Journal of Memory and Language 76 (2014) 158–173

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Retrieval dynamics of the strength based mirror effect in recognition memory

Aslı Kılıç *, Ilke Öztekin

Department of Psychology, Koç University, Turkey

ARTICLE INFO

Article history:
Received 9 December 2013
revision received 10 June 2014

Keywords:
Item recognition
The strength based mirror effect
The response-deadline speed-accuracy trade-off procedure
Memory models
Response bias

ABSTRACT

The strength based mirror effect (SBME) refers to an increase in hit rates (HR) and a decrease in false alarm rates (FAR) for the test lists that follow a strongly encoded study list. Earlier investigation of accuracy and reaction time distributions by fitting the diffusion model indicated a mirror effect in the drift rate parameter, which was interpreted as an indication of more conservative responses due to a shift in the drift criterion. Additionally, the starting point for the evidence accumulation was found to be more liberal for the strong test lists. In order to further investigate this paradoxical effect of list strength on these two kinds of bias estimated from the diffusion model, we employed the response-deadline procedure which provided a direct assessment of response bias early in retrieval, prior to evidence accumulation. Results from the retrieval functions indicated more liberal response bias in the list strength paradigm with both pure- and mixed-strength study lists. On the contrary, the SBME was observed at the asymptotic accuracy, suggesting that the conservative response bias might be observed later in retrieval when memory evidence has fully accumulated. In addition, comparison of the SBME across pure and mixed lists revealed that the SBME was most prominent in the pure-list paradigm, suggesting that both the differentiation and criterion shift accounts jointly explain the SBME in recognition memory.

Introduction

Episodic memory is often tested in the laboratory by presenting participants a list of items to study. In an item recognition task, participants are asked to endorse the items they have recently studied (targets) and reject the new items (foils). In recognition memory, when a list of items is strengthened via increasing the number of repetitions or manipulations during encoding, the probability to correctly endorse targets (hit rate) increase and the probability to incorrectly endorse foils (false alarm rate) decrease, producing a *strength based mirror effect* (SBME, Glanzer & Adams, 1985; Ratcliff, Clark, & Shiffrin, 1990; Stretch & Wixted, 1998). This subjective memory strength can be defined as a global match between the test item and traces in memory or alternatively as familiarity, based on the signal detection framework.

Previous research employed reaction time distributions to study the SBME. For instance, Criss (2010) and Starns, Ratcliff, and White (2012) applied the diffusion model (Ratcliff, 1978), a dynamic version of the signal detection framework, in a list-strength paradigm. In the diffusion model, memory evidence is assumed to accumulate over time and a response is given when enough evidence is accumulated towards one of the two responses ("yes" and "no") in an item-recognition task. The two responses are represented as two boundaries and the separation between the two boundaries can measure the speed-accuracy

* Corresponding author. Address: Department of Psychology, Koç University, Rumeli Feneri Yolu, Sariyer 34450, Istanbul, Turkey.
E-mail address: akilic@ku.edu.tr (A. Kılıç).

http://dx.doi.org/10.1016/j.jml.2014.06.009
0749-596X/

trade-off. The placement of the boundaries depends on the participant and can be manipulated by experimental conditions. For example, if the participants are instructed to give accurate responses rather than fast responses, they place their response boundaries far apart from each other and thus, giving a response requires more time and evidence is more likely to accumulate towards the correct response (Ratcliff, 1985). The *starting point* parameter measures the tendency towards one of the responses by indicating the point at which the evidence begins to accumulate towards one of the boundaries. For example, if the starting point is closer to the "no" boundary, the frequency of the "no" responses will be higher and the mean reaction time of the "no" responses will be shorter while the mean reaction time of the "yes" responses will be longer. The parameter that indicates the rate of evidence accumulation is the drift rate parameter (v). At each time point, the sampled evidence is compared to a criterion (*drift criterion*) and if the sampled evidence exceeds the criterion, evidence accumulates towards the "yes" boundary; if it fails to exceed the criterion, evidence accumulates towards the "no" boundary. In summary, there are three different types of criterion that determines the decisions made in the diffusion model: Boundary separation, starting point and the drift criterion.

Criss (2010) manipulated list-strength in item recognition and the parameters of the diffusion model showed that when speed-accuracy trade-off (boundary separation parameter) was taken into account, a mirror effect was observed in the drift rate parameters for the items tested in strong lists. The responses were more accurate and the average reaction time of the correct responses was faster for the foils tested along with strong targets (strong foils) compared with the responses of the foils tested along with weak targets (weak foils). Thus, faster and more accurate correct responses ("no") to the strong foils have produced lower drift rate (higher in absolute value) and the decrease in the drift rate has been interpreted as a decrease in the overall memory strength for the strong foils. This explanation depends on the differentiation mechanisms, which causes the foils to become less similar to the targets when items are strengthened during encoding. Accordingly, the differentiation models propose that foils that are compared to strong targets become less confusable at retrieval (Criss, 2006, 2009, 2010; Criss & McClelland, 2006; Criss et al., 2013).

The decrease in the drift rate of the strong foils could be alternatively interpreted as a shift in the drift criterion in the diffusion model (see Starns, Ratcliff, et al., 2012). That is because the drift rate is defined in relation to the drift criterion, as the distance from the drift criterion determines the drift rate. The exact placement of the drift criterion cannot be estimated in the diffusion model and it is arbitrarily set to the zero point of the drift rate. Starns, Ratcliff, et al. (2012) suggested that when items were strengthened during encoding, participants required more evidence to endorse the probe, thus the drift criterion shifts hypothetically to some positive value. In the diffusion model, this shift is manifested as faster accrual of evidence towards the "no" boundary, as the sampled evidence for the strong foils at each time step will more likely fail to exceed the drift criterion. In addition to the mirror effect observed in the drift rates, both studies reported that the starting point parameter was more liberal, meaning that participants were more biased towards the "yes" response boundary when tested with the strong targets.

Critical evidence for a shift in the drift criterion comes from the SBME observed when list strength is manipulated only during test (Starns, Ratcliff, et al., 2012; Starns, White, & Ratcliff, 2010; 2012). Different from previous studies in which strength was manipulated in pure lists (i.e. strength was manipulated across lists), Starns et al. presented participants with mixed lists of items (i.e. strength was manipulated within lists). However, in the subsequent test lists, either weak or strong targets were tested along with foils. The SBME observed in the drift rates after studying a mixed-list could only be explained by a shift in the drift criterion. That is because the memory evidence for foils after a mixed study list would be comparable across test strength conditions, and as a result, a decrease in the drift rates for strong foils would not be expected due to a differentiation mechanism. Similar to the findings from the pure-list paradigm, the starting point for evidence accumulation was closer to the "yes" boundary for strong targets.

In the current study, we tested whether list-strength has opposite effects on these two kinds of criterion, namely starting point and drift criterion. To do so, we employed the response-deadline speed-accuracy trade-off (SAT) procedure, which provides an in-depth investigation of different types of response bias by controlling for the speed-accuracy tradeoffs over the course of retrieval.

The response-deadline SAT procedure

The SAT procedure provides conjoint and unbiased measures of retrieval speed and retrieval success (Benjamin & Bjork, 2000; Hintzman & Curran, 1994; McElree & Dosher, 1989; Öztekin, Gungor, & Badre, 2012; Öztekin & McElree, 2007, 2010). In contrast to traditional reaction time measures, which are subject to speed-accuracy trade-offs and hence cannot provide pure measures of processing speed, by providing the full time-course of retrieval, SAT procedure yields independent assessment of accuracy and speed of processing (see McElree, 2006 for an overview). In SAT, participants are cued to respond with a response signal (a tone) presented at one of several time points, typically ranging from 60 to 3000 ms after the probe onset. The lag between the probe onset and the response signal is assigned randomly to test trials and participants are trained to give a response within 300 ms after the response cue. Although the diffusion model can also quantitatively account for the speed-accuracy trade-off by measuring the criterion to terminate the evidence accumulation (boundary separation), experimental manipulation of response deadline has the further advantage of providing the full time course of retrieval for each experimental condition, in addition to eliminating the bias related to speed-accuracy trade-off.

SAT retrieval functions can describe changes in accuracy as a function of total processing time, the total time that passes from probe onset to the response after the response deadline. The SAT functions typically start with a period of chance performance where the information retrieved is not adequate to discriminate between targets and foils. Later, a rapid increase in accuracy follows and shows the rate of information accrual over additional processing time. Finally, an asymptote is observed, indicating the overall accuracy, which does not further improve by additional retrieval time (Fig. 1). The shape of this function is usually well fit by an exponential approach to a limit. Three parameters describe these functions: (a) an asymptote, reflecting overall limitations of memory (total available memory strength), (b) an intercept, indicating the point in time at which performance departs from chance, and (c) a rate of rise from chance, reflecting retrieval speed. The asymptote parameter indicates the retrieval success, while the intercept and the rate parameters jointly constitute retrieval speed measures.

Current study

In this study, we employed the response-deadline SAT procedure in a list-strength paradigm to investigate how strengthening items across (pure-list) and within (mixed-list) lists increase accuracy by simultaneously increasing hit rates and decreasing false alarm rates. Employing the SAT procedure allowed us to further examine the response bias early in retrieval. In the SAT procedure, the accuracy of the responses that follow earlier lags (e.g., 60 ms, 100 ms) tends to be at chance. This means that the hit rates and the false alarm rates would be comparable and consequently, the proportion of "yes" responses would potentially indicate the starting point of the evidence accumulation in terms of the diffusion model. We name this type of bias as *prior bias*, referring to participants' tendency towards one of the responses even before they are presented with the test probe. That corresponds to a bias that is observed when participants are given incentives to prefer one response to the other or when the change in the base rate of targets and foils result in a preference over one of the responses. We refer to the bias that is moderated by the drift criterion as *evidentiary bias*, which can be observed once performance reaches asymptote. The evidentiary bias can be explained as the willingness to accumulate evidence towards one of the responses (e.g., in the strong condition, participants were claimed to require more evidence to endorse a test probe). One advantage of employing the response-deadline procedure is to investigate whether the strength manipulation has differential effects on the prior and evidentiary bias. More liberal starting points could be due to greater drift rates in the strong lists, as these two parameters tend to be correlated in the diffusion model fits of the simulated data (Ratcliff & Tuerlinckx, 2002), or alternatively, strengthening a list of items cause opposite effects on the starting point of evidence accumulation and the criterion to endorse a probe. Thus, the current study provides an empirical investigation as to whether more liberal starting points would be observed as a more liberal prior bias in the retrieval function, and consequently whether there was a real effect of strength on the starting point parameter as opposed to a misestimation of the diffusion model.

