Notes on EBRW Extension

Takao Noguchi
Department of Experimental Psychology, University College London

Abstract

Here I write down description and math for our EBRW extension. Bits and pieces from this document may appear in the later manuscript.

Preliminary

In this note, I consider a binary classification problem: whether to classify Item i into Category $\mathbb{A}$ or $\mathbb{B}$. This problem is approached as evidence accumulation processes: at each time step, evidence is accumulated toward responding as $\mathbb{A}$ or $\mathbb{B}$, and as soon as the accumulated evidence reaches the boundary, θ_A or θ_B, a response is made.

The math derivations below are informed by Chapters 2 and 4 of Kijima (1997) and Chapters 2 and 7 of Bhat and Miller (2002).

Notation

Free parameters are indicated with Greek letters (e.g., ρ), and a matrix and a vector are indicated with English capital letters (e.g., Q), whose element is denoted as a corresponding small letters (e.g., q). The element at ith row, jth column of Q matrix, for example, is indicated as q_{ij}.

Model Specifications

GCM

According to Nosofsky (1986)'s *generalized context model* (GCM), people represent categories by storing individual exemplars in memory. Classification decisions are based on summing the similarity of an object to the exemplars of the alternative categories. In the GCM, exemplars are represented as points in a multidimensional psychological space, and similarity between exemplars is a decreasing function of their distance in the space.

Assume the exemplars reside in an multidimensional psychological space, and let x_{im} denote the value of exemplar i on psychological dimension m. The distance between

Last updated: November 6, 2015.

exemplars i and j is given by

$$d_{ij} = \left(\sum_m \omega_m |x_{im} - x_{jm}|^\rho \right)^{1/\rho}, \tag{1}$$

where ω_m $(0 \leq \omega_m, \sum \omega_m = 1)$ represents the attention weight given to dimension m. The x_{im} psychological coordinate values for the exemplars are generally derived by conducting multidimensional scaling studies or else are assumed to be given by the physical coordinate values used for constructing the stimuli. The attention weights (ω_m) are free parameters in the model.

The distances (d_{ij}) are transformed to similarity measures (s_{ij}) by using an exponential decay function:

$$s_{ij} = \exp(-\gamma\, d_{ij}), \tag{2}$$

where γ is an overall sensitivity parameter for scaling distances in the space.

Because of factors such as recency of presentation, exemplars may reside in memory with differing strengths. Let η_j denote the memory strength for exemplar j. The degree to which exemplar j is activated when presented with item i is determined jointly by the exemplar's strength in memory and by its similarity to item i. Specifically, the activation for exemplar j given presentation of item i (a_{ij}) is given by

$$a_{ij} = \eta_j\, s_{ij}. \tag{3}$$

Finally, the evidence for Category $\mathbb{A}$ given presentation of item i is found by summing the activations for all stored exemplars of Category $\mathbb{A}$. The conditional probability with which item i is classified into Category $\mathbb{A}$ is found by dividing this evidence by the summed evidence for all the categories:

$$p(\mathbb{A}) = \frac{\sum_{j \in \mathbb{A}} a_{ij}}{\sum_{j \in \mathbb{A}} a_{ij} + \sum_{j \in \mathbb{B}} a_{ij}} \tag{4}$$

The GCM does not provides an account of how the exemplar-based similarity comparison process unfolds over time. Such an account is provided by EBRW.

EBRW

In the *exemplar-based random walk* (EBRW; Nosofsky & Palmeri, 1997), there is a random walk counter that accrues information pointing to either Category $\mathbb{A}$ or $\mathbb{B}$. The counter has a starting value of zero; positive increments move it in the direction of Category $\mathbb{A}$, and negative increments move it in the direction of Category $\mathbb{B}$. The observer establishes criteria representing the amount of evidence that is needed to execute either a Category A response (θ_A) or a Category B response $(-\theta_B)$. Once the counter reaches either of these criteria, the appropriate categorization response is made.

