDMEM. This allows us to rewrite the model as a statistical one defined as a parametric family of probability distributions, so as to facilitate application of the FT we carry out next.

2.1. Memory retrieval as a multinomial process

Given the assumption that features are independently sampled during recognition, the matching counts, n_{00}, n_{01}, n_{10}, and n_{11}, follow a multinomial probability distribution because they are discrete random variables. Specifically, for a test item that is old (i.e., presented during study), the probability distribution of the matching counts is given by

$$f(\boldsymbol{n}|\boldsymbol{p}(\theta)) = v! \prod_{\lambda} \frac{p_\lambda(\theta)^{n_\lambda}}{n_\lambda!}. \quad (3)$$

In the equation, $\lambda = \{00, 01, 10, 11\}$ is an index variable representing the four types of pattern matches, and $\boldsymbol{p}(\theta) = (p_{00}(\theta), p_{01}(\theta), p_{10}(\theta), p_{11}(\theta))$ is a vector consisting of four multinomial probability parameters, one for each matching type, defined as: $p_{00}(\theta) = (1 - s)(1 - p) + sd(1 - r)(1 - p)$, $p_{01}(\theta) = (1 - s)p + sd(r + p - rp)$, $p_{10}(\theta) = s(1 - d)(1 - r)(1 - p)$, $p_{11}(\theta) = s(1 - d)(r + p - rp)$. Note that $p_{00}(\theta) + p_{01}(\theta) + p_{10}(\theta) + p_{11}(\theta) = 1$ for all θ, by definition, and that $p_\lambda(\theta = (d, p, r, s))$ does not depend upon the vector length parameter v.

Similarly, for a new item that was not presented during the study phase, the matching counts follow another multinomial probability distribution of sample size v,

$$f(\boldsymbol{n}|\boldsymbol{q}(\theta)) = v! \prod_{\lambda} \frac{q_\lambda(\theta)^{n_\lambda}}{n_\lambda!}. \quad (4)$$

In this equation, the multinomial probability vector $\boldsymbol{q}(\theta) = (q_{00}(\theta), q_{01}(\theta), q_{10}(\theta), q_{11}(\theta))$ is defined as: $q_{00}(\theta) = (1 - s(1 - d))(1 - p)$, $q_{01}(\theta) = (1 - s(1 - d))p$, $q_{10}(\theta) = s(1 - d)(1 - p)$, $q_{11}(\theta) = s(1 - d)p$.

Using the multinomial probabilities $\boldsymbol{p}(\theta)$ and $\boldsymbol{q}(\theta)$, the logarithm of the likelihood ratio, called the *loglikelihood* *ratio*, is re-expressed as

$$\log(L(\boldsymbol{n}|\theta, v)) = \sum_{\lambda} \beta_{\lambda}(\theta)\, n_{\lambda}, \tag{5}$$

where $\beta_{\lambda}(\theta) = \log(p_{\lambda}(\theta)/q_{\lambda}(\theta))$. Note that the loglikelihood ratio is a linear combination of multinomial counts n_{λ} weighted by $\beta_{\lambda}(\theta)$.

2.2. Hit and false alarm (FA) rates as expectations

The model makes predictions using two measures of recognition accuracy, hit and FA rates. The hit rate is defined as the proportion of the time that an old item is correctly recognized as old, whereas the FA rate as the proportion of the time that a new item is *incorrectly* recognized as old. Statistically speaking, $P(\text{Hit}|\theta, v)$, the probability of recognizing an old item as old, is an expectation under the multinomial distribution in Eq. (3):

$$\begin{aligned} P(\text{Hit}|\theta, v) &= Prob.[\log(L(\boldsymbol{n}\,|\theta, v)) > 0 | f(\boldsymbol{n}|\boldsymbol{p}(\theta))] \\ &= \sum_{n_{00}+n_{01}+n_{10}+n_{11}=v} S(\log(L(\boldsymbol{n}|\theta, v))) f(\boldsymbol{n}|\boldsymbol{p}(\theta)), \end{aligned} \tag{6}$$