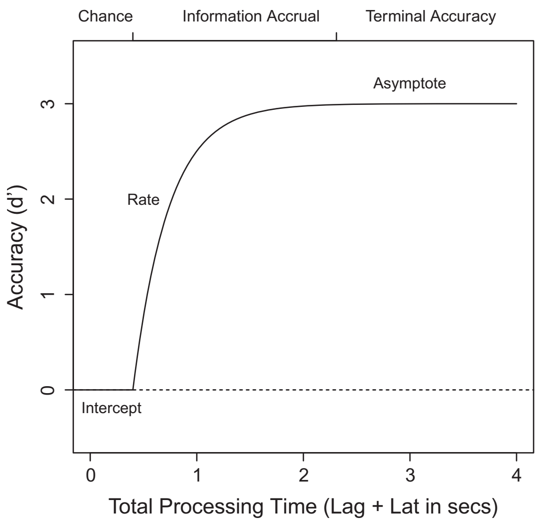

Fig. 1. Illustration of hypothetical SAT function which plots accuracy in d' units as a function of total processing time. In SAT function there are three phases, (a) a period when performance is at chance ($d' = 0$), (b) a period of information accrual (when performance departs from chance and increases with a constant rate) and (c) a period of terminal accuracy (when maximum level of accuracy is reached).

We manipulated strength via a levels-of-processing task (Craik & Lockhart, 1972), which has been shown to produce the SBME similar to strengthening via repetition or study duration (Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Shimizu, Velanova, & Rhodes, 2005). Participants were presented with a pure-list paradigm (Experiment 1) where the SBME would be observed at the asymptote but not in the speed parameters (e.g., Dosher, 1984). The rate parameter in the SAT function should not be confused with the drift rate parameter of the diffusion model. The strength effect observed in the mean drift rate parameter changes the asymptote but not the rate parameter of the SAT function. In the retrieval function derived from the diffusion model, the rate of information accumulation is defined as the ratio of within-trial variance to across-trial variance of the drift rate (Gronlund & Ratcliff, 1989; Ratcliff, 1978). In other words, for a given mean drift rate, if the variance of evidence accumulation in a single trial is less than the variance across trials, then information reaches asymptote faster. Otherwise, it will take longer to reach the asymptotic drift rate. This assumption indicates that the drift rate is independent of the rate at which accuracy approaches to its limits. This suggests that the strength effect will be observed as an increase in the asymptote rather than an increase in the speed of information accrual in the SAT function.

In addition to the standard pure-list paradigm, we also manipulated strength in a mixed-list paradigm in which half of the items in a list were randomly strengthened during study. In Experiment 2 participants were not informed of the strength condition at test contrary to the mixed-list paradigm employed in Starns, Ratcliff, et al. (2012) study, and thus the SBME would not be predicted in the asymptotic FAR because of a null effect on the drift criterion.

The aim of Experiment 2 was to test whether strength would have an effect on prior bias while no effect was expected on the evidentiary bias. In Experiment 3, we sought to test the drift criterion shift account by presenting participants the strength condition of the targets before proceeding with the recognition task, so that participants could shift their criterion accordingly. Finally, Experiment 4 investigated whether the size of the strength effect differ significantly across pure- and mixed-list paradigms, which aimed to investigate the contribution of criterion shifts and the differentiation mechanism in the list-strength paradigm.

Experiment 1

In order to investigate the retrieval dynamics of the SBME, the SAT procedure was first applied to the traditional pure-study paradigm where the items were strengthened across lists in Experiment 1. Increasing the study time of items has been shown to increase the asymptotic accuracy but no significant effect on the retrieval speed (Dosher, 1984). Application of the diffusion model to the list-strength paradigm showed that evidence accumulation starts from a point closer to the "yes" boundary, suggesting an adoption of a more liberal prior bias in the retrieval function of the strong items.

Method

Participants

Twelve undergraduate students from Koç University took part in the experiment and received monetary compensation for their participation. One participant who dropped out of the experiment after the first session and two participants who had low overall accuracy ($d' < 0.35$) were excluded from the subsequent analysis. Five of the remaining 9 participants were female and 2 of them were left-handed.

Materials

The word pool consisted of 902 words. Six hundred of the words were from Turkish Word Norms (Tekcan & Göz, 2005). The remaining 302 words were randomly selected from the association sets of the Turkish Word Norms[1].

Procedure and design

Participants completed two 50-min sessions with an additional 10-min practice session for the SAT procedure at the beginning of the first session. Each session consisted of four study-test blocks. In the study block, participants were presented with 140 words[2] and were administered a levels-of-processing task as the strength manipulation (Craik & Lockhart, 1972). For the strong condition, they were required to make a semantic judgment ("Does the word have a pleasant meaning?") and for the weak condition, they made an orthographic judgment ("Does the word contain the letter 'e'?"). The study block was self-paced, as the word was displayed on the screen until the participant responded with the 'z' ("yes") or the '?/' ("no") keys, and a 100-ms ISI followed each response. In each session, half of the study-test blocks (2) were strong and the other half (2) were weak, which were presented in random order. The test list consisting of 70 targets and 70 foils immediately followed the study list. Each test item was presented for the duration of the response-deadline after a 500-ms presentation of a visual mask consisting of non-word symbols (see Fig. 2). The visual mask was presented right before the test item in order to prepare the participant to the new test trial. The response-deadline was cued with a 50-ms tone at the 60, 200, 300, 500, 800, 1500, 3000 ms after the stimulus onset. Participants were trained to respond within 300-ms of the tone and received a feedback of their response time. If they fail to respond within the allotted time or respond earlier then the tone, they received a feedback ("your response is too slow" or "you responded before the tone", respectively). Early responses and the responses that were longer than 600-ms, were excluded from the subsequent analysis.

The experiment was a 2 (Strength) $\times$ 7 (Response Lag) within-subjects design. There were 40 responses for targets and 40 responses for foils at each strength and lag condition. Lag condition was assigned randomly within each strength and test item type condition over the course of testing. After the removed trials, the number of responses for each strength and item condition was 33.56 ($SD = 5.29$), 36.76 ($SD = 3.07$), 37.78 ($SD = 2.86$), 38.05 ($SD = 2.13$), 37.86 ($SD = 2.07$), 36.80 ($SD = 2.73$) and 35.77 ($SD = 3.90$) on average, for 60, 200, 300, 500, 800, 1500 and 3000 ms response lag conditions respectively.

Results and discussion

Accuracy

As the accuracy measure, d's were obtained for each strength and lag condition for each participant. The perfect performances of hit rates (HR) and false alarm rates (FAR) were adjusted as follows: HR greater than .99 were adjusted to .98 and FR lower than .01 were adjusted to .02, as an approximation to the Snodgrass & Corwin (1988) adjustment which was previously used in previous studies employing the response deadline SAT procedure (e.g., Öztekin & McElree, 2007, 2010; Öztekin et al., 2012). The full time-course of the strength effect on accuracy was examined by fitting data at the group and individual level as an exponential approach to a limit (Dosher, 1981; McElree & Dosher, 1993; Nobel & Shiffrin, 2001; Wickelgren, 1977; Öztekin & McElree, 2007, 2010; Öztekin et al., 2012):

$$d'(t) = \lambda(1 - e^{-\beta(t-\delta)}), \quad t < \delta, \quad \text{else } 0 \tag{1}$$

where $d'(t)$ is the predicted d' at time t, λ is the asymptotic accuracy reflecting the overall performance of recognition

[1] Association sets include the words that were generated from the words in Turkish Word Norms by a free association task (see Tekcan & Göz, 2005 for the association sets).

[2] In the following experiments, participants study strong and weak targets in the same list but later tested on either strong or weak targets only. In order to eliminate the length of the study list as confound, we presented participants with targets (70) and filler items (70).

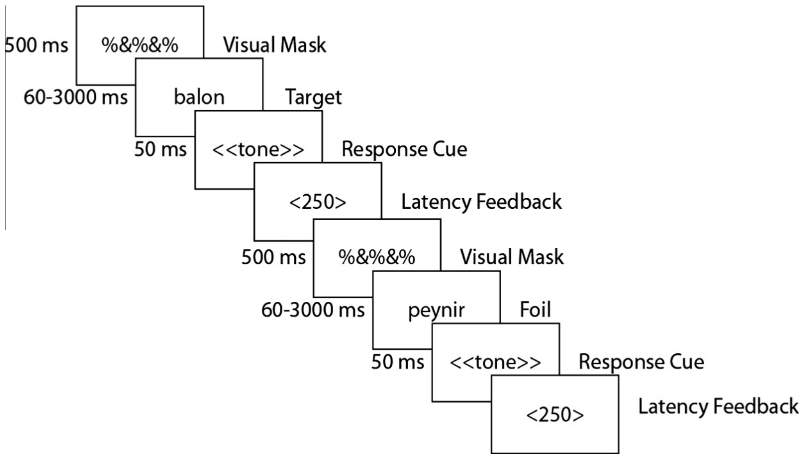

Fig. 2. Illustration of the test trials. The first test probe is a target and the second test probe is a foil in the above illustration. In the experiment, 140 test probes were presented in each block, half of which were targets.

in d' units, δ is the intercept reflecting the time at which accuracy departs from chance and β is the rate of the information accrual from chance to asymptote.