When Item i is presented, it sets off a race among all the exemplars stored in memory. For simplicity in getting started, and for reasons of analytic tractability, we assume that the race times are exponentially distributed. Thus, the expected time that Exemplar j

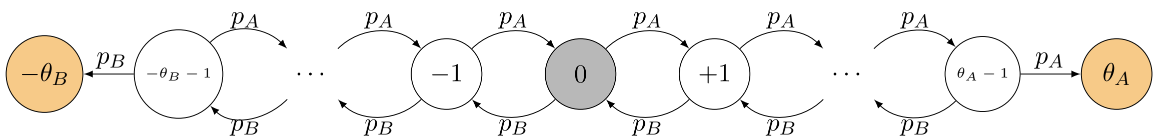

Figure 1. States in the EBRW model. A gray node is a starting state, and yellow nodes indicate absorbing states.

completes its race is given by $\frac{1}{a_{ij}}$, where the activation value (a_{ij}) is computed as in the GCM.

Also from the GCM, when Item i is presented, the probability that Exemplar j is retrieved is given by

$$p(j|i) = \frac{a_{ij}}{\sum_{j\in\mathbb{A}} a_{ij} + \sum_{j\in\mathbb{B}} a_{ij}.} \tag{5}$$

Original formulation (discrete time). The evidence accumulation in the EBRW can be formulated as Markov chain (see Figure 1). This Markov chain has two absorbing states (yellow nodes in Figure 1) and $\theta_A + \theta_B - 1$ transient states (white and gray nodes).

First, we calculate the probability of taking a step toward Category $\mathbb{A}$ by summing the probabilities that any one of the exemplars from Category $\mathbb{A}$ is retrieved, yielding

$$p_A = \frac{\sum_{j\in\mathbb{A}} a_{ij}}{\sum_{j\in\mathbb{A}} a_{ij} + \sum_{j\in\mathbb{B}} a_{ij}} \tag{6}$$

Then, the transition matrix for this Markov chain is written with four sub-matrices:

$$P = \left[\begin{array}{c|c} I & 0 \\ \hline R & Q \end{array}\right] \tag{7}$$

$$
= \begin{array}{cl}
\text{state} & \begin{array}{cc|ccccccccccc} \theta_A & -\theta_B & -\theta_B{-}1 & -\theta_B{-}2 & \cdots & -2 & -1 & 0 & 1 & 2 & \cdots & \theta_A{-}2 & \theta_A{-}1 \end{array} \\
\begin{array}{c} \theta_A \\ \theta_B \\ -\theta_B{-}1 \\ -\theta_B{-}2 \\ \vdots \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ \vdots \\ \theta_A{-}2 \\ \theta_A{-}1 \end{array} &
\left[\begin{array}{cc|ccccccccccc}
1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\hline
0 & p_B & 0 & p_A & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & p_B & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
 & & & & & & & & & & & & \\
0 & 0 & 0 & 0 & \cdots & 0 & p_A & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & \cdots & p_B & 0 & p_A & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & p_B & 0 & p_A & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & p_B & 0 & p_A & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & p_B & 0 & \cdots & 0 & 0 \\
 & & & & & & & & & & & & \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & p_A \\
p_A & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & p_B & 0
\end{array}\right]
\end{array}. \tag{8}
$$

A cell located at row k column j of this P matrix represents the probability of going from state k to state j (note each row in P matrix sums to 1). The size of P matrix depends on the number of transient states: R is a matrix of size $(\theta_A + \theta_B - 1) \times 2$, and Q is of size $(\theta_A + \theta_B - 1) \times (\theta_A + \theta_B - 1)$. Here, R is a transition matrix from the transient states to the absorbing states, and Q is a transition matrix within the transient states. In particular, Q has p_A at row k column $k+1$ and p_B at row k column $k-1$, and the other cells in Q are all 0.

Also, we construct a vector, labeled Z, representing the starting state of the evidence accumulation:

$$Z = \begin{array}{c} \begin{matrix} \theta_A & -\theta_B & -\theta_B-1 & -\theta_B-2 & \cdots & -2 & -1 & 0 & 1 & 2 & \cdots & \theta_A-2 & \theta_A-1 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \end{array}. \tag{9}$$

The derivation below uses the fundamental matrix (M) of this Markov chain, which is given by

$$M = \left(\sum_{t=1}^{\infty} Q^t \right) \tag{10}$$

$$= (I - Q)^{-1}. \tag{11}$$

Each element in this M matrix is the expected number of visit to each state: the value at the kth row jth column (m_{kj}) is the expected number of transitions to state j, when the chain started at state k.