where the summation is over all values of n_{00}, n_{01}, n_{10}, and n_{11} satisfying $n_{00} + n_{01} + n_{10} + n_{11} = v$ and $S(x)$ is the step function defined as $S(x) = 1$ if $x > 0$ and 0 otherwise. In the above equation, the likelihood-based decision rule in (1) has, for mathematical convenience, been translated into an equivalent, loglikelihood based rule since $x > 1 \Longleftrightarrow \log(x) > 0$ for all $x > 0$. Similarly, $P(\text{FA}|\theta, v)$, the probability of recognizing a new item as old, can be written as another expectation under the multinomial distribution in Eq. (4):

$$P(\text{FA}|\theta, v) = \sum_{n_{00}+n_{01}+n_{10}+n_{11}=v} S(\log(L(\boldsymbol{n}|\theta, v))) f(\boldsymbol{n}|\boldsymbol{q}(\theta)). \tag{7}$$

To derive the model's predictions on hit and FA rates given a specific set of parameter values, all that we need to do is to evaluate the two equations (6) and (7). This can, however, be computationally very expensive to calculate because of the combinatorial explosion problem. The number of terms that must be summed rapidly increases as v increases (in proportion to v^3), and thus becomes prohibitively large even for a reasonable value of v.[1] For instance, for $v = 200$, there are over one million terms to sum! In short, a direct evaluation of the multinomial sums is generally not possible in practice unless v is trivially small (e.g., 5–10).

Faced with this obstacle, the modeler must rely on a less-than-satisfactory shortcut: brute-force simulations of the model are run in which for each chosen set of parameter values, the entire process of the model is simulated on computer using random number generators. This is obviously an extremely time-consuming procedure such that finding parameter values that best fit observed data can be next to impossible.

3. Integral expressions for recognition probabilities

With the statistical formulation of BCDMEM in hand, we are ready to apply the FT technique and show how it enables one to derive easy-to-compute integral expressions for hit and FA rates, thus freeing the modeler from cumbersome and restrictive simulation methods.

The FT, a mathematical tool widely used in many scientific and engineering applications, is an integral transformation technique that decomposes a function into an infinite sum (an integral) of its complex-valued frequency components with coefficients determined specific to the function (see, e.g., Arfken & Weber, 2001; Boas, 1983).

3.1. Fourier integral expressions for hit and FA rates

Using the FT technique, the step function $S(x)$ defined in Eq. (6) can be equivalently written as

$$S(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{ikx}}{ik}\, dk, \tag{8}$$

where $e^{ikx} \equiv \cos(kx) + i \sin(kx)$ is a complex number consisting of the real and imaginary parts, with the latter containing the mathematical symbol $i = \sqrt{-1}$ or equivalently $i^2 = -1$.

By replacing $S(x)$ in (6) with its FT in (8) and the loglikelihood ratio with the expression in (5), we rewrite $P(\text{Hit}|\theta, v)$ as

$$\begin{aligned} P(\text{Hit}|\theta, v) &= \sum_{n_{00}+n_{01}+n_{10}+n_{11}=v} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{ik\, \log(L(\boldsymbol{n}|\theta, v))}}{ik}\, dk \right] f(\boldsymbol{n}|\boldsymbol{p}(\theta)) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{dk}{ik} \left[\sum_{n_{00}+n_{01}+n_{10}+n_{11}=v} e^{ik \sum_{\lambda} \beta_{\lambda}(\theta) n_{\lambda}}\, v! \prod_{\lambda} \frac{p_{\lambda}(\theta)^{n_{\lambda}}}{n_{\lambda}!} \right]. \end{aligned} \tag{9}$$

The sum in the hard bracket [...] reduces to

$$\text{Sum} = \sum_{n_{00}+n_{01}+n_{10}+n_{11}=v} v! \prod_{\lambda} \frac{(e^{ik\beta_{\lambda}} p_{\lambda}(\theta))^{n_{\lambda}}}{n_{\lambda}!}$$

using the relation $e^{ik \sum_{\lambda} \beta_{\lambda}(\theta) n_{\lambda}} = \prod_{\lambda} (e^{ik\beta_{\lambda}(\theta)})^{n_{\lambda}}$. We then notice that this sum corresponds to the multinomial expansion

$$\text{Sum} = \left(\sum_{\lambda} e^{ik\beta_{\lambda}}\, p_{\lambda}(\theta) \right)^{v}.$$