The data were fit with the exponential function using the *optim* function in R (R Core Team., 2012) to estimate the three parameters with the maximum likelihood estimation (MLE) method. In the MLE method, data is assumed to be normally distributed and as a result, the variance of d's of each condition for each individual participant can be estimated with the following function:

$$sd(d') = \sqrt{\frac{HR(1-HR)}{n_t\phi^2[z(HR)]} - \frac{FR(1-FR)}{n_F\phi^2[z(FR)]}}, \quad (2)$$

where n_T is the number of targets, n_F is the number of foils, ϕ is the normal density function and z is the inverse normal transformation (Gourevitch & Galanter, 1967; Liu & Smith, 2009). For group fits, d's were obtained by averaging d's at each strength and lag condition across participants and the standard deviation that was fed into MLE method was the standard error of the mean d's in the group fits[3].

In order to select the most parsimonious model, we fit the data with nested models ranging from a 3-parameter null model (a common asymptote [λ], a common rate [β], and a common intercept [δ] parameter for both of the strength conditions) to a 6-parameter full model (a unique asymptote [λ], a unique rate [β], and a unique intercept [δ] parameter for each strength condition). The best fitting model was selected based on three criteria: (1) The value of BIC, AICc and adjusted R^2 statistics from the group data fit; (2) the consistency of parameter estimates across participants; (3) evaluation of whether the fit yielded systematic deviations that could be accounted for by additional parameters (Öztekin et al., 2012; Öztekin & McElree, 2007, 2010). In order to achieve the latter two criteria, statistical tests were conducted on the best fitting parameter values across participants.

[3] In addition to MLE estimates, models were also fit by minimizing the squared errors (least squares estimation). The results that will be presented below did not change based on the estimation method used.

The exponential function fit comparison values of the group data are presented in Table 1. The list-strength effect is best explained by the 2λ–1β–1δ model that allocates separate parameters for asymptotic accuracy and common parameters for retrieval speed of the weak and the strong list conditions (Fig. 3A). In order to assess the parameter consistency across participants, the full model (2λ–2β–2δ) was fit to individual participants' data. Parameter estimates derived from the individual model fits indicate that asymptotic accuracy (λ) is significantly greater for the strong condition (M=1.78, SD = 0.64) compared with the weak condition (M = 0.75, SD = 0.38), $t(8)$ = 5.53, p < .001, supporting the conclusions derived from the model fit comparisons of the averaged data. Consistent with the best fitting model of the average data, the retrieval speed parameters (β and δ) did not differ significantly across list-strength conditions (1/β, t = 0.52; δ, t = −1.74). Parameter values of the best fitting model to group data and individual data are presented in Table 2.

These results suggest that strengthening a list of items in an item recognition task might increase the availability of information in memory; but crucially does not have a measurable effect on the retrieval dynamics. However, what is more critical for the SBME is that the increase in accuracy is due to a simultaneous increase in HR and a decrease in FAR. Thus, it is also important to investigate the time-course of how HR and FAR individually contribute to memory performance and identify the effect of list-strength on the time-course of HR and FAR separately.

Hit and false alarm analysis

Fig. 4A plots the probability of endorsing a test item as a function of total processing time for each strength condition and item type (target vs. foil). A 7 (lag) × 2 (strength) repeated-measures ANOVA on HR showed a main effect of strength, $F(1,8)$ = 16.35, p < .01, with HR increasing as a function of strength. A main effect of lag indicates that HR increased as a function of total processing time, $F(6,48)$ = 11.34, p < .001. The interaction between list-strength and lag was not significant, suggesting that the

Table 1
Exponential function fit statistics for the candidate models.

Model	Experiment 1			Experiment 2			Experiment 3		
	Adj R^2	BIC	AICc	Adj R^2	BIC	AICc	Adj R^2	BIC	AICc
$1\lambda–1\beta–1\delta$	.464	29.440	29.922	.708	15.465	15.948	.632	19.827	20.31
$2\lambda–1\beta–1\delta$	.986	0.353	2.241	.951	6.314	8.202	.984	−1.931	−0.043
$1\lambda–2\beta–1\delta$	.886	3.850	5.738	.784	12.152	14.041	.905	3.682	5.57
$1\lambda–1\beta–2\delta$	.572	18.583	20.471	.747	13.279	15.168	.638	18.912	20.8
$2\lambda–2\beta–1\delta$	.889	5.285	9.833	.765	14.595	18.900	.859	7.027	11.332
$2\lambda–1\beta–2\delta$	.989	2.755	7.059	.949	8.896	13.201	.983	0.706	5.01
$1\lambda–2\beta–2\delta$	.984	2.934	7.238	.946	8.845	13.149	.986	0.49	4.795
$2\lambda–2\beta–2\delta$	.980	5.817	13.983	.949	10.829	18.995	.982	3.123	11.289

Note: According to the model selection criteria (lowest value of BIC and AICc, parameter consistency across participants and parsimony), $2\lambda–1\beta–1\delta$ was selected as the best fitting model in all three experiments.

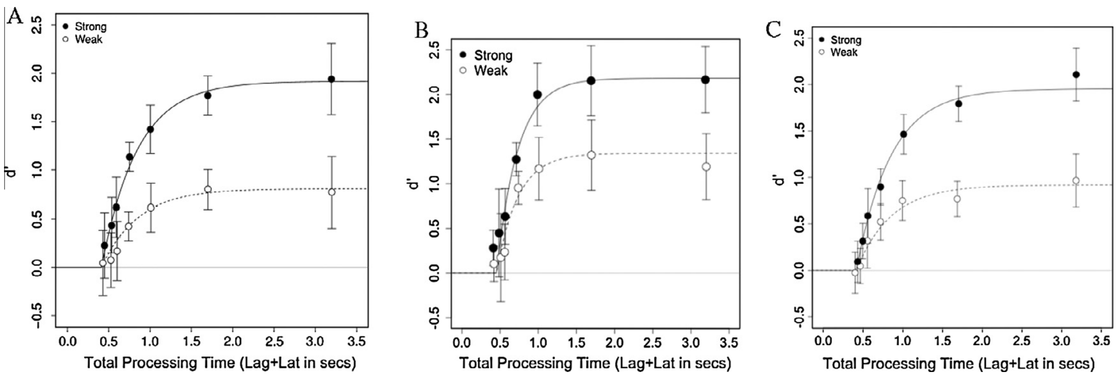

Fig. 3. SAT functions of the strong and weak list conditions averaged over participants. Accuracy (in d' units) plotted as a function of total processing time (duration of the response deadline plus response latency in secs). Points represent the empirical data and lines represent the best fitting exponential function ($2\lambda–1\beta–1\delta$) derived from Eq. (1). Error bars indicate the standard error of the mean d's at each lag and strength condition. Panel A: Encoding strength is manipulated across lists (Experiment 1). All the words in a given study list is encoded either strongly or weakly and at test, those targets are tested along with foils. Panel B: Encoding strength is manipulated within lists (Experiment 2). Half of the items were encoded strongly and the other half was encoded weakly. Strength was manipulated across lists during test. Participants were tested either with strong targets or weak targets along with foils. Participants were not informed of the strength of the list that they would be tested on. Panel C: Encoding strength is manipulated within lists (Experiment 3). Different from the second experiment, participants were informed of the test list condition by being told which judgment they had made during encoding.

Table 2
Parameter values of the best fitting exponential function in Experiment 1.

Parameter	Average	Participants								
		1	2	3	4	5	6	7	8	9
λ_{strong}	1.92	1.42	1.27	1.53	2.81	1.15	2.63	1.39	1.54	2.62
λ_{weak}	0.81	0.37	0.30	1.13	1.32	0.59	1.44	0.31	0.41	0.87
β	2.29	268.47	18.14	3.58	2.66	1.67	1.86	5.80	4.20	2.73
δ	0.41	0.72	0.47	0.44	0.42	0.30	0.41	0.71	0.43	0.57
Adj R^2	.986	.385	.660	.452	.818	.684	.91	.828	.787	.835

Note: Average parameter values are not the average of parameters obtained from participants' data, they are rather obtained from the fits to the data averaged across participants.

HR of the strongly encoded items was greater than that of the weakly encoded items even when the total processing time was relatively short (e.g. 0.5 s). FAR analysis indicated a different pattern: A 7 (lag) × 2 (strength) repeated-measures ANOVA on FAR did not reveal a main effect of list-strength, but a significant strength by lag interaction, $F(6,48) = 5.45$, $p < .001$. That is, FAR of the test lists composed of strongly encoded items decrease significantly at late retrieval; 1 s, $t(8) = -2.78$, $p = 0.02$; 1.7 s, $t(8) = 2.90$, $p = 0.02$; 3.2 s, $t(8)$, $p = 0.02$. Consistent with the HR data, a main effect of lag was also significant, $F(6,48) = 10.63$, $p < .001$, suggesting that the FAR decreased as a function of total processing time.

Fig. 4A plots the time course of HR and FAR, showing a pattern similar to the accuracy results. Early in retrieval, HR and FAR are comparable to each other, indicating a performance at chance level. This early period in retrieval corresponds to the period in which d' is equal to zero in the

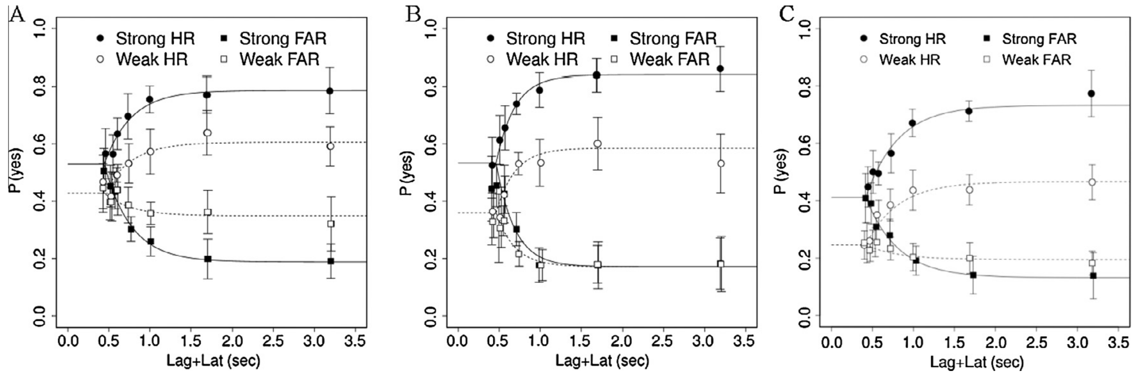

Fig. 4. Probability of endorsing the test item as a function of total processing time (duration of the response deadline plus response latency in secs). Time course of hit rates (HR) false alarm rates (FAR) are averaged over participants for each response lag and list-strength condition. Points represent the empirical data and lines represent the best fitting descriptive function derived from Eq. (3). Error bars are the standard error of the mean *P*(*"yes"*). Panel A: Encoding strength is manipulated across lists (Experiment 1). Panel B: Encoding strength is manipulated within lists (Experiment 2). Panel C: Encoding strength is manipulated within lists and participants were informed of the test list-strength (Experiment 3).

