Optimizing Human Learning

Behzad Tabibian[1,2], Utkarsh Upadhyay[1], Abir De[1], Ali Zarezade[3], Bernhard Schölkopf[2], and Manuel Gomez-Rodriguez[1]

[1]MPI for Software Systems, manuelgr@mpi-sws.org, utkarshu@mpi-sws.org, ade@mpi-sws.org
[2]MPI for Intelligent Systems Systems, btabibian@tue.mpg.de, bs@tue.mpg.de
[3]Sharif University, zarezade@ce.sharif.edu

Abstract

Spaced repetition is a technique for efficient memorization which uses repeated, spaced review of content to improve long-term retention. Can we find the optimal reviewing schedule to maximize the benefits of spaced repetition? In this paper, we introduce a novel, flexible representation of spaced repetition using the framework of marked temporal point processes and then address the above question as an optimal control problem for stochastic differential equations with jumps. For two well-known human memory models, we show that the optimal reviewing schedule is given by the recall probability of the content to be learned. As a result, we can then develop a simple, scalable online algorithm, MEMORIZE, to sample the optimal reviewing times. Experiments on both synthetic and real data gathered from Duolingo, a popular language-learning online platform, show that our algorithm may be able to help learners memorize more effectively than alternatives.

1 Introduction

Our ability to remember a piece of information depends critically on the number of times we have reviewed it and the time elapsed since the last review, as first shown by a seminal study by Ebbinghaus [10]. The effect of these two factors have been extensively investigated in the experimental psychology literature [16, 9], particularly in second language acquisition research [2, 5, 7, 21]. Moreover, these empirical studies have motivated the use of *flashcards*, small pieces of information a learner repeatedly reviews following a schedule determined by a *spaced repetition* algorithm [6], whose goal is to ensure that learners spend more (less) time working on forgotten (recalled) information.

In recent years, spaced repetition software and online platforms such as Mnemosyne[1], Synap[2], SuperMemo[3], or Duolingo[4] have become increasingly popular, often replacing the use of physical flashcards. The promise of these softwares and online platforms is that automated fine-grained monitoring and greater degree of control will result in more effective spaced repetition algorithms. However, most of these algorithms are simple rule-based heuristic with a few hard-coded parameters [6]—principled data-driven models and algorithms with provable guarantees have been largely missing until very recently [19, 22]. Among these recent notable exceptions, the work most closely related to ours is by Reddy et al. [22], who proposed a queueing network model for a particular spaced repetition method—the Leitner system [12] for reviewing flashcards—and then developed a heuristic approximation for scheduling reviews. However, their heuristic does not have provable guarantees, it does not adapt to the learner's performance over time, and it is specifically designed for Leitner systems.

[1]http://mnemosyne-proj.org/
[2]http://www.synap.ac
[3]https://www.supermemo.com
[4]http://www.duolingo.com

In this paper, we first introduce a novel, flexible representation of spaced repetition using the framework of marked temporal point processes [1]. For two well-known human memory models, we use this presentation to express the dynamics of a learner's forgetting rates and recall probabilities for the content to be learned by means of a set of stochastic differential equations (SDEs) with jumps. Then, we can find the optimal reviewing schedule for spaced repetition by solving a stochastic optimal control problem for SDEs with jumps [11, 25, 26]. In doing so, we need to introduce a proof technique of independent interest (refer to Appendices 8 and 9).

The solution uncovers a linear relationship between the optimal reviewing intensity and the recall probability of the content to be learned, which allows for a simple, scalable online algorithm, which we name MEMORIZE, to sample the optimal reviewing times (Algorithm 1). Finally, we experiment with both synthetic and real data gathered from Duolingo, a popular language-learning online platform, and show that our algorithm may be able to help learners memorize more effectively than alternatives. To facilitate research in this area within the machine learning community, we are releasing an implementation of our algorithm at http://learning.mpi-sws.org/memorize/.

Further related work. There is a rich literature which tries to ascertain which model of human memory predicts performance best [23, 24]. Our aim in this work is to provide a methodology to derive an optimal reviewing schedule given a choice of human memory model. Hence, we apply our methodology to two different memory models—exponential and power-law forgetting curve models.

The task of designing reviewing schedules also has a rich history, starting with the Leitner system itself [12]. In this context, Metzler-Baddeley et al. [18] have recently shown that adaptive reviewing schedules perform better than non-adaptive ones using data from SuperMemo. In doing so, they proposed an algorithm that schedules reviews just as the learner is *about* to forget an *item*, *i.e.*, when the probability of recall falls below a threshold. Lindsey et al. [14] have also used a similar heuristic for scheduling reviews, albeit with a model of recall inspired by ACT-R and Multiscale Context Model [20]. In this work, we use such heuristic as a baseline ("Threshold") in our experiments, with a memory model inspired by Settles et al. [23].

Finally, another line of research has pursued locally optimal scheduling by identifying which item would benefit the most from a review. Pavlik et al. [21] have used the ACT-R model to make locally optimal decisions about which item to review by greedily selecting the item which is closest to its maximum *learning rate* as a heuristic. Mettler et al. [17] have also employed a similar heuristic (ARTS system) to arrive at a reviewing schedule by taking response time into account. In this work, our goal is devising strategies which are globally optimal and allow for explicit bounds on the rate of reviewing.

2 Problem Formulation

In this section, we first briefly revisit two popular memory models we will use in our work. Then, we describe how to represent spaced repetition using the framework of marked temporal point processes. Finally, we conclude with a statement of the spaced repetition problem.

Modeling human memory. Following previous work in the psychology literature [10, 15, 24, 3], we consider the exponential and the power-law forgetting curve models with binary recalls (*i.e.*, a user either completely recalls or forgets an item).

The probability of recalling item i at time t for the exponential forgetting curve model is given by

$$m_i(t) := \mathbb{P}(r_i(t) = 1) = \exp\left(-n_i(t)(t - t_r)\right), \tag{1}$$

where t_r is the time of the last review and $n_i(t) \in \mathbb{R}^+$ is the forgetting rate[5] at time t, which may depend on many factors, *e.g.*, item difficulty and number of previous (un)successful recalls of the item. The probability of recalling item i at time t for the power-law forgetting curve model is given by

$$m_i(t) := \mathbb{P}(r_i(t) = 1) = (1 + \omega(t - t_r))^{-n_i(t)}, \tag{2}$$

[5]Previous works often use the inverse of the forgetting rate, referred as memory strength or half-life, $s(t) = n^{-1}(t)$ [22, 23]. However, it will be more tractable for us to work with the forgetting rates.

where t_r is the time of the last review, $n_i(t) \in \mathbb{R}^+$ is the forgetting rate and ω is a time scale parameter. Remarkably, despite their simplicity, the above functional forms have been recently shown to provide accurate quantitative predictions at a user-item level in large scale web data [22, 23].