Finally, plugging this result into Eq. (9), we obtain the following one-dimensional integral expressions for hit rate,

[1] The total number of terms to sum in (6) is obtained by evaluating the sum $(\sum_{n_{11}=0}^{v} \sum_{n_{10}=0}^{v-n_{11}} \sum_{n_{01}=0}^{v-n_{11}-n_{10}} 1)$ with the help of an algebraic computing package such as *Mathematica* or using Faulhaber's Formula (Gradshteĭn & Ryzhik, 1990). The number obtained is equal to $(v^3 + 6v^2 + 11v + 6)/6$, that can be approximated to $v^3/6$ for large v.

and a similar expression for FA rate, as follows:

$$P(\mathrm{Hit}|\theta, v) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum_{\lambda} e^{ik\beta_{\lambda}(\theta)}p_{\lambda}(\theta)\right)^{v}\frac{dk}{ik},$$

$$P(\mathrm{FA}|\theta, v) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum\nolimits_{\lambda} e^{ik\beta_{\lambda}(\theta)}q_{\lambda}(\theta)\right)^{v}\frac{dk}{ik}. \quad (10)$$

The above FT integrals can be easily evaluated numerically. To understand why, note that the integral for $P(\mathrm{Hit}|\theta, v)$ is a line integral along the real axis in the complex plane that contains the singularity point at $k = 0$. From the Residue Theorem of complex analysis (Arfken & Weber, 2001), the integration around the singularity yields the value of $\frac{1}{2}$, and the integration along the rest of the real axis can be calculated just for the real part of the integrand function because the real part of the function is even around $k = 0$ (i.e., $f(k) = f(-k)$). Also, because the imaginary part of the function is odd, its integral along the real axis cancels out. In short, the FT integral for the hit rate is rewritten as

$$P(\mathrm{Hit}|\theta, v) = \frac{1}{2} + \frac{1}{\pi}\int_{0+}^{\infty}\Re\left[\frac{1}{ik}\left(\sum_{\lambda} e^{ik\beta_{\lambda}(\theta)}p_{\lambda}(\theta)\right)^{v}\right]dk, \quad (11)$$

where $\Re[z]$ is the real part of a complex number z and 0+ is an infinitesimally small positive number.[2] It is straightforward to evaluate numerically the above expression using *Mathematica* or *Matlab*.

We verified the correctness of the FT results by comparing them with simulation data of Dennis and Humphreys (2001). In one simulation study, they evaluated BCDMEM's ability to reproduce the data of Glanzer and Adams (1990), who first had participants perform a lexical decision task using low- and high-frequency words presented auditorily or visually. This was followed by a standard old–new recognition memory task. The empirical data from the four conditions are in the second column of Table 1. To the right are the best-fit predictions from Dennis and Humphreys (2001) simulations, which are quite good. To validate that our implementation of BCDMEM was correct, we ran the same simulation using the same parameter values as Dennis and Humphreys (2001). The data, shown in the fourth column, are very similar. Next we evaluated the FT integral expression in (10) to determine if equivalent results would be obtained. As shown in the rightmost column, this is in fact the case.

A second test was performed to verify further the close correspondence between the two forms of the model (simulation vs FT integration). Three parameters d, p, s were fixed at values different from the preceding simulation. The learning rate parameter, r, was varied across its range and model predictions were compared across this range. As r increases, discriminability of old from new items should increase, causing the hit and FA rates to diverge from one another. The results are shown in Fig. 1, with the circles representing the simulation predictions and the lines the FT predictions. As is clear, the values from the analytic expression are in almost perfect agreement with the simulation values. The results from this and the preceding test demonstrate the accuracy and correctness of the FT integral.

[2]The second term of the right-hand side of the equation is evaluated in practice by approximating the term as $\lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{\varepsilon}^{\infty}\Re[\ldots]dk$, provided that the result converges to a finite value.