SAT function, and also provides an estimate for the prior bias observed at the beginning of retrieval. For example, the probability to endorse a probe that is greater than 0.5, implies a bias towards the "yes" response. If the probability to endorse a probe is lower than 0.5, then the bias is towards the "no" response. After a point in time that is similar to the intercept in the SAT function, HRs increase and FARs decrease, indicating an increase in the accuracy level. The important pattern observed here is the mirror effect for HRs and FARs, which becomes more prominent as the total processing time increases and later reaches an asymptote. This increase in the mirror effect can be further estimated with an exponential function, which provides independent and unbiased estimates of asymptotic strength and processing speed for HR and FAR. The following exponential function can be employed for further investigation of the prior bias observed early in retrieval:

$$P(\text{"}yes\text{"}) = \lambda_1 + (\lambda_2 - \lambda_1)(e^{-\beta(t-\delta)}), \quad t > \delta, \quad \text{else } \lambda_2, \qquad (3)$$

where P("yes") is the predicted probability to endorse a test item (HR or FAR at time t), λ_1 is the asymptotic probability to endorse a test item, λ_2 is the probability to endorse a test item before information begins to accumulate (i.e. the prior bias when performance is at chance early in retrieval), β is the growth rate (HR) or the decay rate (FAR) and δ is the time point when the information begins to accumulate (i.e. performance departs from chance). HR and FAR data, averaged over participants were fit with the function separately for each strength condition.

In a nested model fitting routine, five models were fit to the group HR and FAR data with the MLE method, ranging from a 12-parameter full model ($4\lambda_1$–$2\lambda_2$–4β–2δ) to a 7-parameter asymptote-only model ($4\lambda_1$–λ_2–β–δ). In the full model, separate λ_1 parameters for asymptotic HR and FAR were allocated for each list-strength condition (4-λ_1). The asymptote during early retrieval when performance is at chance is assumed to be equal for HR and FAR (observed as d' = 0 in SAT functions), and thus λ_2 represents the prior bias. However, the prior bias at the beginning of retrieval was allowed to differ for each list-strength condition (2-λ_2). Then, four different rate parameters were allocated for each list-strength and item-type condition (4-β). Finally, the time point parameter was fixed for item type due to fixed prior bias, but allowed to vary across list-strength conditions (2-δ).

The model fit statistics suggest that the model could be reduced by fixing the growth rate parameter across list-strength and item-type conditions ($4\lambda_1$–$2\lambda_2$–β–2δ), and further by fixing the intercept parameter across list-strength conditions ($4\lambda_1$–$2\lambda_2$–β–δ). This shows that processing speed measures, both the rate parameters and the intercepts, do not differ significantly across strength conditions, consistent with the results from the SAT functions. This was also supported by a null effect of strength on δ observed from the fits to the individuals' HR and FAR data. The results of the best fitting model showed that separate asymptotes for the overall strength (λ_1) of HR and FAR was required to describe the SBME. Consistent with the mirror effect, the mean FAR of the strong foils (M = .20, SD = .1) was lower than the mean FAR of the weak foils (M = .33, SD = .21), and this difference approached significance, $t(8) = -1.75$, p = 0.06. On the other hand, mean HR for the strong targets (M = .79, SD = .14) was significantly greater than the mean HR for the weak targets (M = .66, SD = .21), $t(8) = 2.29$, p = .025. These results further support the necessity to allocate four separate λ_1 parameters for each strength and item type condition. In summary, the asymptotic memory availability is significantly different for both HR and FAR across strength conditions.

In order to further test whether list-strength has a significant effect on prior bias, HR and FAR data were fit with an asymptote-only model in which a common parameter for the prior bias (asymptote at early retrieval) was allocated along with four separate asymptote parameters ($4\lambda_1$–λ_2–β–δ). An increase in BIC, AICc and a decrease in Adjusted-R^2 values show that separate prior bias parameters are required to explain the group data, suggesting that participants have different tendencies towards the "no" response when tested on lists with different strength. A pair-wise comparison on the prior bias parameter (λ_2) observed from individuals' full model fit revealed a

Table 3
Parameter values of the best fitting exponential function of hit rates and the false alarm rates (Experiment 1).

Parameters		Average	Participants								
			1	2	3	4	5	6	7	8	9
λ_1	Strong HR	.79	.57	.89	.72	.89	.71	.90	.64	.79	.97
	Weak HR	.61	.28	.90	.56	.64	.50	.71	.58	.44	.73
	Strong FAR	.19	.14	.41	.09	.08	.30	.11	.19	.27	.21
	Weak FAR	.35	.22	.78	.23	.17	.30	.24	.42	.31	.40
λ_2	Strong	.53	.13	1.00	.42	.40	.62	.47	.44	.60	.70
	Weak	.43	.15	.87	.59	.07	.23	.57	.33	.48	.70
β		3.04	6.40	3.22	2.57	3.62	4.70	3.05	4.40	10.00	3.21
δ		0.44	0.53	0.44	0.44	0.43	0.45	0.48	0.52	0.46	0.60
Adj R^2		.972	.691	.813	.822	.914	.893	.794	.771	.715	.847

Note: Average parameter values are not the average of parameters obtained from participants' data, they are rather obtained from the fits to the data averaged across participants. λ_1 is the asymptotic probability to endorse a test item for a given item type (target or foil) and strength (strong or weak). λ_2 is the probability to endorse a test item when performance is at chance and indicates the prior bias at early retrieval. β is the rate of evidence accumulation towards asymptotic values and δ is the time point at which the performance departs from chance.

Table 4
Parameter values of the best fitting exponential function in Experiment 2.

Parameter	Average	Participants							
		1	2	3	4	5	6	7	8
λ_{strong}	2.18	1.51	2.27	1.58	2.48	1.76	2.96	2.31	2.34
λ_{weak}	1.34	0.81	0.86	1.05	2.34	1.05	1.31	1.07	1.35
β	3.58	1.03	0.36	1.95	3.04	5.31	4.92	4.77	1.38
δ	0.45	0.29	0.55	0.35	0.46	0.42	0.47	0.47	0.24
Adj R^2	.951	.588	.812	.856	.920	.690	.821	.808	.552

Note: Average parameter values are not the average of parameters obtained from participants' data, they are rather obtained from the fits to the data averaged across participants.

strength effect such that participants were more likely to endorse a strong test item ($M = .54$, $SD = .24$) compared with a weak test ($M = .42$, $SD = .28$) item when performance is at chance, $t(8) = 2.05$, $p = .04$. As Fig. 4 shows, when participants were tested on weakly encoded lists, they were more likely to respond "no" compared with being tested on the strongly encoded lists.

The parameters of the best model ($4\lambda_1$–$2\lambda_2$–β–δ) are presented in Table 3. The best fitting parameter values show the SBME in asymptotic HR and FAR, as λ_1 of strong HR was greater than λ_1 of weak HR and λ_1 of strong FAR was lower than the λ_1 of weak FAR. These findings suggest that the total available memory strength that is used to endorse a target and to reject a foil increases simultaneously as a function of encoding strength. The difference in the parameter estimate of λ_2 indicates a difference in prior bias across the strength conditions. For example, if participants were not biased towards any of the responses, the asymptote at early retrieval when performance is at chance, λ_2 would be .5. That is because participants would equally respond "yes" and "no" when memory evidence is not adequate to discriminate between targets and foils. However, λ_2 was .43 in the weak condition of the group data fit, indicating a tendency towards "no" response, as 57% of the responses were "no". In the strong condition of the group data fit, λ_2 was .53, showing that 53% of the responses were "yes" (see Table 3).

In summary, strengthening a list of items increases the total availability of information in memory but does not have a significant effect on the rate of information accrual. In addition to the SBME observed in asymptotic HR and FAR, the results showed that participants had a tendency to endorse probes in the strong lists, consistent with the more liberal starting point for the strong lists observed in the diffusion model applications (Criss, 2010; Starns, Ratcliff, et al., 2012). If the SBME observed in strong lists is due to more conservative evidentiary bias as stated by the criterion shift account, then these findings suggest contradictory effects of strengthening on the prior and the evidentiary bias. On the other hand, the differentiation account would not require such contradictory effects because in the differentiation models, the SBME is caused by lower memory strength of strong foils rather than meta-cognitive processes. In the following experiment, strength was manipulated in a mixed-list paradigm in which the differentiation account does not predict the SBME. Different from Starns, Ratcliff (2012) study, participants were not informed of the strength condition of the test list-strength. Thus, participants were not expected to adopt a more conservative drift criterion prior to testing. The goal of Experiment 2 was to test whether tes strength would have an effect on prior bias while null effect was expected on the evidentiary bias. Another goal was to investigate whether deep encoding would have similar strengthening effects as item repetition, which does not predict the SBME when participants were not informed.

Experiment 2

In Experiment 2, strength was manipulated in a mixed-list paradigm in which only half of the study items were strengthened via deep encoding. In half of the study-test cycles, only the strong targets were tested along with foils, and in the other half, only the weak targets were tested along with foils. Different from the Starns, Ratcliff, et al. (2012) study, participants were not informed of the strength condition at the beginning of the recognition task. In other words, participants were not given an explicit opportunity to change their strategy based on the strength condition of the test list before beginning the test phase. Thus, the drift criterion would not be expected to be different across strength conditions, as participants would not adopt different drift criterion. As mentioned earlier, differentiation of the foils in the strong condition would be similar to that of in the weak test lists, thus the SBME would not be predicted if deep encoding strengthens memory similar to item repetition. One other important goal of this experiment was to investigate whether prior bias would be affected even when the evidentiary bias was not expected to be different across strength conditions.