In the remainder of the paper, for ease of exposition, we derive the optimal reviewing schedule and report experimental results only for the exponential forgetting curve model. Appendix 10 contains the derivation of the optimal reviewing schedule for the power-law forgetting curve model as well as its experimental validation.

Modeling spaced repetition. Given a learner who wants to memorize a set of items $\mathcal{I}$ using spaced repetition, *i.e.*, repeated, spaced reviews of the items, we represent each reviewing event as a triplet

$$e \coloneqq (\underset{\text{item}}{i}, \overset{\text{time}}{t}, \underset{\text{recall}}{r}),$$

which means that the learner reviewed item $i \in \mathcal{I}$ at time t and either recalled it ($r = 1$) or forgot it ($r = 0$). Here, note that each reviewing event includes the outcome of a test (*i.e.*, a recall) and this is a key difference from the paradigm used by several laboratory studies [7, 8], which consider a sequence of reviewing events followed by a single test. In other words, our data consists of TEST/REVIEW-...-TEST/REVIEW sequences, in contrast, the data in those studies consists of REVIEW-...-REVIEW-TEST sequences[6].

In the above representation, we model the recall r using the memory model defined by Eq. 1, *i.e.*, $r \sim \text{Bernoulli}(m_i(t))$, and we keep track of the reviewing times using a multidimensional counting process $\boldsymbol{N}(t)$, in which the i-th entry, $N_i(t)$, counts the number of times the learner has reviewed item i up to time t. Following the literature on temporal point processes [1], we characterize these counting processes using their corresponding intensities, *i.e.*, $\mathbb{E}[d\boldsymbol{N}(t)] = \boldsymbol{u}(t)dt$, and think of the recall r as their binary *marks*. Moreover, every time a learner reviews an item, the recall r has been experimentally shown to have an effect on the forgetting rate of the item [9, 22, 23]. In particular, using large scale web data from Duolingo, Settles et al. [23] have provided strong empirical evidence that (un)successful recalls of an item i during a review have a multiplicative effect on the forgetting rate $n_i(t)$—a successful recall at time t_r changes the forgetting rate by $(1 - \alpha_i)$, *i.e.*, $n_i(t) = (1 - \alpha_i)n_i(t_r)$, $\alpha_i \leq 1$, while an unsuccessful recall changes the forgetting rate by $(1 + \beta_i)$, *i.e.*, $n_i(t) = (1 + \beta_i)n_i(t_r)$, $\beta_i \geq 0$, where α_i and β_i are item specific parameters which can be found using historical data. In this context, the initial forgetting rate, $n_i(0)$, captures the difficulty of the item, with more difficult items having higher initial forgetting rates compared to easier items.

Hence, we express the dynamics of the forgetting rate $n_i(t)$ for each item $i \in \mathcal{I}$ using the following stochastic differential equation (SDE) with jumps:

$$dn_i(t) = -\alpha_i n_i(t) r_i(t) dN_i(t) + \beta_i n_i(t)(1 - r_i(t))dN_i(t), \tag{3}$$

where $N_i(t)$ is the corresponding counting process and $r_i(t) \in \{0, 1\}$ indicates whether item i has been successfully recalled at time t. Here, we would like to highlight that: (i) the forgetting rate, as defined above, is a Markov process and this will be useful in the derivation of the optimal reviewing schedule; (ii) the Leitner system [12] with exponential spacing can also be cast using this formulation with particular choices of α_i and β_i and the same initial forgetting rate, $n_i(0) = n(0)$, for all items; and, (iii) several laboratory studies, in which learners follow sequences REVIEW-...-REVIEW-TEST, suggest the parameters α_i and β_i should be time-varying since the retention rate follows an *inverted U-shape* [8], however, we found that in our dataset, in which learners follow sequences TEST/REVIEW-...-TEST/REVIEW, considering constant α_i and β_i is a valid approximation (refer to Appendix 13).

Given the above definition, one can also express the dynamics of the recall probability $m_i(t)$, defined by Eq. 1, by means of a SDE with jumps using the following Proposition (proven in Appendix 6):

Proposition 1 *Given an item $i \in \mathcal{I}$ with reviewing intensity $u_i(t)$, the recall probability $m_i(t)$, defined by Eq. 1, is a Markov process whose dynamics can be defined by the following SDE with jumps:*

$$dm_i(t) = -n_i(t)m_i(t)dt + (1 - m_i(t))dN_i(t), \tag{4}$$

[6]In most spaced repetition software and online platforms such as Mnemosyne, Synap, or Duolingo, the learner is tested in each review, *i.e.*, the learner follows TEST/REVIEW-...-TEST/REVIEW sequences.

where $N_i(t)$ is the counting process associated to the reviewing intensity $u_i(t)$.

Expressing the dynamics of the forgetting rates and recall probabilities as SDEs with jumps will be very useful for the design of our stochastic optimal control algorithm for spaced repetition.

The spaced repetition problem. Given a set of items $\mathcal{I}$, our goal is to find the optimal item reviewing intensities $\boldsymbol{u}(t) = [u_i(t)]_{i\in\mathcal{I}}$ that minimize the expected value of a particular convex loss function $\ell(\boldsymbol{m}(t), \boldsymbol{n}(t), \boldsymbol{u}(t))$ of the recall probability of the items, $\boldsymbol{m}(t) = [m_i(t)]_{i\in\mathcal{I}}$, the forgetting rates, $\boldsymbol{n}(t) = [n_i(t)]_{i\in\mathcal{I}}$, and the intensities themselves, $\boldsymbol{u}(t)$, over a time window $(t_0, t_f]$, *i.e.*,

$$\begin{aligned} &\underset{\boldsymbol{u}(t_0,t_f]}{\text{minimize}} \;\; \mathbb{E}_{(\boldsymbol{N},\boldsymbol{r})(t_0,t_f]}\left[\phi(\boldsymbol{m}(t_f), \boldsymbol{n}(t_f)) + \int_{t_0}^{t_f} \ell(\boldsymbol{m}(\tau), \boldsymbol{n}(\tau), \boldsymbol{u}(\tau))d\tau\right] \\ &\text{subject to} \;\; \boldsymbol{u}(t) \geq 0 \;\forall t \in (t_0, t_f), \end{aligned} \tag{5}$$