Response probability. Then, the probability of classifying Item i into each category can be computed as follows.

$$[p(\mathbb{A}),\ p(\mathbb{B})] = Z \ \left(\sum_{t=1}^{\infty} Q^t \right) \ R \tag{12}$$

$$= Z \ M \ R. \tag{13}$$

Alternative response probability. Nosofsky and Palmeri (1997) derived the response probability as follows:

$$p(\mathbb{A}) = \begin{cases} \dfrac{1 - (p_B/p_A)^{\theta_B}}{1 - (p_B/p_A)^{\theta_A + \theta_B}} & \text{if } p_A \neq p_B \\ \\ \dfrac{\theta_B}{\theta_A + \theta_B} & \text{if } p_A = p_B. \end{cases} \tag{14}$$

Proof is in Chapter 16 of the book by Feller (1968), but I haven't seen it yet.

Response time. Before reaching the absorbing states, several exemplars are retrieved. Here, we derive the expected number of exemplars retrieved by summing elements of M row-wise to derive the total number of exemplars retrieved for each starting state:

$$M_p = M \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \tag{15}$$

and then, the expectation is given by

$$\mathcal{E}(n) = Z\ M_p. \tag{16}$$

To compute its variance, we first square each element in M_p

$$M_{p2} = diag(M_p)\ M_p, \tag{17}$$

and then the variance is given by

$$\mathcal{V}(n) = Z\ (2\,M - I)\ M_p - M_{p2}. \tag{18}$$

The expected response time is given by multiplying the expected time to accumulate one piece of evidence with the expected number of times evidence is accumulated. Given the property of exponential distribution, the expected time that at least one exemplar completes its race and accumulate evidence is

$$\mathcal{E}(\Delta t) = \alpha + \frac{1}{\sum_{j\in\mathbb{A}} a_{ij} + \sum_{j\in\mathbb{B}} a_{ij},} \tag{19}$$

where α is a step-time constant. Then,

$$\mathcal{E}(t) = \mathcal{E}(\Delta t)\ \mathcal{E}(n). \tag{20}$$

<u>*Alternative response time.*</u> Nosofsky and Palmeri (1997) derived the response time as follows:

$$\mathcal{E}(n) = \begin{cases} \dfrac{\theta_B}{p_B - p_A} - \dfrac{\theta_A + \theta_B}{p_B - p_A}\left[\dfrac{1-(p_B/p_A)^{\theta_B}}{1-(p_B/p_A)^{\theta_A+\theta_B}}\right] & \text{if } p_A \neq p_B \\ \theta_A\,\theta_B & \text{if } p_A = p_B. \end{cases} \tag{21}$$

Proof is in Chapter 16 of the book by Feller (1968), but I haven't seen it yet.

Nosofsky and Palmeri (1997) also derived the response time conditioned on a response: the response time given a category response of $\mathbb{A}$ is

$$\mathcal{E}(n|\mathbb{A}) = \begin{cases} \dfrac{1}{p_A - p_B}\left[\dfrac{(p_A/p_B)^{\theta_A+\theta_B}+1}{(p_A/p_B)^{\theta_A+\theta_B}-1}(\theta_A+\theta_B) - \dfrac{(p_A/p_B)^{\theta_B}+1}{(p_A/p_B)^{\theta_B}-1}\theta_B\right] & \text{if } p_A \neq p_B \\ \dfrac{\theta_A}{3}\,(2\,\theta_B + \theta_A) & \text{if } p_A = p_B. \end{cases} \tag{22}$$

Similarly,

$$\mathcal{E}(n|\mathbb{B}) = \begin{cases} \dfrac{1}{p_B - p_A}\left[\dfrac{(p_A/p_B)^{-\theta_A+\theta_B}+1}{(p_A/p_B)^{-\theta_A-\theta_B}-1}(\theta_A+\theta_B) - \dfrac{(p_A/p_B)^{-\theta_A}+1}{(p_A/p_B)^{-\theta_A}-1}\theta_B\right] & \text{if } p_A \neq p_B \\ \dfrac{\theta_B}{3}\,(2\,\theta_A + \theta_B) & \text{if } p_A = p_B. \end{cases} \tag{23}$$

Proof is in Busemeyer (1982), but I haven't checked it yet.