Method

Participants

Thirteen undergraduate students from Koç University took part in experiment in exchange for monetary compensation. Three participants dropped out of the experiment after the first session and two participants did not comply with the instructions, thus their data were not included in the subsequent analysis. Seven of the remaining 8 participants were female and only one of them was left-handed.

Materials, procedure and design

The materials and the procedure of this experiment were identical to those of Experiment 1. The only difference was that list-strength was manipulated only during test. During study, participants completed a pleasantness task for half of the items (70 words) as the strong condition and an orthographic judgment task for the other half (70 words) as the weak condition. Later, in two of the study-test cycles, only the strong targets were tested along with foils (strong foils) and in the remaining two cycles, only the weak targets were tested along with foils (weak foils). As in Experiment 1, participants completed two sessions.

The experiment was a 2 (Strength) $\times$ 7 (Response Lag) within-subjects design. There were 40 responses for targets and 40 responses for foils at each strength and lag condition. Lag condition was assigned randomly within each strength and test item type condition over the course of testing. After the removed trials, the number of responses for each strength and item condition was 35.93 ($SD = 4.81$), 38.00 ($SD = 2.39$), 38.90 ($SD = 1.40$), 38.94 ($SD = 1.58$), 38.75 ($SD = 1.07$), 38.25 ($SD = 1.13$) and 3753 ($SD = 2.05$) on average, for 60, 200, 300, 500, 800, 1500 and 3000 ms response lag conditions respectively.

Results and discussion

Accuracy

The exponential function fit comparison values of the group data are presented in Table 1. The list-strength effect is best explained by the 2λ–1β–1δ model that allocates separate parameters for asymptotic accuracy and common parameters for retrieval speed of the weak and the strong list conditions (Fig. 3B). In order to assess the parameter consistency across participants, the full model (2λ–2β–2δ) was fit to individual participants' data. Parameter estimates derived from the individual model fits indicate that asymptotic accuracy (λ) is significantly greater for the strong condition ($M = 2.23$, $SD = 0.48$) compared with the weak condition ($M = 1.26$, $SD = 0.69$), $t(7) = 3.89$, $p < .01$, supporting the conclusions derived from the model fit comparisons of the averaged data. Consistent with the best fitting model of the average data, the retrieval speed parameters (β and δ) did not differ significantly across list-strength conditions ($1/\beta$, $t = 1.40$; δ, $t = -1.22$). Parameter values of the best fitting model to group data and individual data are presented in Table 4.

Hit and false alarm analysis

Fig. 4B plots probability of endorsing a test item as a function of total processing time for each strength condition and item type (target vs. foil). A 7 (lag) $\times$ 2 (strength) repeated-measures ANOVA on HR showed a main effect of strength, $F(1,7) = 25.25$, $p < .01$, with HR increasing as a function of strength. A main effect of lag indicates that HR increased as a function of total processing time, $F(6,42) = 14.04$, $p < .001$. The interaction between list-strength and lag was not significant, suggesting that the HR of the strongly encoded items was greater than that of the weakly encoded items even when the total processing time was relatively short (e.g. 0.5 s). FAR analysis indicated a different pattern: A 7 (lag) $\times$ 2 (strength) repeated-measures ANOVA on FAR did not reveal a main effect of list-strength or a strength by lag interaction, but a significant main effect of lag, $F(6,42) = 11.93$, $p < .001$. That is, the FAR decreased as a function of total processing time but the SBME was not observed, as the asymptotic FAR was comparable over the course of retrieval (see Table 5).

To further investigate the retrieval dynamics of HR and FAR, the exponential function defined in Eq. (3) was fitted in a nested-models routine with models ranging from 12-parameter full model ($4\lambda_1$–$2\lambda_2$–4β–2δ) model to 6 parameter null model ($3\lambda_1$–λ_2–β–δ). The model fit statistics suggested that a 7-parameter asymptote-only model ($3\lambda_1$–$2\lambda_2$–β–δ) explained the strength effect in a mixed-list paradigm. Similar to the results from Experiment 1, separate λ_1 parameters were required for asymptotic HR across the two strength conditions. Statistical comparisons of the individual parameter estimates supported the asymptotic HR in the strong condition ($M = .86$, $SD = 0.13$) was significantly greater than the asymptotic HR in the weak condition ($M = .58$, $SD = 0.15$), $t(7) = 5.52$, $p < .001$. However, the strength effect on the asymptotic FAR failed to reach statistical significance, $t(7) = 0.90$.

The model fit statistics favored the model that allocates separate asymptotes early in retrieval (λ_2), which indicates a strength effect on the prior bias similar to the pattern observed in Experiment 1. The comparison of parameter estimates from the fits of individuals' data also showed that the asymptotic P("yes") at early retrieval was significantly greater for the strong condition ($M = .53$, $SD = 0.19$)

Table 5
Parameter values of the best fitting exponential function of hit rates and the false alarm rates (Experiment 2).

Parameters		Average	Participants							
			1	2	3	4	5	6	7	8
λ_1	Strong HR	.84	.74	.78	.79	.86	.89	.96	.96	.80
	Weak HR	.59	.60	.26	.55	.70	.68	.52	.76	.67
	FAR	.17	.30	.03	.18	.07	.29	.07	.28	.12
λ_2	Strong	.53	.73	.27	.50	.55	.75	.73	.49	.21
	Weak	.36	.48	.10	.49	.46	.59	.15	.46	.14
β		4.41	3.79	5.45	2.02	4.15	5.62	5.44	5.00	2.62
δ		0.46	0.48	0.52	0.36	0.47	0.43	0.47	0.45	0.40
Adj R^2		.974	.808	.939	.669	.809	.912	.935	.887	.831

Note: Average parameter values are not the average of parameters obtained from participants' data, they are rather obtained from the fits to the data averaged across participants. λ_1 is the asymptotic probability to endorse a test item for a given item type (target or foil) and strength (strong or weak). λ_2 is the probability to endorse a test item when performance is at chance and indicates the prior bias at early retrieval. β is the rate of evidence accumulation towards asymptotic values and δ is the time point at which the performance departs from chance.

Table 6
Parameter values of the best fitting exponential function in Experiment 3.

Parameter	Average	Participants									
		1	2	3	4	5	6	7	8	9	10
λ_{strong}	1.95	2.00	1.39	2.54	1.00	1.60	2.54	2.58	1.72	2.66	1.72
λ_{weak}	0.92	0.61	0.40	1.18	0.86	0.81	0.82	2.03	0.47	1.53	1.01
β	2.29	3.02	39.04	1.95	2.67	3.34	1.95	3.41	2.07	0.55	1.72
δ	0.41	0.49	0.58	0.27	0.25	0.53	0.38	0.47	0.43	0.63	0.31
Adj R^2	.984	.808	.696	.829	.591	.725	.858	.938	.708	.882	.752

Note: Average parameter values are not the average of parameters obtained from participants' data, they are rather obtained from the fits to the data averaged across participants..

Table 7
Parameter values of the best fitting exponential function of hit rates and the false alarm rates (Experiment 3).

Parameters		Average	Participants									
			1	2	3	4	5	6	7	8	9	10
λ_1	Strong HR	.73	.60	.48	.93	.56	.64	.93	.84	.81	.87	.57
	Weak HR	.47	.29	.13	.69	.28	.29	.56	.60	.51	.77	.54
	Strong FAR	.13	.02	.05	.11	.16	.14	.10	.07	.22	.24	.12
	Weak FAR	.19	.12	.06	.27	.09	.10	.25	.05	.35	.37	.24
λ_2	Strong	.41	.18	.06	.47	.33	.27	.68	.58	.60	.72	.21
	Weak	.25	.02	.02	.40	.01	.07	.26	.27	.50	.65	.25
β		2.54	3.11	10.00	2.09	3.78	3.24	2.47	4.37	2.91	1.30	4.15
δ		0.42	0.52	0.56	0.24	0.39	0.52	0.42	0.45	0.41	0.70	0.39
Adj R^2		.984	.691	.813	.822	.914	.893	.794	.771	.715	.847	

Note: Average parameter values are not the average of parameters obtained from participants' data, they are rather obtained from the fits to the data averaged across participants. λ_1 is the asymptotic probability to endorse a test item for a given item type (target or foil) and strength (strong or weak). λ_2 is the probability to endorse a test item when performance is at chance and indicates the prior bias at early retrieval. β is the rate of evidence accumulation towards asymptotic values and δ is the time point at which the performance departs from chance.

compared with the weak condition ($M = .37$, $SD = 0.19$), $t(7) = 2.83$, $p = 0.01$ (see Table 5).

In summary, strengthening items within lists increased the asymptotic accuracy and, similar to the results in Experiment 1, did not have a measurable effect on the speed of retrieval. Further, the FAR analysis showed a null effect of strength at the asymptote, suggesting that the total available memory evidence of the foils did not differ significantly across strength condition. Since the participants were not informed of the strength of the test list, they would not adopt a more stringent drift criterion when tested with strong lists. Similarly, the differentiation account would not predict lower FAR for the strong lists because the memory evidence of the foils would be comparable across test list-strength conditions. One interesting finding of this experiment was that participants had different prior bias across strength conditions even though they were not informed of which test condition they would receive. When participants were tested with strong targets, they had a tendency towards the "yes" response compared with the weak test condition. These strength effects on the prior bias will be further discussed later. In the following experiment, the effect of changing the strategy before test was further investigated by informing participants of the strength of the test list.

Experiment 3

The aim of Experiment 3 was to test the effect of informing participants about the strength of the targets on the SBME in a mixed-list paradigm. If changing the response strategy causes the SBME, then informing participants would result in a decrease in the asymptotic FAR contrary to the findings in Experiment 2.

Method

Participants

Fourteen undergraduate students from Koç University took part in experiment in exchange for monetary compensation. Two participants dropped out of the experiment after the first session and two participants did not comply with the instructions, thus their data were not included in the subsequent analysis. Six of the remaining 10 participants were male and all of the participants were right-handed.