where $\boldsymbol{u}(t_0, t_f]$ denotes the item reviewing intensities from t_0 to t_f, the expectation is taken over all possible realizations of the associated counting processes and (item) recalls, denoted as $(\boldsymbol{N}, \boldsymbol{r})(t_0, t_f]$, the loss function is nonincreasing (nondecreasing) with respect to the recall probabilities (forgetting rates and intensities) so that it rewards long-lasting learning while limiting the number of item reviews, and $\phi(\boldsymbol{m}(t_f), \boldsymbol{n}(t_f))$ is an arbitrary penalty function. Finally, note that the forgetting rates $\boldsymbol{n}(t)$ and recall probabilities $\boldsymbol{m}(t)$, defined by Eq. 3 and Eq. 4, depend on the reviewing intensities $\boldsymbol{u}(t)$ we aim to optimize since $\mathbb{E}[d\boldsymbol{N}(t)] = \boldsymbol{u}(t)dt$.

3 The MEMORIZE Algorithm

In this section, we tackle the spaced repetition problem defined by Eq. 5 from the perspective of stochastic optimal control of jump SDEs [11]. More specifically, we first derive a solution to the problem considering only one item, provide an efficient practical implementation of the solution, and then generalize it to the case of multiple items.

Optimizing for one item. Given an item i with reviewing intensity $u_i(t) = u(t)$ and associated counting process $N_i(t) = N(t)$, recall outcome $r_i(t) = r(t)$, recall probability $m_i(t) = m(t)$ and forgetting rate $n_i(t) = n(t)$, we can rewrite the spaced repetition problem defined by Eq. 5 as:

$$\begin{aligned} &\underset{u(t_0,t_f]}{\text{minimize}} \;\; \mathbb{E}_{(N,r)(t_0,t_f]}\left[\phi(m(t_f), n(t_f)) + \int_{t_0}^{t_f} \ell(m(\tau), n(\tau), u(\tau))d\tau\right] \\ &\text{subject to} \;\; u(t) \geq 0 \;\forall t \in (t_0, t_f), \end{aligned} \tag{6}$$

where, using Eq. 3 and Eq. 4, the forgetting rate $n(t)$ and recall probability $m(t)$ is defined by the following two coupled jump SDEs:

$$\begin{aligned} dn(t) &= -\alpha n(t) r(t) dN(t) + \beta n(t)(1 - r(t))dN(t) \\ dm(t) &= -n(t)m(t)dt + (1 - m(t))dN(t) \end{aligned}$$

with initial conditions $n(t_0) = n_0$ and $m(t_0) = m_0$.

Next, we will define an optimal cost-to-go function J for the above problem, use Bellman's principle of optimality to derive the corresponding Hamilton-Jacobi-Bellman (HJB) equation [4], and exploit the unique structure of the HJB equation to find the optimal solution to the problem.

Definition 2 *The optimal cost-to-go $J(m(t), n(t), t)$ is defined as the minimum of the expected value of the cost of going from state $(m(t), n(t))$ at time t to the final state at time t_f.*

$$J(m(t), n(t), t) = \min_{u(t,t_f]} \mathbb{E}_{(N,r)(t,t_f]}\left[\phi(m(t_f), n(t_f)) + \int_{t}^{t_f} \ell(m(\tau), n(\tau), u(\tau))d\tau\right] \tag{7}$$

Algorithm 1: The MEMORIZE Algorithm

Input: Parameter q, α, β, n_0, t_f
Output: Next reviewing time t

$(u(t), n(t)) \leftarrow (q^{-1/2}, n_0)$;
$s \leftarrow Sample(u(t))$ // sample initial reviewing time
while $s < t_f$ **do**
  $r(s) \leftarrow ReviewItem(s)$ // review item, $r(s) \in \{0, 1\}$
  $n(t) \leftarrow (1 - \alpha)n(s)r(s) + (1 + \beta)n(s)(1 - r(s))$ // update forgetting rate
  $m(t) \leftarrow \exp(-(t - s)n(t))$ // update recall probability
  $u(t) \leftarrow q^{-1/2}(1 - m(t))$ // update reviewing intensity
  $s \leftarrow Sample(u(t))$ // sample next reviewing time
end while
return t ;

Now, we use Bellman's principle of optimality, which the above definition allows[7], to break the problem into smaller subproblems, and rewrite Eq. 7 as:

$$J(m(t), n(t), t) = \min_{u(t,t+dt]} \mathbb{E}[J(m(t + dt), n(t + dt), t + dt)] + \ell(m(t), n(t), u(t))dt$$
$$0 = \min_{u(t,t+dt]} \mathbb{E}[dJ(m(t), n(t), t)] + \ell(m(t), n(t), u(t))dt, \tag{8}$$

where $dJ(m(t), n(t), t) = J(m(t+dt), n(t+dt), t+dt) - J(m(t), n(t), t)$. Then, we differentiate J with respect to time t, $m(t)$ and $n(t)$ using the following Lemma (proven in Appendix 7).

Lemma 3 *Let $x(t)$ and $y(t)$ be two jump-diffusion processes defined by the following jump SDEs:*

$$dx(t) = f(x(t), y(t), t)dt + g(x(t), y(t), t)z(t)dN(t) + h(x(t), y(t), t)(1 - z(t))dN(t)$$
$$dy(t) = p(x(t), y(t), t)dt + q(x(t), y(t), t)dN(t)$$

where $N(t)$ is a jump process and $z(t) \in \{0, 1\}$. If function $F(x(t), y(t), t)$ is once continuously differentiable in $x(t)$, $y(t)$ and t, then,

$$dF(x, y, t) = (F_t + fF_x + pF_y)(x, y, t)dt + [F(x + g, y + q, t)z(t) + F(x + h, y + q, t)(1 - z(t)) - F(x, y, t)]dN(t),$$

where for notational simplicity we dropped the arguments of the functions f, g, h, p, q and argument of state variables.