Alternative response time 2. Diederich and Busemeyer (2003) used an alternative derivation to compute the expected number of exemplars, conditioned on a response. Proof is supposedly in Diederich (1995), but I haven't found it yet.

$$[\mathcal{E}(n|\mathbb{A}),\ \mathcal{E}(n|\mathbb{B})] = \left(\sum_{t=1}^{\infty} t\, Z\, Q^t\ R\right) \begin{bmatrix} p(A|i) & 0 \\ 0 & p(B|i) \end{bmatrix}^{-1} \tag{24}$$

$$= Z\ \left(\sum_{t=1}^{\infty} t\, Q^t\right)\ R \begin{bmatrix} p(A|i) & 0 \\ 0 & p(B|i) \end{bmatrix}^{-1} \tag{25}$$

$$= Z\ (I-Q)^{-2}\ R \begin{bmatrix} p(A|i) & 0 \\ 0 & p(B|i) \end{bmatrix}^{-1}. \tag{26}$$

EBRWd: Markov chain formulation (discrete time). One potential limitation of the above Markov chain formulation is in response time computation. It considers only the expected retrieval time for all the exemplars. This limitation may be significant when the retrieval time is systematically different between exemplars: for example, one exemplar from Category $\mathbb{A}$ is much faster to retrieve while on average, the exemplars from Category $\mathbb{A}$ is much slower to retrieve. To overcome this limitation, we consider time to retrieve individual exemplars in the EBRW model.

As we consider retrieval of individual exemplars here, a state in Markov chain is now characterized by both evidence accumulated and the last retrieved exemplar. These states are illustrated in Figure 2. This Markov chain has $(n+n')\,(\theta_B+\theta_A-1)+1$ transient states, where n is the number of exemplars in Category $\mathbb{A}$ and n' is the number of exemplars in Category $\mathbb{B}$. Similarly, the number of absorbing states is now $n+n'$.

The transition matrix is formulated as above:

$$P_d = \left[\begin{array}{c|c} I & 0 \\ \hline R_d & Q_d \end{array}\right], \tag{27}$$

where R_d is of size $((n+n')\,(\theta_B+\theta_A-1)+1)\times(n+n')$ and Q_d is a square matrix. Specif-