Materials, procedure and design

The materials and the procedure of this experiment were identical to those of Experiment 2. The only difference was that participants were informed of the strength of the targets that they would be tested on before proceeding with the test list. For the weak test lists, participants were instructed that they would only be tested on words for which they had decided whether they contain the letter 'e'. For the strong test list, they were instructed that they would be tested only on the words for which they had made a pleasantness judgment.

The experiment was a 2 (Strength) × 7 (Response Lag) within-subjects design. There were 40 responses for targets and 40 responses for foils at each strength and lag condition. Lag condition was assigned randomly within each strength and test item type condition over the course of testing. After the removed trials, the number of responses for each strength and item condition was 37.55 (SD = 2.05), 38.80 (SD = 1.04), 39.32 (SD = 0.99), 39.05 (SD = 1.17), 38.75 (SD = 1.67), 37.75 (SD = 2.00) and 36.17 (SD = 2.66), on average, for 60, 200, 300, 500, 800, 1500 and 3000 ms response lag conditions respectively.

Results and discussion

Accuracy

Goodness of fit measures of the group data are presented in Table 1. The list-strength effect is best explained by the 2λ–1β–1δ model that allocates separate parameters for asymptotic accuracy and common parameters for retrieval speed of the weak and the strong list conditions (Fig. 3C). In order to assess the parameter consistency across participants, the full model (2λ–2β–2δ) was fit to individual participants' data. Parameter estimates derived from the individual model fits indicate that asymptotic accuracy (λ) is significantly greater for the strong condition (M = 1.84, SD = 0.50) compared with the weak condition (M = 0.96, SD = 0.42), $t(9) = 4.34$, $p < .01$, consistent with the model fit comparisons of the average data. Retrieval speed parameters (β and δ) also did not differ significantly across list-strength conditions ($1/\beta$, $t = -0.04$; δ, $t = -1.23$). Parameter values of the best fitting model to group data and individual data are presented in Table 6.

Hit and false alarm analysis

Fig. 4C plots the probability of endorsing a test item as a function of total processing time for each strength condition and item type (target vs. foil). A 7 (lag) × 2 (strength) repeated-measures ANOVA on HR showed a main effect of strength, $F(1,9) = 46.87$, $p < .001$, with HR increasing as a function of strength. A main effect of lag indicates that HR increased as a function of total processing time, $F(6,54) = 22.47$, $p < .001$. The list-strength by lag interaction was also significant, $F(6,54) = 2.61$, $p = 0.03$, indicating an increase in the strength effect later at retrieval. A 7 (lag) × 2 (strength) repeated-measures ANOVA on FAR revealed a main effect of lag $F(6,54) = 7.72$, $p < .001$, along with a strength by lag interaction, $F(6,54) = 10.67$, $p < .001$, but no main effect of list-strength. These results suggest that FAR was lower for the weak condition early in retrieval (e.g., until 600 ms), but later, this effect reversed such that the FAR became lower for the strong foils. However, this mirror effect was less effective (Lag 5, $t = -0.42$; Lag 6, $t = -1.95$; Lag 7, $t = -1.18$). This strength by lag interaction on FAR might suggest that strengthening targets have opposite effects over the course of retrieval. More specifically, when participants were tested on strong targets, they tended to be more liberal prior to retrieval and later they adopt a more stringent criterion when accumulating evidence.

To further investigate the retrieval dynamics of HR and FAR, the exponential function defined in Eq. (3) was fitted in a nested-models routine with models ranging from 12-parameter full model ($4\lambda_1$–$2\lambda_2$–4β–2δ) model to 6 parameter null model ($3\lambda_1$–λ_2–β–δ). The model fit statistics suggested that an 8-parameter asymptote-only model ($4\lambda_1$–$2\lambda_2$–β–δ) explained the strength effect best in a mixed-list paradigm where the participants were informed of the test list-strength. Similar to the results from previous experiments, separate λ_1 parameters were required for asymptotic HR across the two strength conditions. Statistical comparisons of the individual parameter estimates supported that the asymptotic HR in the strong condition (M = .75, SD = 0.18) was significantly greater than the asymptotic HR in the weak condition (M = .49, SD = 0.26), $t(9) = 4.19$, $p < .01$. Crucially, in contrast to results of Experiment 2, separate λ_1 parameters were also required for strong and weak asymptotic FAR. The comparisons of λ_1 from the fits of the individuals' data further supported the strength effect on the asymptotic FAR. The FAR of weak foils (M = .19, SD = 0.13) was significantly greater than the FAR of strong foils (M = .12, SD = 0.07), $t(9) = -2.22$, $p = .03$. This finding indicates a SBME at asymptotic FAR in contrast to the null strength effect observed from the analysis on the FAR. The λ_1 parameters of the FAR might represent a trend towards a SBME (see Table 7).

In order to investigate the prior bias, we next examined how strengthening affect asymptote at early retrieval (λ_2) that is the probability of endorsing the test item when performance is at chance. The model fit statistics favored the model that allocates separate λ_2, which indicates a strength effect on the prior bias similar to the pattern observed in the previous experiments. The comparison of parameter estimates from the fits of individuals' data also showed that the asymptotic p("yes") at early retrieval was significantly greater for the strong condition ($M = .42$, $SD = 0.23$) compared with the weak condition ($M = .26$, $SD = 0.23$), $t(9) = 3.63$, $p < 0.01$. Similar to the findings of the previous experiments, participants were less likely to respond with "yes" at a given accuracy level.

In summary, different from Experiment 2, the results of Experiment 3 showed that information regarding the strength condition of the test list created an opportunity for the participants to adopt a different strategy to endorse test probes. Considering the criterion shift account, these results suggest that strength has reverse effects on prior and evidentiary bias. When tested with strong targets, participants had a tendency towards the "yes" response early in retrieval. However, later in retrieval, this tendency becomes more conservative, meaning that participants require more evidence to endorse a test item. Differentiation models do not predict the SBME when items are strengthened in a mixed-list paradigm in contrast to the criterion shift models. Thus, a trend for the SBME observed in a mixed-list paradigm might suggest that the differentiation account is not necessary to explain the SBME observed in a pure-list paradigm. However, a comparison between the size of the SBME might suggest that a combination of the differentiation and the criterion shift account explains the greater SBME in the pure-list paradigm. In the following experiment, we tested this hypothesis by manipulating strength and study list condition in a single experiment and compared the SBME across the study list conditions.

Experiment 4

The goal of this experiment was to investigate the contribution of criterion shift and differentiation mechanisms in the list-strength effect observed in pure lists. In order to examine the size of the SBME caused by these two mechanisms, a standard "yes/no" item recognition task was administered in a pure- and mixed-list paradigm. Participants were randomly assigned to one of the list condition, but list strength was manipulated within participants. The SBME observed after studying a mixed-list can only be attributed to the criterion shift account, as the differentiation account does not predict a SBME when the study list is composed of mixed strength items. On the other hand, both accounts predict a SBME when participants are tested on the pure-study paradigm. Accordingly, if the SBME in the pure-list paradigm is due to both criterion shift and differentiation accounts, the size of the effect should be greater than the SBME observed in the mixed-list paradigm in which only the criterion shift predicts the SBME.

Method

Participants

Fifty-six undergraduate students from Koç University participated in the experiment in exchange for monetary compensation. Participants were randomly assigned to each of the study conditions in which targets were strengthened in either pure or mixed lists. Twenty-nine participants were tested in the pure-list condition and 27 participants were tested in the mixed-list condition.

Materials, procedure and design

The materials used in this experiment were identical to the materials used in the previous experiments. The only difference from Experiment 1 and 3 was that reaction time was measured rather than employing the response deadline procedure. The design was a 2 (study condition) × 2 (strength) mixed factorial, with study condition manipulated between subjects and item strength manipulated within subjects. In the pure-list condition, participants were presented with 140 words, 70 of which were targets and the other 70 were fillers presented randomly. In 2 of 4 study-test cycles, participants made a semantic judgment to strongly encode the words and in the remaining 2 cycles, participants made an orthographic judgment for shallow encoding. At test, participants were presented with targets along with 70 lures, and they were required to discriminate the targets from the lures. In the mixed-list condition, participants were asked to make a semantic judgment for 70 targets (strong) and an orthographic judgment for the remaining 70 words (weak). In 2 of 4 study-test cycles, only the strong targets were tested along with 70 lures, and in the remaining 2 cycles, only the weak targets were tested along with 70 lures. As in Experiment 3, participants were informed of the strength of the targets that they would be tested prior to the test in the mixed-list condition.

Results and discussion

A 2 (study condition: pure and mixed) × 2 (strength: strong and weak) mixed factor repeated measures ANOVA was conducted on d', hit rates, and false alarm rates. Discriminability was significantly better for strong items ($M = 2.19$, $SD = 0.79$) compared to weak items ($M = 0.94$, $SD = 0.39$), $F(1,54) = 196.72$, $p < .001$, $\eta_p^2 = .785$. Consistent with the null list-strength effect in recognition memory (Ratcliff et al., 1990), the main effect of study condition and the interaction between study condition and list-strength were unreliable. This confirms the finding that discriminability did not depend on whether the targets were strengthened within or across study lists.

The effect of strengthening targets produced a mirror effect in the pure condition but not when the target words were strengthened within study lists. Results from ANOVA conducted on hit rates revealed a significant main effect of strength, $F(1,54) = 262.68$, $p < .001$, $\eta_p^2 = .829$, showing that the hit rates were greater for the strong items ($M = .82$, $SD = 0.24$) when compared to the weak items ($M = .51$, $SD = 0.14$). No other effects on hit rates were reliable. We observed that the false alarm rate for the strong foils ($M = .15$, $SD = 0.12$) were significantly lower than those of weak foils ($M = .21$, $SD = 0.12$), $F(1,54) = 16.92$, $p < .001$, $\eta_p^2 = .239$. Central to our main question, the interaction between test list strength and study condition was significant, $F(1,54) = 7.12$, $p = .01$, $\eta_p^2 = .117$. This suggested that the strength effect was more prominent in the pure list condition ($M = -.09$, $SD = 0.08$), $t(28) = -5.63$, $p < .001$, whereas the false alarm rate did not differ significantly across foil strength when participants studied strong and weak targets in the same list ($M = -.02$, $SD = 0.11$), $t(26) = -0.89$. Fig. 5 presents the mean hit rate and the mean false alarm rate as a function of strength and study condition.