Specifically, consider $x(t) = n(t)$, $y(t) = m(t)$, $z(t) = r(t)$ and $J = F$ in the above Lemma, then,

$$dJ(m, n, t) = J_t(m, n, t) - nmJ_m(m, n, t) + [J(1, (1 - \alpha)n, t)r + J(1, (1 + \beta)n, t)(1 - r) - J(m, n, t)]dN(t).$$

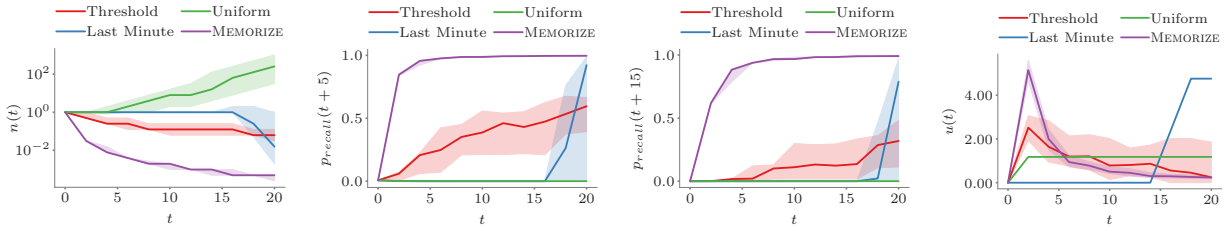
Then, if we substitute the above equation in Eq. 8, use that $\mathbb{E}[dN(t)] = u(t)dt$ and $\mathbb{E}[r(t)] = m(t)$, and rearrange terms, the HJB equation follows:

$$0 = J_t(m, n, t) - nmJ_m(m, n, t) + \min_{u(t,t+dt]} \big\{\ell(m, n, u)[J(1, (1-\alpha)n, t)m + J(1, (1+\beta)n, t)(1-m) - J(m, n, t)]u(t)\big\} \tag{9}$$

To solve the above equation, we need to define the loss ℓ. Following the literature on stochastic optimal control [4], we consider the following quadratic form, which is nonincreasing (nondecreasing) with respect to the recall probabilities (intensities) so that it rewards learning while limiting the number of item reviews:

$$\ell(m(t), n(t), u(t)) = \frac{1}{2}(1 - m(t))^2 + \frac{1}{2}qu^2(t). \tag{10}$$

[7] Bellman's principle of optimality readily follows using the Markov property of the recall probability $m(t)$ and forgetting rate $n(t)$.



(a) Forgetting rate (b) Short-term recall prob. (c) Long-term recall prob. (d) Reviewing intensity

Figure 1: Performance of MEMORIZE in comparison with several baselines. The solid lines are median values and the shadowed regions are 30% confidence intervals. Short-term recall probability corresponds to $m(t+5)$ and long-term recall probability to $m(t+15)$. In all cases, we use $\alpha = 0.5$, $\beta = 1$, $n(0) = 1$ and $t_f - t_0 = 20$. Moreover, we set $q = 3 \cdot 10^{-4}$ for MEMORIZE, $\mu = 0.6$ for the uniform reviewing schedule, $t_{lm} = 5$ and $\mu = 2.38$ for the last minute reviewing schedule, and $m_{th} = 0.7$ and $c = \zeta = 5$ for the threshold based reviewing schedule. Under these parameter values, the total number of reviewing events for all methods are equal (with a tolerance of 5%).

where q is a given parameter, which trade-offs recall probability and number of item reviews. This particular choice of loss function does not directly place a hard constraint on number of reviews—instead, it limits the number of reviews by penalizing high reviewing intensities.

Under these definitions, we can find the relationship between the optimal intensity and the optimal cost by taking the derivative with respect to $u(t)$ in Eq. 9:

$$u^*(t) = q^{-1} \left[J(m(t), n(t), t) - J(1, (1-\alpha)n(t), t)m(t) - J(1, (1+\beta)n(t), t)(1 - m(t)) \right]_+ .$$

Finally, we plug in the above equation in Eq. 9 and find that the optimal cost-to-go J needs to satisfy the following nonlinear differential equation:

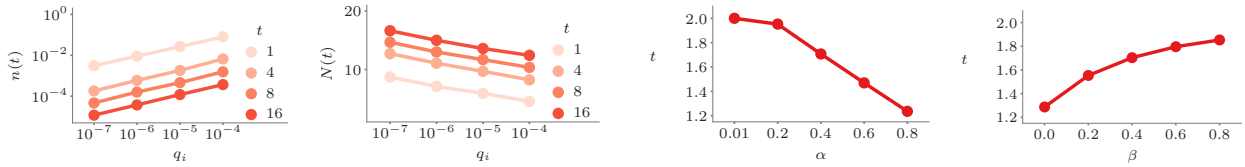
$$\begin{aligned} 0 = {} & J_t(m(t), n(t), t) - n(t)m(t)J_m(m(t), n(t), t) + \frac{1}{2}(1 - m(t))^2 \\ & - \frac{1}{2} q^{-1} \big[J(m(t), n(t), t) - J(1, (1-\alpha)n(t), t)m(t) - J(1, (1+\beta)n(t), t)(1 - m(t)) \big]_+^2 . \end{aligned} \tag{11}$$

with $J(m(t_f), n(t_f), t_f) = \phi(m(t_f), n(t_f))$ as terminal condition. To continue further, we rely on a technical Lemma (refer to Appendix 8), which derives the optimal cost-to-go J for a general family of losses ℓ. Using this Lemma, the optimal reviewing intensity is readily given by following Theorem (proven in Appendix 9):

Theorem 4 *Given a single item, the optimal reviewing intensity for the spaced repetition problem, defined by Eq. 6, under quadratic loss, defined by Eq. 10, is given by* $u^*(t) = q^{-1/2}(1 - m(t))$.

Note that the optimal intensity only depends on the recall probability, whose dynamics are given by Eqs. 3 and 4, and thus allows for a very efficient procedure to sample reviewing times. Algorithm 1 summarizes our sampling method, which we name MEMORIZE. Within the algorithm, $ReviewItem(s)$ returns the recall outcome $r(s)$ of an item review at time s, where $r(s) = 1$ indicates the item was recalled successfully and $r(s) = 0$ indicates it was not recalled, and $Sample(u(t))$ samples from an inhomogeneous poisson process with intensity $u(t)$ and it returns the sampled time. In practice, we sample from an inhomogeneous poisson process using a standard thinning algorithm [13].

Optimizing for multiple items. Given a set of items $\mathcal{I}$ with reviewing intensities $\boldsymbol{u}(t)$ and associated counting processes $\boldsymbol{N}(t)$, recall outcomes $\boldsymbol{r}(t)$, recall probabilities $\boldsymbol{m}(t)$ and forgetting rates $\boldsymbol{n}(t)$, we can solve the spaced repetition problem defined by Eq. 5 similarly as in the case of a single item.