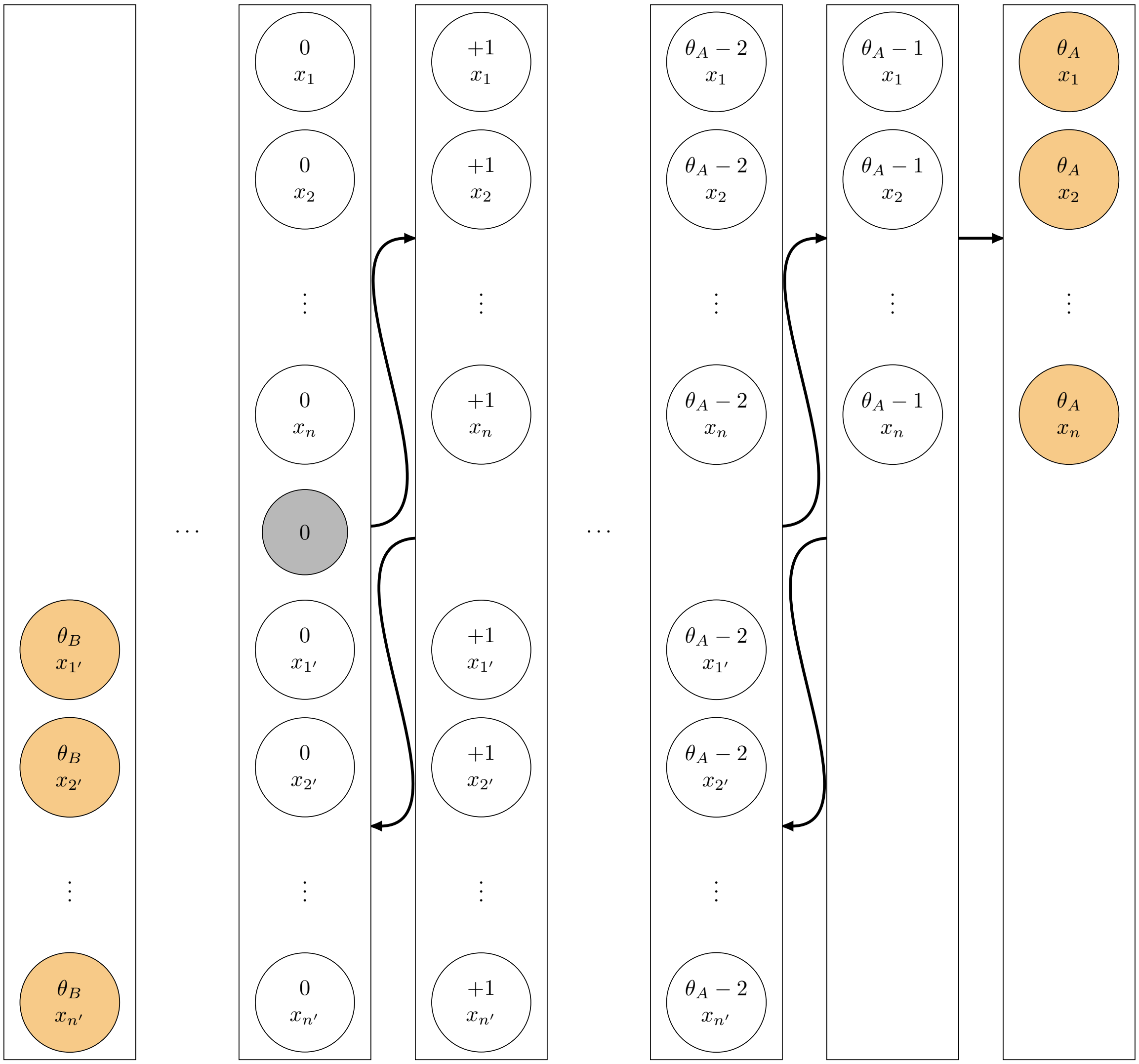

Figure 2. A state is characterized by accumulated evidence and the last exemplar retrieved. A gray node is a starting state, and yellow nodes indicate absorbing states.

ically,

$$
R_d = \begin{array}{c c}
\text{state} & \begin{array}{cccccccc} \theta_A,x_1 & \theta_A,x_2 & \cdots & \theta_A,x_n & \theta_B,x_{1'} & \theta_B,x_{2'} & \cdots & \theta_B,x_{n'} \end{array} \\
\begin{array}{c}
\theta_B-1,x_{1'} \\ \theta_B-1,x_{2'} \\ \vdots \\ \theta_B-1,x_{n'} \\ \theta_B-2,x_1 \\ \theta_B-2,x_2 \\ \vdots \\ \theta_A-2,x_{n'} \\ \theta_A-1,x_1 \\ \theta_A-1,x_2 \\ \vdots \\ \theta_A-1,x_n
\end{array} &
\begin{bmatrix}
0 & 0 & \cdots & 0 & p(1'|i) & p(2'|i) & \cdots & p(n'|i) \\
0 & 0 & \cdots & 0 & p(1'|i) & p(2'|i) & \cdots & p(n'|i) \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & 0 & \cdots & 0 & p(1'|i) & p(2'|i) & \cdots & p(n'|i) \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
p(1|i) & p(2|i) & \cdots & p(n|i) & 0 & 0 & \cdots & 0 \\
p(1|i) & p(2|i) & \cdots & p(n|i) & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
p(1|i) & p(2|i) & \cdots & p(n|i) & 0 & 0 & \cdots & 0
\end{bmatrix}
\end{array}, \tag{28}
$$

and

$$Q_d = \begin{array}{cc}
\text{state} & \begin{array}{ccccccccccccccccc} \cdots & -1,x_{1'} & -1,x_{2'} & \cdots & -1,x_{n'} & \cdots & 0,x_1 & 0,x_2 & \cdots & 0,x_n & \cdots & 1,x_1 & 1,x_2 & \cdots & 1,x_n & \cdots \end{array} \\
\begin{array}{c} \vdots \\ 0,x_1 \\ 0,x_2 \\ \vdots \\ 0,x_n \\ \vdots \end{array} & \left[\begin{array}{ccccccccccccccc}
\vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
p(1'|i) & p(2'|i) & \cdots & p(n'|i) & \cdots & 0 & 0 & \cdots & 0 & \cdots & p(1|i) & p(2|i) & \cdots & p(n|i) \\
p(1'|i) & p(2'|i) & \cdots & p(n'|i) & \cdots & 0 & 0 & \cdots & 0 & \cdots & p(1|i) & p(2|i) & \cdots & p(n|i) \\
\vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
p(1'|i) & p(2'|i) & \cdots & p(n'|i) & \cdots & 0 & 0 & \cdots & 0 & \cdots & p(1|i) & p(2|i) & \cdots & p(n|i) \\
\vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots
\end{array}\right.
\end{array} \tag{29}$$