These results indicate that the SBME observed after studying a pure list is greater than the SBME observed in a mixed-list paradigm. We propose that both differentiation and criterion shift accounts contributed to the SBME in the pure-list paradigm, while in the mixed-list paradigm, the SBME was observed only due to a more conservative criterion. Thus, the results from this experiment support the contribution of the differentiation mechanism in the pure-list paradigm.

General discussion

The current study examined the full time-course of how SBME unfolds during retrieval for both pure and mixed-list paradigm, and provided the first investigation of how strengthening produces paradoxical effects on the two kinds of bias. To do so, a response-deadline speed-accuracy trade-off (SAT) procedure was employed in a list-strength paradigm. The items were strengthened via a levels-of-processing approach both in pure- and mixed-study list paradigm. Overall, our data implicated that strengthening a list of items increased the maximum level of accuracy, with no measurable impact on retrieval speed measures, either the rate of information accrual or when information first starts to accrue during retrieval. Additional analyses of the retrieval functions of targets and foils revealed that this enhancement in the maximum level of accuracy resulted from an increase in HR and a decrease in FAR in the pure-study list paradigm when performance saturates. In the mixed-list paradigm, the retrieval functions of foils produced different results depending on whether participants received information regarding the strength of the targets in the subsequent test list. Specifically, when participants were informed of the test list-strength, the

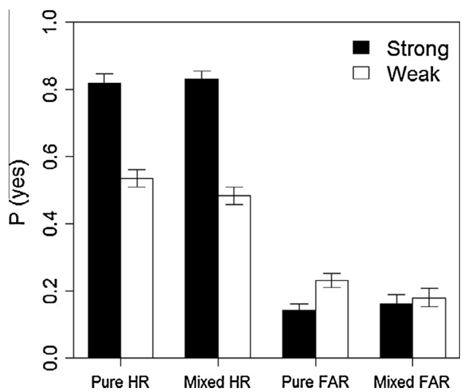

Fig. 5. Hit rates and false alarm rates as a function of study condition. Pure condition refers the pure-list paradigm in which items are strengthened across lists and mixed condition refers to the mixed-list paradigm in which items are strengthened within lists.

retrieval function of the foils preferred different asymptote parameters for strong and weak FARs, whereas the actual FAR of the responses following the longer response deadlines (e.g., 0.8–3 s) did not show a SBME. The asymptotic effect in the retrieval functions could potentially characterize a trend towards the SBME that was originally observed in Starns, Ratcliff, et al. (2012) data. However, an investigation of the size of the SBME across study-list conditions shows that the SBME in the pure-list paradigm is greater than the SBME in the mixed-list paradigm (cf. Starns, Ratcliff, et al., 2012). In line with the predictions of both the differentiation and the criterion shift account, when participants were not informed in the mixed-list paradigm, the asymptotic FAR was comparable across strength conditions. In addition to the implications regarding the theoretical explanation for the SBME, the results from the retrieval functions of HR and FAR indicate that strengthening a list of items produced reverse effects on the prior and the evidentiary bias. More specifically, participants were more liberal prior to evidence accumulation but they set a more conservative criterion to accumulate evidence in the strong test lists. Below we discuss the theoretical implications of these findings.

Prior and evidentiary bias

In addition to providing unbiased measures of speed and accuracy, response-deadline procedure has facilitated an empirical distinction between the two kinds of bias in item recognition, that is prior bias and evidentiary bias. Prior bias can be measured by the probability to endorse a probe earlier in retrieval when accuracy is at chance, while evidentiary bias is inferred from the mirror effect observed in HR and FAR when accuracy reaches an asymptote. Data from the response-deadline experiments in the current study reveal that testing only the strong targets caused participants to become more liberal in their responses when performance is at chance. This indicates that participants have a tendency to endorse a test item even before evidence starts to accumulate. This tendency towards the “yes” response has also been observed in the results from the diffusion model applications, as the starting point of the evidence accumulation was found to be closer to the “yes” boundary in strong test lists (e.g., Criss, 2010; Starns, Ratcliff, et al., 2012). The important finding from the response-deadline experiments is that this strength effect on the starting point is also observed with an experimental manipulation.

One interesting finding of the current study is that the prior bias is also evident in Experiment 2, which was the mixed-list paradigm when participants were not informed of the strength condition of the test lists. There could be two possible explanations regarding this tendency towards the “yes” response in strong lists, namely the sequential dependencies that are observed in recognition memory, and perceived strength of the test list. Malmberg and Annis (2012) showed that hits were more likely to be followed by hits than misses, and false alarms were more likely to be followed by false alarms than correct rejections. Regarding the current study, the increase in the frequency of "yes" responses in the strong lists might have caused a momentum for the "yes" response when accuracy was at chance early in retrieval. In other words, before information started to accumulate, participants might have repeated their previous response, which was more likely to be "yes" because the overall probability of a "yes" response was greater in the strong test lists. Alternatively, this tendency for the "yes" response could have been due to the perceived strength of the test list. In other words, participants might have recognized the strength of the test list even without being explicitly informed and consequently adopted a more conservative criterion. However, this does not explain why only the prior bias but not the evidentiary bias was affected from this perceived strength in Experiment 2.

Evidence for the evidentiary bias arises from the SBME observed in the mixed-list paradigm. More specifically, when accuracy reached asymptote in the mixed-list paradigm the hit rate increased and the false alarm rates decreased, especially when participants were informed of the strength condition. Although this finding does not directly measure the evidentiary bias, as it is the criterion that is set to moderate the amount of evidence that would be required to endorse the probe, the fact that a mirror effect is observed in a mixed-list paradigm suggests that evidentiary bias contributes to the SBME. In other words, in line with the criterion shift account, a more conservative evidentiary bias can explain the SBME observed when differentiation was controlled and participants were informed of the strength conditions.

The differentiation models

One potential mechanism that has been proposed to explain the SBME is the differentiation models (Criss & McClelland, 2006; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). According to these models, when items are strengthened during encoding, the foils become less similar to the targets and thus, they become less confusable at retrieval. That is achieved by updating the traces of targets in memory when items are strongly encoded. Thus, strong targets will have more complete memory traces and they will be more distinct from the foils at test. It has been suggested that the differentiation mechanism is in effect in other discrimination tasks (see Gibson & Gibson, 1955 for perceptual discrimination; see Wagenmakers et al., 2004 for lexical decision) and not particular to item recognition.

In the Retrieving Effectively from Memory model (REM, Shiffrin & Steyvers, 1997), items are represented as vectors of feature values that are drawn from the geometric distribution. The memory traces of the studied items are stored as vectors during encoding and these memory traces are noisy such that some of the feature values are encoded incorrectly or not encoded at all. When a probe is presented, it is compared to all of the traces in memory and a subjective likelihood is calculated for each comparison which is later averaged across all of the traces in memory. If this average likelihood (odds value) exceeds a criterion as in the signal detection framework, the probe is endorsed. Otherwise, it is rejected. When a list of items is strengthened during encoding, the memory traces of those items will be more complete as more feature values are encoded. Then, during test, foils will be less likely to match with the traces in memory and accordingly will produce lower odds value, which will be manifested as a decrease in the FAR. Targets, on the other hand, will match better with the more complete traces in memory and the odds value will be more likely to exceed the criterion, which will result in an increase in the HR (e.g., Criss, 2006).

Although REM was originally developed for accuracy measures, recent versions were adapted to account for reaction time data as well as full time-course retrieval functions derived from the response-deadline procedures (see Malmberg, 2008; Nobel & Shiffrin, 2001; see also Wagenmakers et al., 2004 for an adaptation to lexical decision). For example, Wagenmakers et al. (2004) interpreted the activation of a probe feature to develop as an exponential function of the total processing time, which would increase monotonically with a rate and reach an asymptote when full activation is achieved. It is also possible that the features of the memory traces rather than the features of the probe are activated over time (e.g., Malmberg, 2008). In either case, the memory performance increases over time by increasing the similarity between the memory traces and targets, and by decreasing the similarity between the memory traces and foils in an item recognition task. Thus, the differentiation mechanism accounts for the SBME by assuming more complete memory traces for strongly encoded items, and this would predict the strength effect only at asymptote not on the speed of retrieval.

The mirror effect observed with deep processing has been recently explained due to differences in retrieval strategy led by the differences in encoding mechanisms (Gallo et al., 2008; Scimeca et al., 2011). To be more specific, when participants were instructed to be tested on only deeply encoded items, they might have restricted their search set to those items and thus, a mirror effect could have been observed even in the mixed-list paradigm (cf. McDonough & Gallo, 2012). Such an explanation requires encoding of context features along with item features (e.g., REM.4 in Shiffrin & Steyvers, 1997). In REM.4, the set of items that were encoded within a similar context (e.g. deep encoding) are first activated, later the global match is determined between the probe and the activated set of memory traces. Thus, this explanation can account for the SBME even in a mixed list paradigm where items are strengthened via a levels-of-processing task similar to the studies reported by Gallo et al. (2008).

Alternatively, the mirror effect observed in a mixed-list paradigm after deep encoding could be explained by qualitative factors, which enhance memory by encoding of distinctive features (e.g., Gallo et al., 2008). Different from strengthening via repetition, encoding of qualitative factors could be modeled by traces composed of distinctive features rather than more complete memory traces. This mechanism has been used to explain the word frequency mirror effects, which refers to greater HR and lower FAR for low frequency items (Criss, 2010; Shiffrin & Steyvers, 1997). Thus, in a mixed-list paradigm, one could assume more distinctive traces for deeply encoded items and similarly this would cause a mirror effect only when items are studied in pure lists. In REM, distinctive features do not predict a mirror effect in the mixed-list paradigm consistent with the results from Experiment 2. It is also important to note that the distinctiveness of the features has been assumed to be a property of the stimuli whereas the strength effect has been assumed to be a property of encoding in REM (Criss, 2010; Shiffrin & Steyvers, 1997; Criss, Aue, & Kılıç, 2014).