Table 1

	G & A (1990) data	D & H (2001) simulation	Present simulation	FT integration
$\mathrm{Hit_L}$	0.7437	0.744±0.024	0.7371	0.7338
$\mathrm{Hit_H}$	0.6785	0.687±0.022	0.6793	0.6792
$\mathrm{FA_L}$	0.2172	0.189±0.022	0.1818	0.1825
$\mathrm{FA_H}$	0.2804	0.299±0.025	0.2935	0.2943

The second column contains observed data from Glanzer and Adams (G & A, 1990), obtained under high (H) and low (L) word-frequency conditions. The third column shows best-fit predictions of BCDMEM as reported in Dennis and Humphreys (D & H, 2001). Parameter values for their simulation were $d = 0.590$, $p_L = 0.094$, $p_H = 0.336$, $r = 0.602$, $s = 0.02$, and $v = 200$. The probability values shown in the third column are obtained by digitally scanning the figure image in Dennis and Humphreys (2001, Fig. 9A). The next two columns show our own results obtained by direct simulation and exact evaluation of the Fourier transformation (FT) integral expressions in (10).

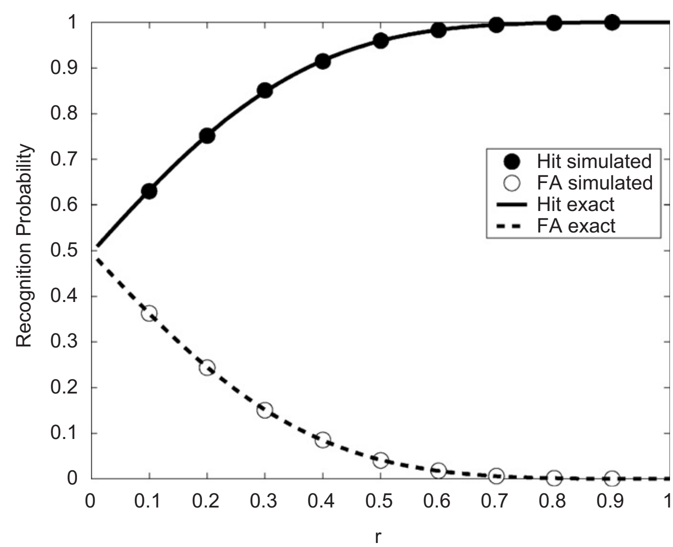

Fig. 1. Comparison between simulated and exact FT results. Solid and open circles represent simulation values of hit and FA rates, respectively, averages based on 5000 independent replications. Solid and dotted curves are exact values obtained by analytically evaluating the FT integral expressions in (10). The parameter values used in both cases were $d = 0.20$, $p = 0.40$, $s = 0.20$, and $v = 200$.

3.2. Asymptotic expressions of Fourier integrals

Although the FT integrals in (10) are readily computable, it is useful to seek their asymptotic expressions to identify fundamental properties of BCDMEM. To derive an asymptotic expression for $P(\mathrm{Hit}|\theta, v)$, we expand the second term on the right-hand side of Eq. (11) for large values of the vector length parameter, v. By plotting the integrand function in the term as a function of k, we observe that the function concentrates its mass around $k = 0$ for large v. The first step of asymptotic expansion is therefore to expand this function around $k = 0$. To do this, we successively apply the two Taylor series expansions of the exponential and logarithmic functions, $e^x \approx 1 + x + \frac{x^2}{2}$ and $\log(1 + x) \approx x - \frac{x^2}{2}$, to obtain the following approximation:

$$\begin{aligned}
&\log\left(\sum_\lambda e^{ik\beta_\lambda(\theta)} p_\lambda(\theta)\right) \\
&\approx \log\left(1 + ik\sum_\lambda \beta_\lambda(\theta)p_\lambda(\theta) - \frac{k^2}{2}\sum_\lambda \beta_\lambda^2(\theta)p_\lambda(\theta)\right) \\
&\approx \left(ik\sum_\lambda \beta_\lambda(\theta)p_\lambda(\theta) - \frac{k^2}{2}\sum_\lambda \beta_\lambda^2(\theta)p_\lambda(\theta)\right) \\
&\quad - \frac{1}{2}\left(ik\sum_\lambda \beta_\lambda(\theta)p_\lambda(\theta) - \frac{k^2}{2}\sum_\lambda \beta_\lambda^2(\theta)p_\lambda(\theta)\right)^2 \\
&= i\mu_\mathrm{p}(\theta)k - \frac{\sigma_\mathrm{p}^2(\theta)}{2}k^2 + O(k^3), \qquad (12)
\end{aligned}$$

where $\mu_\mathrm{p}(\theta)$ and $\sigma_\mathrm{p}^2(\theta)$ are the mean and variance of the single-trial "payoff" $\beta(\theta)$ defined as $\mu_\mathrm{p}(\theta) = \sum_\lambda \beta_\lambda(\theta)p_\lambda(\theta)$ and $\sigma_\mathrm{p}^2(\theta) = \sum_\lambda \beta_\lambda^2(\theta)p_\lambda(\theta) - \mu_\mathrm{p}^2(\theta)$. From the above approximation and remembering the identity $z^v = \exp(v\log(z))$, we obtain

$$\begin{aligned}
\left(\sum_\lambda e^{ik\beta_\lambda}p_\lambda\right)^v &= \exp\left(v\log\left(\sum_\lambda e^{ik\beta_\lambda}p_\lambda\right)\right) \\
&\approx \exp\left(iv\mu_\mathrm{p}k - \frac{v\sigma_\mathrm{p}^2}{2}k^2\right), \qquad (13)
\end{aligned}$$

where, for simplicity, we employed the shorthand symbols β_λ, p_λ, μ_p, and σ_p^2 by dropping the explicit θ in the argument. This convention will be used in the remainder of this section, whenever convenient.

Now, using the approximation in (13), the second term on the right-hand side of (11) is rewritten as

$$\begin{aligned}
&\frac{1}{\pi}\int_{0+}^{\infty} \Re\left[\frac{1}{ik}\left(\sum_\lambda e^{ik\beta_\lambda(\theta)}p_\lambda(\theta)\right)^v\right]dk \\
&\approx \frac{1}{\pi}\int_{0+}^{\infty} \Re\left[\frac{1}{ik}\exp\left(iv\mu_\mathrm{p}k - \frac{v\sigma_\mathrm{p}^2}{2}k^2\right)\right]dk \\
&= \frac{1}{\pi}\int_{0+}^{\infty}\left(\frac{1}{2}\int_{-v\mu_\mathrm{p}}^{v\mu_\mathrm{p}}\cos(xk)\,dx\right)\exp\left(-\frac{v\sigma_\mathrm{p}^2}{2}k^2\right)dk \\
&= \mathrm{ncdf}\left(\frac{\mu_\mathrm{p}}{\sigma_\mathrm{p}/\sqrt{v}}\right) - \frac{1}{2}, \qquad (14)
\end{aligned}$$

where $\mathrm{ncdf}(x) \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{y^2}{2}\right)dy$ is the unit normal cumulative distribution function. By plugging the above result into (11), we obtain the desired asymptotic hit rate as

$$P(\mathrm{Hit}|\theta, v) \approx \mathrm{ncdf}\left(\frac{\mu_\mathrm{p}(\theta)}{\sigma_\mathrm{p}(\theta)/\sqrt{v}}\right) \qquad (15)$$

for large v. Similarly, the asymptotic FA rate for large v is obtained as

$$P(\mathrm{FA}|\theta, v) \approx \mathrm{ncdf}\left(\frac{\mu_q(\theta)}{\sigma_q(\theta)/\sqrt{v}}\right), \qquad (16)$$

where $\mu_q(\theta) = \sum_\lambda \beta_\lambda(\theta)q_\lambda(\theta)$ and $\sigma_q^2(\theta) = \sum_\lambda \beta_\lambda^2(\theta)q_\lambda(\theta) - \mu_q^2(\theta)$.[3]