As defined above, $p(j|i)$ is the probability of retrieving Exemplar j using Item i as a cue. Each row in P_d matrix sums to 1.

Response probability. Then as above, the fundamental matrix is given by

$$M_d = (I - Q_d)^{-1}, \tag{30}$$

and the probability of reaching the absorbing states is

$$[(\theta_A, x_1), (\theta_A, x_2), \ldots, (\theta_A, x_n), (\theta_B, x_{1'}), (\theta_B, x_{2'}), \ldots, (\theta_B, x_{n'})] = Z\ M_d\ R_d. \tag{31}$$

The probability of classifying Item i into each category is given by summing the probabilities of appropriate absorbing states:

$$[p(\mathbb{A}),\ p(\mathbb{B})] = Z\ M_d\ R_d \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}. \tag{32}$$

Response time. As the fundamental matrix (M_d) is the expected number of times each exemplar is retrieved, we first multiply each element in M_d with the expected retrieval time for each exemplar:

$$\mathcal{E}(t_{\cdot(z,j)}) = \left(\alpha + \frac{1}{a_{ij}}\right)\ m_{\cdot(z,j)}. \tag{33}$$

Then the expected response time is given by summing the relevant elements:

$$\mathcal{E}(t) = Z\ T \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{34}$$

EBRWc: Markov process formulation (continuous time). In addition, we formulate the EBRW with Markov process (continuous time). As in the EBRWd, the EBRWc has $(n+n')(\theta_B+\theta_A-1)+1$ transient states, and $n+n'$ absorbing states.

In the EBRWc, we consider transition rates, instead of transition probabilities. The rate matrix is written with four sub-matrices:

$$P_c = \left[\begin{array}{c|c} 0 & 0 \\ \hline R_c & Q_c \end{array}\right], \tag{35}$$

where R_c is of size $((n+n')(\theta_B+\theta_A-1)+1)\times(n+n')$ and Q_c is a square matrix. Specifically,

$$R_c = \begin{array}{c c} \text{state} & \begin{array}{cccccccc} \theta_A,x_1 & \theta_A,x_2 & \cdots & \theta_A,x_n & \theta_B,x_{1'} & \theta_B,x_{2'} & \cdots & \theta_B,x_{n'} \end{array} \\ \begin{array}{c} \theta_B-1,x_{1'} \\ \theta_B-1,x_{2'} \\ \vdots \\ \theta_B-1,x_{n'} \\ \theta_B-2,x_1 \\ \theta_B-2,x_2 \\ \vdots \\ \theta_A-2,x_{n'} \\ \theta_A-1,x_1 \\ \theta_A-1,x_2 \\ \vdots \\ \theta_A-1,x_n \end{array} & \left[\begin{array}{cccccccc} 0 & 0 & \cdots & 0 & a_{i1'} & a_{i2'} & \cdots & a_{in'} \\ 0 & 0 & \cdots & 0 & a_{i1'} & a_{i2'} & \cdots & a_{in'} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & a_{i1'} & a_{i2'} & \cdots & a_{in'} \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & a_{i1'} & a_{i2'} & \cdots & a_{in'} \\ a_{i1} & a_{i2} & \cdots & a_{in} & 0 & 0 & \cdots & 0 \\ a_{i1} & a_{i2} & \cdots & a_{in} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} & 0 & 0 & \cdots & 0 \end{array}\right] \end{array}, \tag{36}$$