(a) Learning effort

(b) Item difficulty and aptitude of the learner

Figure 2: Learning effort, aptitude of the learner and item difficulty. Panel (a) shows the average forgetting rate $n(t)$ and number of reviewing events $N(t)$ for different values of the parameter q, which controls the learning effort. Panel (b) shows the average time the learner takes to reach a forgetting rate $n(t) = \frac{1}{2}n(0)$ for different values of the parameters α and β, which capture the aptitude of the learner and the item difficulty. In Panel (a), we use $\alpha = 0.5$, $\beta = 1$, $n(0) = 1$ and $t_f - t_0 = 20$. In Panel (b), we use $n(0) = 20$ and $q = 0.02$. In both panels, error bars are too small to be seen.

More specifically, consider the following quadratic form for the loss ℓ:

$$\ell(m(t), n(t), u(t)) = \frac{1}{2}\sum_{i\in\mathcal{I}}(1 - m_i(t))^2 + \frac{1}{2}\sum_{i\in\mathcal{I}} q_i u_i^2(t).$$

where $\{q_i\}_{i\in\mathcal{I}}$ are given parameters, which trade-off recall probability and number of item reviews and may *favor* the learning of one item over another. Then, one can exploit the independence among items assumption to derive the optimal reviewing intensity for each item, proceeding similarly as in the case of a single item:

Theorem 5 *Given a set of items $\mathcal{I}$, the optimal reviewing intensity for each item $i \in \mathcal{I}$ in the spaced repetition problem, defined by Eq. 5, under quadratic loss is given by $u_i^*(t) = q_i^{-1/2}(1 - m_i(t))$.*

Finally, note that we can easily sample item reviewing times simply by running $|\mathcal{I}|$ instances of MEMORIZE (Algorithm 1), one per item.

4 Experiments

4.1 Experiments on synthetic data

In this section, our goal is analyzing the performance of MEMORIZE under a controlled setting using metrics and baselines that we cannot compute in the real data we have access to.

Experimental setup. We evaluate the performance of MEMORIZE using two quality metrics: recall probability $m(t + \tau)$ at a given time in the future $t + \tau$ and forgetting rate $n(t)$. Here, by considering high (low) values of τ, we can assess long-term (short-term) retention. Moreover, we compare the performance of our method with three baselines: (i) a *uniform* reviewing schedule, which sends item(s) for review at a constant rate μ; (ii) a *last minute* reviewing schedule, which only sends item(s) for review during a period $[t_{lm}, t_f]$, at a constant rate μ therein; and (iii) a *threshold* based reviewing schedule, which increases the reviewing intensity of an item by $c \exp((t - s)/\zeta)$ at time s, when its recall probability reaches a threshold m_{th}. The threshold baseline is similar to the heuristics proposed by Metzler-Baddeley et al. [18] and Lindsey et al. [14]. We do not compare with the algorithm proposed by Reddy et al. [22] because, as it is specially designed for Leitner system, it assumes a discrete set of forgetting rate values and, as a consequence, is not applicable to our (more general) setting. Unless otherwise stated, we set the parameters of the baselines and our method such that the total number of reviewing events during $(t_0, t_f]$ are equal.

Solution quality. For each method, we run 100 independent simulations and compute the above quality metrics over time. Figure 1 summarizes the results, which show that our model: (i) consistently outperforms all the baselines in terms of both quality metrics; (ii) is more robust across runs both in terms of quality

metrics and reviewing schedule; and (iii) reduces the reviewing intensity as times goes by and the recall probability improves, as one could have expected.

Learning effort. The value of the parameter q controls the learning effort required by MEMORIZE—the lower its value, the higher the number of reviewing events. Intuitively, one may also expect the learning effort to influence how quickly a learner memorizes a given item—the lower its value, the quicker a learner will memorize it. Figure 2a confirms this intuition by showing the average forgetting rate $n(t)$ and number of reviewing events $N(t)$ at several times t for different q values.

Aptitude of the learner and item difficulty. The parameters α and β capture the aptitude of a learner and the difficulty of the item to be learned—the higher (lower) the value of α (β), the quicker a learner will memorize the item. In Figure 2b, we evaluate quantitatively this effect by means of the average time the learner takes to reach a forgetting rate of $n(t) = \frac{1}{2}n(0)$ using MEMORIZE for different parameter values.

4.2 Experiments on real data

In this section, our goal is to evaluate how *well* each reviewing schedule spaces the reviews leveraging a real dataset[8]. Unlike the synthetic experiments, we cannot intervene and determine *what would have happened* if a user would follow MEMORIZE or any of the baselines in the real dataset. As a consequence, measuring the performance of different algorithms is more challenging. We overcome this difficulty by relying on likelihood comparisons to determine *how closely* a (user, item) pair followed a particular reviewing schedule and compute quality metrics that do not depend on the choice of memory model.

Datasets description. We use data gathered from Duolingo, a popular language-learning online platform[9]. This dataset consists of ∼12 million *sessions* of study, involving ∼5.3 million unique (user, word) pairs, which we denote by $\mathcal{D}$, collected over the period of two weeks. In a single session, a user answers multiple questions, each of which contains multiple words. Each word maps to an item i and the fraction of correct recalls of sentences containing a word i in the session is used as an estimate of its recall probability at the time of the session, as in previous work [23]. If a word is recalled perfectly during a session then it is considered as a successful recall, *i.e.*, $r_i(t) = 1$, and otherwise it is considered as an unsuccessful recall, *i.e.*, $r_i(t) = 0$. Since we can only expect the estimation of the model parameters to be accurate for users and items with enough number of reviewing events, we only consider users with at least 30 reviewing events and words that were reviewed at least 30 times. After this preprocessing step, our dataset consists of ∼5.2 million unique (user, word) pairs.

Experimental setup and methodology. As pointed out previously, we cannot intervene in the real datasets and thus rely on likelihood comparisons to determine *how closely* a (user, item) pair followed a particular reviewing schedule. More in detail, we proceed as follows.

First, we estimate the parameters α and β using half-life regression[10], where we fit a single set of parameters for all items, but a different initial forgetting rate $n_i(0)$ per item (refer to Appendix 12 for more details). Then, for each user, we use maximum likelihood estimation to fit the parameter q in MEMORIZE and the parameter μ in the uniform reviewing schedule. For the threshold based schedule, we fit one set of parameters for all users using maximum likelihood estimation for the parameter c and grid search for the parameter ζ.