Three properties of BCDMEM directly fall out of these asymptotic expressions. First, the mean $\mu_\mathrm{p}(\theta)$ turns out to be equal to the Kullback–Leibler distance measure between two distributions in information theory (Cover & Thomas, 1991):

$$\mu_\mathrm{p}(\theta) = \sum_\lambda p_\lambda(\theta)\log\left(\frac{p_\lambda(\theta)}{q_\lambda(\theta)}\right) = D(p||q) \geqslant 0$$

and similarly $\mu_q(\theta) = -D(q||p) \leqslant 0$. The Kullback–Leibler distance is not symmetric, i.e., $D(p||q) \neq D(q||p)$, unless $p = q$. One can also easily show the inequality $\mu_q(\theta) \leqslant 0 \leqslant \mu_\mathrm{p}(\theta)$. These observations together imply that the hit rate will always be equal to or greater than the FA rate; that is, $P(\mathrm{FA}|\theta, v) \leqslant 0.5 \leqslant P(\mathrm{Hit}|\theta, v)$ for all θ and v, at least asymptotically.

Second, the hit and FA rates that the model predicts depend on the value of the vector length parameter, v. In particular, as v is increased to infinity, hit and FA rates converge to 1 and 0, respectively, regardless of the values of the other four parameters $\theta = (d, p, r, s)$. Formally, we observe

$$P(\mathrm{FA}|\theta, v) \to 0 \quad \text{and} \quad P(\mathrm{Hit}|\theta, v) \to 1 \quad \text{as } v \to \infty$$

for all θ.

Third, related to the second point, inspection of the asymptotic expressions reveals a trade-off between the parameter vector $\theta = (d, p, r, s)$ and the vector length parameter, v. The same hit rate is predicted for two different choices of model parameters, (θ_1, v_1) and (θ_2, v_2), as long as they are chosen such that the value of the argument of the cumulative normal distribution function $\mathrm{ncdf}(\cdot)$ remains unchanged:

$$\frac{\mu_\mathrm{p}(\theta_1)}{\sigma_\mathrm{p}(\theta_1)/\sqrt{v_1}} = \frac{\mu_\mathrm{p}(\theta_2)}{\sigma_\mathrm{p}(\theta_2)/\sqrt{v_2}}.$$

The same can be said about the model's prediction of FA rate. An implication of this behavior for BCDMEM is that

[3]Interestingly, the same asymptotic expressions as in (15) and (16) can also be obtained from the multivariate normal approximation of the multinomial distribution. That is, the normal approximation of the multinomial distribution implies that the sampling distribution of the loglikelihood ratio $\log(L(\boldsymbol{n}|\theta, v))$ is approximately normally distributed for large v. The derivation of the asymptotic approximations for hit and FA rates in this manner is shown in the Appendix.

the model is unidentifiable to the extent that different sets of parameter values for which the above equality holds can be found. When the model becomes unidentifiable, one cannot meaningfully interpret the parameter values, as there exist multiple sets of parameter values that fit the data equally well.

Another benefit of the asymptotic expressions for $P(\mathrm{Hit}|\theta, v)$ and $P(\mathrm{FA}|\theta, v)$ is that one can derive an explicit equation describing the receiver operating curve (ROC) in signal detection theory (Green & Swets, 1974). To show this, first note that the asymptotic expressions are obtained under a fixed decision criterion: Decide "old" if $\log L(\boldsymbol{n}|\theta, v) > 0$ and "new" otherwise. Under a more general decision rule assuming a variable criterion c instead of $c = 0$, the asymptotic expressions in (15) and (16) are then modified to

$$P(\mathrm{Hit}|\theta, v, c) \approx \mathrm{ncdf}\left(\frac{\mu_{\mathrm{p}}(\theta) - c}{\sigma_{\mathrm{p}}(\theta)/\sqrt{v}}\right),$$
$$P(\mathrm{FA}|\theta, v, c) \approx \mathrm{ncdf}\left(\frac{\mu_q(\theta) - c}{\sigma_q(\theta)/\sqrt{v}}\right)$$

for $-\infty < c < \infty$. From this, the ROC plotted on Gaussian axes (z_{Hit} vs z_{FA}) is obtained as

$$z_{\mathrm{Hit}}(\theta, v, c) = \left(\frac{\sigma_q(\theta)}{\sigma_{\mathrm{p}}(\theta)}\right) z_{\mathrm{FA}}(\theta, v, c) + \left(\frac{\mu_{\mathrm{p}}(\theta) - \mu_q(\theta)}{\sigma_{\mathrm{p}}(\theta)/\sqrt{v}}\right) \tag{17}$$

which is known as the *z-ROC*.