and

$$Q_c = \begin{array}{c c} \text{state} & \begin{array}{ccccccccccccccccc} \cdots & -1,x_{1'} & -1,x_{2'} & \cdots & -1,x_{n'} & \cdots & 0,x_1 & 0,x_2 & \cdots & 0,x_n & \cdots & 1,x_1 & 1,x_2 & \cdots & 1,x_n & \cdots \end{array} \\ \begin{array}{c} \vdots \\ 0,x_1 \\ 0,x_2 \\ \vdots \\ 0,x_n \\ \vdots \end{array} & \left[\begin{array}{ccccccccccccccccc} & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots & \\ & a_{i1'} & a_{i2'} & \cdots & a_{in'} & \cdots & \lambda & 0 & \cdots & 0 & \cdots & a_{i1} & a_{i2} & \cdots & a_{in} & \\ & a_{i1'} & a_{i2'} & \cdots & a_{in'} & \cdots & 0 & \lambda & \cdots & 0 & \cdots & a_{i1} & a_{i2} & \cdots & a_{in} & \\ & \vdots & \vdots & & \vdots & & \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & & \vdots & \\ & a_{i1'} & a_{i2'} & \cdots & a_{in'} & \cdots & 0 & 0 & \cdots & \lambda & \cdots & a_{i1'} & a_{i2} & \cdots & a_{in} & \\ & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots & \end{array}\right] \end{array}. \tag{37}$$

A cell located at row k column j of this P_c matrix represents the rate of going to state j, that is, the rate of retrieving Exemplar j, a_{ij}. Also, the diagonal entries in Q_c is a negated sum of all the rates:

$$\lambda = -\left(\sum_{j=1}^{n} a_{ij} + \sum_{j=1'}^{n'} a_{ij}\right). \tag{38}$$

Thus, each row in P_c matrix sums to 0.

Response probability. To compute the response probability, we first note that the matrix multiplication leads us to

$$P_c^n = \begin{bmatrix} 0 & 0 \\ Q_c^{n-1} R_c & Q_c^n \end{bmatrix}, \quad n = 1, 2, \ldots. \tag{39}$$

Then we derive probability of each state after time t ($t > 0$) has passed:

$$\begin{aligned} F(t) &= \exp(P_c\, t) \quad \because \text{Solution to Kolmogorov equations} &(40)\\ &= \sum_{n=0}^{\infty} \frac{P_c^n\, t^n}{n!} \quad \because \text{Taylor approximation} &(41)\\ &= I + \sum_{n=1}^{\infty} \frac{P_c^n\, t^n}{n!} &(42)\\ &= \begin{bmatrix} I & 0 \\ R_c(t) & Q_c(t) \end{bmatrix}, &(43) \end{aligned}$$

where the transition probability within the transient states is

$$\begin{aligned} Q_c(t) &= \sum_{n=0}^{\infty} \frac{Q_c^n\, t^n}{n!} &(44)\\ &= \exp(Q_c\, t), &(45) \end{aligned}$$

and the transition probability from the transient states to the abosorbing states is

$$\begin{aligned} R_c(t) &= \int_0^t Q_c(u)\, R_c\, du &(46)\\ &= \left(\int_0^t Q_c(u)\, du \right) R_c. &(47) \end{aligned}$$

The probability for the absorbing state is derived by letting $t \to \infty$.

$$\lim_{t\to\infty} Q_c(t) = 0 \tag{48}$$

and

$$\begin{aligned} \lim_{t\to\infty} R_c(t) &= \left(\int_0^\infty Q_c(u)\, du \right) R_c &(49)\\ &= -Q_c^{-1}\, R_c. &(50) \end{aligned}$$

Thus

$$\lim_{t\to\infty} F(t) = \begin{bmatrix} I & 0 \\ -Q_c^{-1}\, R_c & 0 \end{bmatrix}. \tag{51}$$

Then, the probability of reaching the absorbing states is given by

$$[(\theta_A, x_1), (\theta_A, x_2), \ldots, (\theta_A, x_n), (\theta_B, x_{1'}), (\theta_B, x_{2'}), \ldots, (\theta_B, x_{n'})] = Z \begin{bmatrix} I \\ -Q_c^{-1}\ R_c \end{bmatrix}. \quad (52)$$

The probability of classifying Item i into each category is given by summing the probabilities of appropriate absorbing states:

$$[p(\mathbb{A}),\ p(\mathbb{B})] = Z \begin{bmatrix} I \\ -Q_c^{-1}\ R_c \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}. \quad (53)$$