Then, we compute the likelihood of the times of the reviewing events for each (user, item) pair under the intensity given by MEMORIZE, *i.e.*, $u(t) = q^{-1/2}(1 - m(t))$, the intensity given by the uniform schedule, *i.e.*, $u(t) = \mu$, and the intensity given by the threshold based schedule, *i.e.*, $u(t) = c\exp((t-s)/\zeta)$. The likelihood $LL(\{t_i\})$ of a set of reviewing events $\{t_i\}$ given an intensity function $u(t)$ can be computed as follows [1]:

$$LL(\{t_i\}) = \sum_i \log u(t_i) - \int_0^T u(t)\, dt.$$

[8]Note that it is not the objective of this paper to evaluate the predictive power of the underlying memory models, we are relying on previous work for that [23, 24]. However, for completeness, we provide a series of benchmarks and evaluations for the models we used in this paper in Appendix 12.

[9]The dataset is available at https://github.com/duolingo/halflife-regression.

[10]Half-life h is the inverse of our forgetting rate $n(t)$ multiplied by a constant.

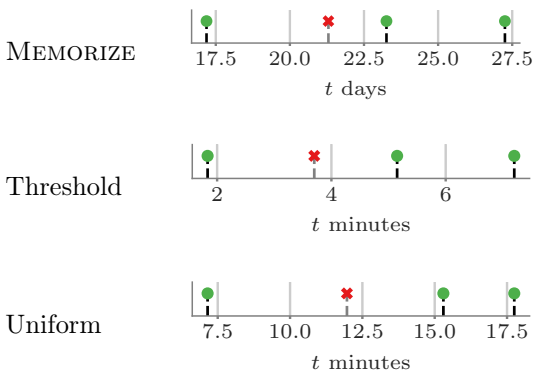


Figure 3: Examples of (user, item) pairs whose corresponding reviewing times have high likelihood under MEMORIZE (top), uniform reviewing schedule (bottom) and threshold based reviewing schedule (middle). In every figure, each candle stick corresponds to a reviewing event with a green circle (red cross) if the recall was (un)successful, and time $t = 0$ corresponds to the first time the user is exposed to the item in our dataset, which may or may not correspond with the first reviewing event. The pairs whose reviewing times follow more closely MEMORIZE or the threshold based schedule tend to increase the time interval between reviews every time a recall is successful while, in contrast, the uniform reviewing schedule does not. MEMORIZE tends to space the reviews more than the threshold based schedule, achieving the same recall pattern with less effort.

This allows us to determine how closely a (user, item) pair follows a particular reviewing schedule[11], as shown in Figure 3. Distribution of the likelihood values under each reviewing schedule is provided in Appendix 12. We do not compare to the last minute baseline since in Duolingo there is no terminal time t_f which users target. Additionally, in many (user, item) pairs, the first review takes place close to $t = 0$ and thus the last minute baseline is equivalent to the uniform reviewing schedule.

Finally, since measurements of the future recall probability $m(t + \tau)$ are not forthcoming and depend on the memory model of choice, we concentrate on the following alternative quality metrics, which do not depend on the particular choice of memory model:

(a) *Effort*: for each (user, item), we measure the effort by means of the empirical estimate of the inverse of the total reviewing period, *i.e.*, $\hat{e} = 1/(t_n - t_1)$. The lower the effort, the less burden on the user, allowing her to learn more items simultaneously.

(b) *Empirical forgetting rate*: for each (user, item), we compute an empirical estimate of the forgetting rate by the time t_n of the last reviewing event, *i.e.*, $\hat{n} = -\log(\hat{m}(t_n))/(t_n - t_{n-1})$. Here, note that the estimate of the forgetting rate only depends on the observed data (not model/methods parameters). For a more fair comparison across items, we normalize each empirical forgetting rate using the average empirical initial forgetting rate of the corresponding item at the beginning of the observation window, *i.e.*, for an item i, $\hat{n}_0 = |\{u : (u, i) \in \mathcal{D}\}|^{-1} \sum_{u:(u,i)\in\mathcal{D}} \hat{n}_{0,u}$ where $\hat{n}_{0,u} = -\log(\hat{m}(t_{u,1}))/(t_{u,1} - t_{u,0})$.

Given a particular recall pattern, the lower the above quality metrics, the more effective the reviewing schedule.

Results. We first group (user, item) pairs by their recall pattern, *i.e.*, the sequence of successful ($r = 1$) and unsuccessful ($r = 0$) recalls over time—if two pairs have the same recall pattern, then they have the same number of reviews and changes in their forgetting rates $n(t)$. For each recall pattern in our observation window, we pick the top 25% pairs in terms of likelihood for each method and compute the average effort and empirical forgetting rate, as defined above. Figure 4 summarizes the results for the most common recall

[11] Duolingo uses hand-tuned spaced repetition algorithms, which propose reviewing times to the users and, thus, the reviewing schedule is expected to be close to the one recommended by Duolingo. However, since users often do not perform reviews exactly at the recommended times, some pairs will be closer to uniform than threshold or MEMORIZE and viceversa.

	Effort, $\hat{e}$		Emp. forgetting, $\hat{n}$	
Pattern	$\frac{\hat{e}_M}{\hat{e}_T}$	$\frac{\hat{e}_M}{\hat{e}_U}$	$\frac{\hat{n}_M}{\hat{n}_T}$	$\frac{\hat{n}_M}{\hat{n}_U}$
●●●	0.00*	0.01*	0.11*	0.11*
●●●●	0.01*	0.04*	0.14*	0.15*
●●●●●	0.02*	0.07*	0.21*	0.22*
×●●	0.00*	0.03*	0.10*	0.10*
×●●●	0.03*	0.06*	0.13*	0.14*
××●●	0.03*	0.05*	0.11*	0.12*
×××●●	0.03*	0.05*	0.12*	0.12*

Figure 4: Performance of MEMORIZE (M) in comparison with the uniform (U) and threshold (T) based reviewing schedules. Each row corresponds to a different recall pattern, depicted in the first column, where markers denote ordered recalls, green circles indicate successful recall and red crosses indicate unsuccessful ones. In the second column, each cell value corresponds to the ratio between the average effort for the top 25% pairs in terms of likelihood for the uniform (or threshold based) schedule and MEMORIZE. In the right column, each cell value corresponds to the ratio between the median empirical forgetting rate for the same pairs for the uniform (or threshold based) schedule and MEMORIZE. In both metrics, if the ratio is smaller than 1, MEMORIZE is more effective than the uniform (or threshold based) schedule for the corresponding pattern. The symbol $*$ indicates whether the change is significant with p-value < 0.01 using the Kolmogrov-Smirnov 2-sample test.

patterns[12], where we report the ratio between the effort and empirical forgetting rate values achieved by the uniform and threshold based reviewing schedules and the values achieved by MEMORIZE. That means, if the reported value is smaller than 1, MEMORIZE is more effective for the corresponding pattern. We find that both in terms of effort and empirical forgetting rate, MEMORIZE outperforms the uniform and threshold based reviewing schedules for all recall patterns. For example, for the recall pattern consisting of two unsuccessful recalls followed by two successful recalls (red-red-green-green), MEMORIZE achieves $0.05\times$ lower effort and $0.12\times$ lower empirical forgetting rate than the second competitor.