4. Discussion

Application of the FT technique makes BCDMEM not only more user-friendly but also more understandable. No longer are extended simulations required to generate reliable predictions. Rather, the analytically exact integral expression enables one to quickly compute model predictions in the form of hit and FA rates. To perform this analysis, we first recast BCDMEM as a multinomial process, characterized by a multinomial probability distribution with hit and FA rates as expectations under that distribution. To our knowledge, this is the first such statistical treatment of the model.

By taking the analysis one step further and deriving the asymptotic expressions of hit and FA rates for large v, we were able to glean previously unknown properties of the model. Most importantly, the context vector length, v, turns out to be a non-ignorable parameter that can significantly affect model predictions. The analysis also showed that the model is unidentifiable, in the sense that multiple sets of parameter values can provide the same fit to a data set. At least one parameter such as v must be fixed to avoid this.[4]

An obvious next question to ask in this line of work is whether one can utilize the asymptotic expressions to address other substantive issues of current interest to memory researchers. For example, it would be interesting if one could show that BCDMEM can predict a reversal or elimination of the low-frequency hit-rate advantage that recent experiments have found (e.g., Criss & Shiffrin, 2004). Another possible analysis, suggested by a reviewer, is to determine whether the model can exhibit the list strength effect (Ratcliff, Clark, & Shiffrin, 1990; Ratcliff, Shen, & Gronlund, 1992). More broadly, one can also ask whether the FT technique can be applied to other simulation-based models (e.g., REM, SAM, MINERVA), and, of course, whether asymptotically they exhibit similar properties as BCDMEM. Unique challenges are faced in applying the technique to each of these models. Work is currently under way that will address many of the above issues.

[4]Because Dennis and Humphreys (2001) always fixed the vector length parameter v in their simulation analyses, they may well have suspected that BCDMEM is unidentifiable when v varies.

Acknowledgments

This work was supported by National Institute of Health Grant R01-MH57472 to J.I.M. and M.A.P. We wish to thank Simon Dennis, Richard Shiffrin, Roger Ratcliff, Mark Steyvers, and an anonymous reviewer for valuable feedback on earlier versions of this paper.

Appendix A. Sampling distribution of loglikelihood ratio

In this Appendix we show, using the normal approximation of the multinomial distribution, that the sampling distribution of the loglikelihood ratio $\log(L(\boldsymbol{n}|\theta, v))$ is approximately normal for large v. From this, we then derive the asymptotic expressions for the Fourier integral in (15).

According to Eq. (5), the loglikelihood ratio, which itself is a random variable, is written as

$$\log(L(\boldsymbol{n}|\theta, v)) = \sum_{\lambda} \beta_{\lambda}(\theta)\, n_{\lambda},$$

where $\lambda = \{00, 01, 10, 11\}$. Suppose that n_{λ} represents the pattern matching count given a target item that is old, and thus it follows the multinomial distribution $f(\boldsymbol{n}|\boldsymbol{p}(\theta))$. The mean of the loglikelihood ratio under this distribution is given by

$$\begin{aligned} E[\log(L(\boldsymbol{n}|\theta, v))] &= \sum_{\lambda} \beta_{\lambda}(\theta)\, E(n_{\lambda}) \\ &= \sum_{\lambda} \beta_{\lambda}(\theta)\, v\, p_{\lambda}(\theta) \\ &= v\mu_{\mathrm{p}}(\theta), \end{aligned} \tag{18}$$

where $\mu_{\mathrm{p}}(\theta)$ is defined in the main text. Next, the variance of the loglikelihood ratio is obtained as