The retrieval dynamics of the SBME can also depend on the methods that are used to strengthen the items. The results of the current study indicated that the retrieval speed was comparable for the items that were deeply encoded and the items that were encoded in the shallow condition. However, an earlier study by Dosher (1984) showed that different methods of strengthening produced different rates of evidence accumulation in an associative recognition task. Strengthening pairs by repetition resulted in faster retrieval compared to the pairs that were strengthened by study time in addition to increasing the asymptotic accuracy. Thus, it could be possible that strengthening by repetition or study time might have different effects on the retrieval dynamics of the SBME as was the case for the retrieval functions of associative recognition task. Thus, future research is required to explore the mechanisms that would account for the effects of strength on retrieval dynamics in general.

The criterion shift account

An alternative account, namely the criterion shift account, proposes that the SBME could be explained by more conservative responses for the strong lists (Benjamin & Bawa, 2004; Hirshman, 1995; Starns, Ratcliff, et al., 2012; Starns, White, et al., 2010, 2012; Stretch & Wixted, 1998; Verde & Rotello, 2007). When a list of items is encoded strongly, participants become aware of the increase in memory accuracy, and as a result, they require more evidence to endorse a test item. This metacognitive process is assumed to be a result of evidentiary bias, which refers to the bias on evidence accumulation. Starns, Ratcliff, et al., (2012) applied the diffusion model in a pure and a mixed-list strength paradigm where the results showed a mirror effect for the drift rate values of the strong targets and foils in line with the pattern observed for hit rates and false alarm rates. They suggested that this pattern could also observed due to a shift in the drift criterion, which is a result of a change in the evidentiary bias. The SBME observed in the pure-list paradigm cannot discriminate between the two accounts. However, the SBME observed in the mixed-list paradigm can only be explained by the criterion shift account (e.g., Starns, Ratcliff, et al., 2012). Consistent with this account, the results from the current study show the SBME in the pure-list paradigm and also a trend for the SBME in the mixed-list paradigm, when accuracy reaches an asymptote. These results indicate that when participants were explicitly informed of the strength condition of the test list, they tended to adopt a more stringent criterion so that they required more evidence to endorse the test item, once the total evidence has accumulated. A comparison of the size of the SBME observed in pure- and mixed-list paradigm suggests that the SBME is more prominent in the pure-list paradigm compared to that in the mixed-list paradigm. Accordingly, we suggest that both differentiation and the criterion shift account contribute to the SBME in the pure-list paradigm and only the criterion shift account causes the mirror effect in the mixed-list paradigm. Thus, a combination of the two mechanisms produce the greater effect observed in the pure-list paradigm.

Conclusions

In this study we evaluated the differential influence of strengthening memory on overall accuracy and processing speed measures. Our results implicate that the strength effect impacts only the total availability of information in memory, with no measurable influence on the processing speed estimates. As for the bias, strengthening items resulted in more liberal responses early in retrieval prior to accrual of evidence, and once evidence reaches its maximum, bias to endorse a strong probe becomes more conservative. Across the four experiments, the SBME was more prominent in the pure-list paradigm when compared with the informed condition of the mixed-list paradigm as predicted by the differentiation account. Taken together, these findings suggest that both the criterion shift and differentiation accounts jointly explain the SBME in recognition memory in the pure-list paradigm.

Acknowledgments

This research was supported by grants from FP7 Marie Curie IRG (PIRG08-GA-2010-277016), TÜBITAK 1001 (111K220), and a Science Academy Young Investigator Award (BAGEP) to I. Öztekin.

References

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language, 51*(2), 159–172. http://dx.doi.org/10.1016/J.Jml.04.001.

Benjamin, A. S., & Bjork, R. A. (2000). On the relationship between recognition speed and accuracy for words rehearsed via rote versus elaborative rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 638–648. http://dx.doi.org/10.1037/0278-7393.26.3.638.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684. http://dx.doi.org/10.1016/s0022-5371(72)80001-x.

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory & Language: Special Issue on Computational Models of Memory, 55*, 461–478.

Criss, A. H. (2009). The Distribution of Subjective Memory Strength: List Strength and Response Bias. *Cognitive Psychology*, 297–319.

Criss, A. H. (2010). Differentiation and response bias in episodic memory: Evidence from reaction time distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2), 484–499. http://dx.doi.org/10.1037/A0018435.

Criss, A. H., Aue, W. R., & Kılıç, A. (2014). Age and response bias: Evidence from the strength-based mirror effect. *The Quarterly Journal of Experimental Psychology*. http://dx.doi.org/10.1080/17470218.2013.874037.

Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language, 55*(4), 447–460. http://dx.doi.org/10.1016/j.jml.2006.06.003.

Criss, A. H., Wheeler, M. E., & McClelland, J. L. (2013). A differentiation account of recognition memory: Evidence from fMRI. *Journal of Cognitive Neuroscience, 25*, 421–435. http://dx.doi.org/10.1162/jocn_a_00292.

Dosher, B. A. (1981). The effects of delay and interference: A speed-accuracy study. *Cognitive Psychology, 13*(4), 551–582. http://dx.doi.org/10.1016/0010-0285(81)90020-7.

Dosher, B. A. (1984). Degree of learning and retrieval speed: Study time and multiple exposures. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 541–574.

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language, 58*, 1095–1111.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning; differentiation or enrichment? *Psychol Rev, 62*(1), 32–41.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*(1), 8–20. http://dx.doi.org/10.3758/Bf03198438.

Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika, 32*, 25–33.

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 846–858. http://dx.doi.org/10.1037/0278-7393.15.5.846.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments – evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33*(1), 1–18. http://dx.doi.org/10.1006/Jmla.1994.1001.

Hirshman, E. (1995). Decision-processes in recognition memory: Criterion shifts and the lists-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 302–313. http://dx.doi.org/10.1037/0278-7393.21.2.302.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12*(5), 852–857. http://dx.doi.org/10.3758/Bf03196776.

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language, 52*(4), 493–504. http://dx.doi.org/10.1016/J.Jml.01.007.

Liu, C. C., & Smith, P. L. (2009). Comparing time-accuracy curves: Beyond goodness-of-fit measures. *Psychonomic Bulletin & Review, 16*(1), 190–203. http://dx.doi.org/10.3758/PBR.16.1.190.

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57*(4), 335–384. http://dx.doi.org/10.1016/J.Cogpsych.02.004.

Malmberg, K. J., & Annis, J. (2012). On the Relationship between Memory and Perception: Sequential dependencies in recognition testing. *Journal of Experimental Psychology: General, 141*(2), 233–259.

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. http://dx.doi.org/10.1037/0033-295x.105.4.734-760.

McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation: Advances in Research and Theory, 46*, 155–200. http://dx.doi.org/10.1016/S0079-7421(06)46005-9.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term-memory: The time course of recognition. *Journal of Experimental Psychology: General, 118*(4), 346–373. http://dx.doi.org/10.1037//0096-3445.118.4.346.

McElree, B., & Dosher, B. A. (1993). Serial retrieval-processes in the recovery of order information. *Journal of Experimental Psychology: General, 122*(3), 291–315. http://dx.doi.org/10.1037/0096-3445.122.3. 291.

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 384–413. http://dx.doi.org/10.1037//0278-7393.27.2.384.

Öztekin, I., Gungor, N. Z., & Badre, D. (2012). Impact of aging on the dynamics of memory retrieval: A time-course analysis. *Journal of Memory and Language, 67*(2), 285–294. http://dx.doi.org/10.1016/j.jml.2012.05.003.

Öztekin, I., & McElree, B. (2007). Proactive interference slows recognition by eliminating fast assessments of familiarity. *Journal of Memory and Language, 57*(1), 126–149. http://dx.doi.org/10.1016/j.jml.2006.08.011.

Öztekin, I., & McElree, B. (2010). Relationship between measures of working memory capacity and the time course of short-term memory retrieval and interference resolution. *Journal of Experimental Psychology: Learning Memory and Cognition, 36*(2), 383–397. http://dx.doi.org/10.1037/A0018029.

R Core Team. (2012). R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing. <http://www.R-project.org>.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108.

Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review, 92*, 212–225.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(2), 163–178. http://dx.doi.org/10.1037//0278-7393.16.2.163.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review, 9*, 438–481.

Shiffrin, R. M., & Steyvers, M. (1997). Model for recognition memory: REM – Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*(2), 145–166. http://dx.doi.org/10.3758/Bf03209391.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50.

Starns, J. J., Ratcliff, R., & White, C. N. (2012). Diffusion model drift rates can be influenced by decision processes: An analysis of the strength-based mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1137–1151. http://dx.doi.org/10.1037/A0028151.

Starns, J. J., White, C. N., & Ratcliff, R. (2010). A direct test of the differentiation mechanism: REM, BCDMEM, and the strength-based mirror effect in recognition memory. *Journal of Memory and Language, 63*(1), 18–34. http://dx.doi.org/10.1016/j.jml.2010.03.004.

Starns, J. J., White, C. N., & Ratcliff, R. (2012). The strength-based mirror effect in subjective strength ratings: The evidence for differentiation can be produced without differentiation. *Memory & Cognition, 40*(8), 1189–1199. http://dx.doi.org/10.3758/s13421-012-0225-1.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1379–1396. http://dx.doi.org/10.1037//0278-7393.24.6.1379.

Tekcan, A. I., & Göz, I. (2005). *Türkçe Kelime Normları (Turkish word norms)*. Istanbul: Bogazici Universitesi Yayinevi.

Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition, 35*(2), 254–262. http://dx.doi.org/10.3758/Bf03193446.

Wagenmakers, E. J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology, 48*(3), 332–367. http://dx.doi.org/10.1016/J.Cogpsych.08.001.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*(1), 67–85. http://dx.doi.org/10.1016/0001-6918(77)90012-9.