Next, we group (user, item) pairs by the number of reviews during a fixed period of time, *i.e.*, we control for the effort, pick the top 25% pairs in terms of likelihood for each method and compute the average empirical forgetting rate. Figure 5 summarizes the results for sequences with up to seven given reviews since the beginning of the observation window, where lower values indicate better performance. The results show that MEMORIZE offers a competitive advantage with respect to the other baselines, which is statistically significant.

5 Conclusions

In this paper, we have first introduced a novel representation of spaced repetition using the framework of marked temporal point processes and SDEs with jumps and then designed a framework that exploits this novel representation to cast the design of spaced repetition algorithms as a stochastic optimal control problem of such SDEs. For ease of exposition, we have considered only two memory models, exponential and power-law forgetting curves, and a quadratic loss function, however, our framework is agnostic to this particular modeling choices and it provides a set of novel techniques to find reviewing schedules that are optimal *under* a given choice of memory model and loss. We experimented on both synthetic and real data gathered from Duolingo, a popular language-learning online platform, and showed that our framework may be able to help learners memorize more effectively than alternatives.

There are many interesting directions for future work. For example, it would be interesting to perform large scale interventional experiments to assess the performance of our algorithm in comparison with existing

[12]Results are qualitatively similar for other recall patterns.

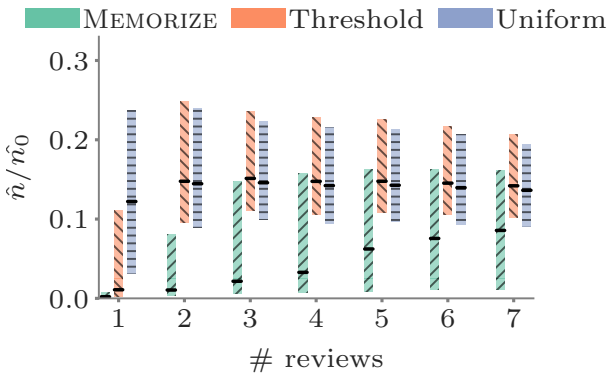


Figure 5: Average empirical forgetting rate for the top 25% pairs in terms of likelihood for MEMORIZE, the uniform reviewing schedule and the threshold based reviewing schedule for sequences with different number of reviews. Boxes indicate 25% and 75% quantiles and solid lines indicate median values, where lower values indicate better performance. In all cases, the competitive advantage MEMORIZE achieves is statistically significant (p-value < 0.01 using the Kolmogrov-Smirnov 2-sample test).

spaced repetition algorithms deployed by, *e.g.*, Duolingo. Moreover, in our work, we consider a particular quadratic loss, however, it would be useful to derive optimal reviewing intensities for other (non-quadratic) losses capturing particular learning goals. We assumed that, by reviewing an item, one can only influence its recall probability and forgetting rate. However, items may be dependent and thus, by reviewing an item, one can influence the recall probabilities and forgetting rates of several items. Finally, it would be very interesting to allow for reviewing events to be composed of groups of items, some reviewing times to be preferable over others, and then derive both the optimal reviewing schedule and optimal grouping of items.

References

[1] O. Aalen, O. Borgan, and H. K. Gjessing. *Survival and event history analysis: a process point of view*. Springer, 2008.

[2] R. C. Atkinson. Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology*, 96(1):124, 1972.

[3] L. Averell and A. Heathcote. The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*, 55(1), 2011.

[4] D. P. Bertsekas. *Dynamic programming and optimal control*. Athena Scientific Belmont, MA, 1995.

[5] K. C. Bloom and T. J. Shuell. Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, 74(4):245–248, 1981.

[6] G. Branwen. Spaced repetition. https://www.gwern.net/Spaced%20repetition, 2016.

[7] N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin*, 132(3):354, 2006.

[8] N. J. Cepeda, E. Vul, D. Rohrer, J. T. Wixted, and H. Pashler. Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological science*, 19(11):1095–1102, 2008.

[9] F. N. Dempster. Spacing effects and their implications for theory and practice. *Educational Psychology Review*, 1(4):309–330, 1989.

[10] H. Ebbinghaus. *Memory: a contribution to experimental psychology.* Teachers College, Columbia University, 1885.

[11] F. B. Hanson. *Applied stochastic processes and control for Jump-diffusions: modeling, analysis, and computation.* SIAM, 2007.

[12] S. Leitner. *So lernt man lernen.* Herder, 1974.

[13] P. A. Lewis and G. S. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval research logistics quarterly*, 26(3):403–413, 1979.

[14] R. V. Lindsey, J. D. Shroyer, H. Pashler, and M. C. Mozer. Improving students' long-term knowledge retention through personalized review. *Psychological science*, 25(3):639–647, 2014.

[15] G. R. Loftus. Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(2), 1985.

[16] A. W. Melton. The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, 9(5):596–606, 1970.

[17] E. Mettler, C. M. Massey, and P. J. Kellman. A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology: General*, 145(7):897, 2016.

[18] C. Metzler-Baddeley and R. J. Baddeley. Does adaptive training work? *Applied Cognitive Psychology*, 23(2):254–266, 2009.

[19] T. P. Novikoff, J. M. Kleinberg, and S. H. Strogatz. Education of a model student. *PNAS*, 109(6):1868–1873, 2012.

[20] H. Pashler, N. Cepeda, R. V. Lindsey, E. Vul, and M. C. Mozer. Predicting the optimal spacing of study: A multiscale context model of memory. In *Advances in neural information processing systems*, pages 1321–1329, 2009.

[21] P. I. Pavlik and J. R. Anderson. Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, 14(2):101, 2008.