$$\begin{aligned} &Var[\log(L(\boldsymbol{n}|\theta, v))] \\ &\quad = Var\left[\sum_{\lambda} \beta_{\lambda}\, n_{\lambda}\right] \\ &\quad = \sum_{\lambda} \beta_{\lambda}^2\, Var(n_{\lambda}) + 2 \sum_{\lambda \neq \kappa} \beta_{\lambda}\beta_{\kappa}\, Cov(n_{\lambda}, n_{\kappa}) \\ &\quad = \sum_{\lambda} \beta_{\lambda}^2 v\, p_{\lambda}(1 - p_{\lambda}) + 2 \sum_{\lambda \neq \kappa} \beta_{\lambda}\beta_{\kappa}\, (-v) p_{\lambda} p_{\kappa} \end{aligned}$$

$$= v\sum_{\lambda}\beta_{\lambda}^2 p_{\lambda} - v\left(\sum_{\lambda}\beta_{\lambda}p_{\lambda}\right)^2$$
$$= v\sigma_{\mathrm{p}}^2(\theta), \tag{19}$$

where $\sigma_{\mathrm{p}}^2(\theta)$ is defined in the main text. In the above equation, for simplicity, we employed the shorthand symbols β_{λ} and p_{λ} for $\beta_{\lambda}(\theta)$ and $p_{\lambda}(\theta)$, respectively.

Next, the multivariate normal approximation of the multinomial distribution implies that each n_{λ} would be approximately normally distributed for large v. Given that the loglikelihood ratio is a linear sum of normals (i.e., n_{λ}'s) as shown in (5), its sampling distribution would also be approximately normal. The *linearity* of the loglikelihood equation in n_{λ} is crucial for the normal approximation to work. Otherwise, the approximation is not justified.

Finally, putting together the above results and observation, we can now rewrite the probability of the loglikelihood ratio exceeding 0 (i.e., $P(\mathrm{Hit}|\theta, v)$) in terms of a unit normal random variable z as

$$Prob.[\log(L(\boldsymbol{n}|\theta, v)) > 0 | f(\boldsymbol{n}|\boldsymbol{p}(\theta))]$$
$$= P\left(z > \frac{0 - v\,\mu_{\mathrm{p}}(\theta)}{\sqrt{v\,\sigma_{\mathrm{p}}^2(\theta)}}\right)$$
$$= P\left(z < \frac{\mu_{\mathrm{p}}(\theta)}{\sigma_{\mathrm{p}}(\theta)/\sqrt{v}}\right)$$
$$= \mathrm{ncdf}\left(\frac{\mu_{\mathrm{p}}(\theta)}{\sigma_{\mathrm{p}}(\theta)/\sqrt{v}}\right) \tag{20}$$

yielding the desired result. The asymptotic expression for $P(\mathrm{FA}|\theta, v)$ can be derived in a similar fashion.

References

Arfken, G. B., & Weber, H. J. (2001). In H. J. Weber (Ed.), *Mathematical methods for physicists* (5th ed.). San Diego: Academic Press.

Boas, M. L. (1983). *Mathematical methods in the physical sciences* (2nd ed.). New York: Wiley.

Cover, T., & Thomas, J. (1991). *Elements of information theory*. New York: Wiley.

Criss, A. H., & Shiffrin, R. M. (2004). Interactions between study task, study time, and the low-frequency hit rate advantage in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 778–786.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452–478.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 5–16.

Gradshteĭn, I. S., & Ryzhik, I. M. (2000). In A. Jeffrey (Ed.), *Tablîŝy integralov, summ, r̂iadov i proizvedeniĩ* (Table of integrals, series, and products) (6th ed.) (Scripta Technica, Inc., Trans.). San Diego: Academic Press.

Green, D., & Swets, J. (1974). *Signal detection theory and psychophysics*. New York: Krieger.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*(4), 724–760.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). The list-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 163–178.

Ratcliff, R., Shen, C.-F., & Gronlund, S. D. (1992). Testing global memory models using roc curves. *Psychological Review*, *99*, 518–535.

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

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 179–195.