Response time. The expected time spent on a transient state is given by

$$\int_0^\infty Q_c(t)\, dt = -Q_c^{-1}. \quad (54)$$

Similarly to the fundamental matrix for the Markov chain (M above), each element in this matrix represents the expected time spent on each state: a cell at the kth row jth column corresponds to the time spent on state j, when the starting state is k. Thus, time required to reach a response boundary is given by

$$\mathcal{E}(t) = -Z\ Q_c^{-1} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (55)$$

SSM: Sequential sampling model

The EBRW model assumes that people use only the target item (Item i in above). However, previous research on free recall has demonstrated that memory retrieval is not an independent process. Rather, memory retrieval is sequence dependent, and here, we implement the first-order Markov chain in retrieval process. That is, the probability of retrieving exemplar j also depends on its similarity to the last exemplar retrieved from memory. In particular, we parameterize the model such that at each step of evidence accumulation, people use the last retrieved exemplar as a cue to retrieve a next exemplar with probability ψ, and with probability $1 - \psi$, people use the target item as a cue.

Unlike the target item, which is to be classified, the last retrieved exemplar has its category label available as well as its features. Thus when the last retrieved exemplar is used as a cue, people may consider its category label and are more likely to retrieve an exemplar from the same category. We implement this possibility by revising the distance between exemplars. Specifically, when the last retrieved exemplar is k,

$$d_{kj} = \left(\omega_l \; 1_{\{y_k = y_j\}} + \sum_m \omega_m \, |x_{im} - x_{jm}|^\rho\right)^{1/\rho}, \tag{56}$$

where y_k is the category label of Exemplar k, and $0 \le \omega_l$, $0 \le \omega_m$, and $\omega_l + \sum \omega_m = 1$.

SSMd: Markov chain formulation (discrete time). As in the EBRWd, each state is characterized by evidence accumulated and the last retrieved exemplar. Thus, the number of transient state is $(n + n')\,(\theta_B + \theta_A - 1) + 1$ and the number of absorbing states is $n + n'$. The probability of transitioning from state $(0, k)$ to state $(+1, j)$ by retrieving Exemplar j in Category $\mathbb{A}$ is now

$$p_{(0,k)(+1,j)} = \frac{\psi \, a_{kj} + (1 - \psi)\, a_{ij}}{\psi \, \sum_j a_{kj} + (1 - \psi) \, \sum_j a_{ij}.} \tag{57}$$

We also divide the transition matrix P into four sub-matrices just as we did for the EBRWd. The rest of derivation is the same as that for the EBRWd.

SSMc: Markov process formulation (continuous time). Similarly with the Markov process formulation, the rate to transition from state $(0, k)$, where Exemplar k is last retrieved, to state $(+1, j)$ by retrieving Exemplar j from Category $\mathbb{A}$ is

$$p_{(0,k)(+1,j)} = \psi \; a_{jk} + (1 - \psi) \; a_{ik}. \tag{58}$$

The rest of derivation is the same as that for the EBRWc.

Model Parameters

GCM

— ρ: similarity kernel parameter
— γ: similarity gradient
— η: memory strength
— θ_A and θ_B: response boundary

EBRW

— ρ: similarity kernel parameter
— γ: similarity gradient
— η: memory strength
— θ_A and θ_B: response boundary
— α: time to accumulate evidence after retrieving an exemplar (only for EBRWd)

SSM

— ρ: similarity kernel parameter
— γ: similarity gradient
— η: memory strength
— θ_A and θ_B: response boundary
— α: time to accumulate evidence after retrieving an exemplar (only for SSMd)
— ψ: the probability of using the last retrieved exemplar as a cue.

Simulation

References

Bhat, U. N. & Miller, G. K. (2002). *Elements of applied stochastic processes* (third). New Jersey: John Wily & Sons, Inc.

Busemeyer, J. R. (1982). Choice behavior in a sequential decision-making task. *Organizational Behavior and Human Performance*, *29*, 175–207.

Diederich, A. (1995). Intersensory facilitation of reaction time: evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, *39*, 197–215.

Diederich, A. & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*, 304–322.

Kijima, M. (1997). *Markov processes for stochastic modeling*. London: Chapman & Hall.

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

Nosofsky, R. M. & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.