[22] S. Reddy, I. Labutov, S. Banerjee, and T. Joachims. Unbounded human learning: Optimal scheduling for spaced repetition. In *KDD*, 2016.

[23] B. Settles and B. Meeder. A trainable spaced repetition model for language learning. In *ACL*, 2016.

[24] J. T. Wixted and S. K. Carpenter. The wickelgren power law and the ebbinghaus savings function. *Psychological Science*, 18(2):133–134, 2007.

[25] A. Zarezade, A. De, H. Rabiee, and M. Gomez-Rodriguez. Cheshire: An online algorithm for activity maximization in social networks. In *arXiv preprint arXiv:1703.02059*, 2017.

[26] A. Zarezade, U. Upadhyay, H. Rabiee, and M. Gomez-Rodriguez. Redqueen: An online algorithm for smart broadcasting in social networks. In *WSDM*, 2017.

Appendix

6 Proof of Proposition 1

According to Eq. 1, the recall probability $m(t)$ depends on the forgetting rate, $n(t)$, and the time elapsed since the last review, $D(t) := t - t_r$. Moreover, we can readily write the differential of $D(t)$ as $dD(t) = dt - D(t)dN(t)$.

We define the vector $\boldsymbol{X}(t) = [n(t), D(t)]^T$. Then, we use Eq. 3 and Itö's calculus [11] to compute its differential:

$$\begin{aligned} d\boldsymbol{X}(t) &\overset{\text{dt}}{=}_{zol} \boldsymbol{f}(\boldsymbol{X}(t), t)dt + \boldsymbol{h}(\boldsymbol{X}(t), t)dN(t) \\ \boldsymbol{f}(\boldsymbol{X}(t), t) &= \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right] \\ \boldsymbol{h}(\boldsymbol{X}(t), t) &= \left[\begin{smallmatrix} -\alpha n(t) r(t) + \beta n(t)(1 - r(t)) \\ -D(t) \end{smallmatrix}\right] \end{aligned}$$

Finally, using again Itö's calculus and the above differential, we can compute the differential of the recall probability $m(t) = e^{-n(t)D(t)} := F(\boldsymbol{X}(t))$ as follows:

$$\begin{aligned} dF(\boldsymbol{X}(t)) &= F(\boldsymbol{X}(t + dt)) - F(\boldsymbol{X}(t)) \\ &= F(\boldsymbol{X}(t) + d\boldsymbol{X}(t)) - F(\boldsymbol{X}(t)) \\ &\overset{\text{dt}}{=}_{zol} (f^T F_{\boldsymbol{X}}(\boldsymbol{X}(t)))dt + F(\boldsymbol{X}(t) + \boldsymbol{h}(\boldsymbol{X}(t), t)dN(t)) - F(\boldsymbol{X}(t)) \\ &\overset{\text{dt}}{=}_{zol} (f^T F_{\boldsymbol{X}}(\boldsymbol{X}(t)))dt + (F(\boldsymbol{X}(t) + h(\boldsymbol{X}(t), t)) - F(\boldsymbol{X}(t)))dN(t) \\ &= (e^{-(D(t)-D(t))n(t)(1+\alpha r_i(t) - \beta(1-r_i(t)))} - e^{-D(t)n(t)})dN(t) - n(t)e^{-D(t)n(t)}dt \\ &= -n(t)e^{-D(t)n(t)}dt + (1 - e^{-D(t)n(t)})dN(t) \\ &= -n(t)F(X(t))dt + (1 - F(X(t)))dN(t) \\ &= -n(t)m(t)dt + (1 - m(t))dN(t). \end{aligned}$$

7 Proof of Lemma 3

According to the definition of differential,

$$\begin{aligned} dF := dF(x(t), y(t), t) &= F(x(t + dt), y(t + dt), t + dt) - F(x(t), y(t), t) \\ &= F(x(t) + dx(t), y(t) + dy(t), t + dt) - F(x(t), y(t), t). \end{aligned}$$

Then, using Itö's calculus, we can write

$$\begin{aligned} dF &\overset{\text{dt}}{=}_{zol} F(x + fdt + g, y + pdt + q, t + dt)dN(t)z + F(x + fdt + h, y + pdt + q, t + dt)dN(t)(1 - z) \\ &+ F(x + fdt, y + pdt, t + dt)(1 - dN(t)) - F(x, y, t) \end{aligned} \tag{12}$$

where for notational simplicity we drop arguments of all functions except F and dN. Then, we expand the first three terms:

$$\begin{aligned} F(x + fdt + g, y + pdt + q, t + dt) &= F(x + g, y + q, t) + F_x(x + g, y + q, t)fdt + F_y(x + g, y + q, t)pdt \\ &+ F_t(x + g, y + q, t)dt \\ F(x + fdt + h, y + pdt + q, t + dt) &= F(x + h, y + q, t) + F_x(x + h, y + q, t)fdt + F_y(x + h, y + q, t)pdt \\ &+ F_t(x + h, y + q, t)dt \\ F(x + fdt, y + pdt, t + dt) &= F(x, y, t) + F_x(x, y, t)fdt + F_y(x, y, t)pdt + F_t(x, y, t)dt \end{aligned}$$

using that the bilinear differential form $dt\, dN(t) = 0$. Finally, by substituting the above three equations into Eq. 12, we conclude that:

$$\begin{aligned} dF(x(t), y(t), t) = (F_t + fF_x + pF_y)(x(t), y(t), t)dt + \big[&F(x + g, y + q, t)z(t) + F(x + h, y + q, t)(1 - z(t)) \\ &- F(x, y, t)\big]dN(t), \end{aligned}$$

8 Lemma 6

Lemma 6 *Consider the following family of losses with parameter $d > 0$,*

$$\begin{aligned}
\ell_d(m(t), n(t), u(t)) &= h_d(m(t), n(t)) + g_d^2(m(t), n(t)) + \frac{1}{2} q u(t)^2, \\
g_d(m(t), n(t)) &= 2^{-1/2} \left[c_2 \frac{\log(d)}{-m(t)^2 + 2m(t) - d} - c_2 \frac{\log(d)}{1-d} + c_1 m(t) \log\left(\frac{1+\beta}{1-\alpha}\right) - c_1 \log(1+\beta) \right]_+, \\
h_d(m(t), n(t)) &= -\sqrt{q} m(t) n(t) c_2 \frac{(-2m(t)+2)\log(d)}{(-m(t)^2 + 2m(t) - d)^2}.
\end{aligned} \tag{13}$$

where $c_1, c_2 \in \mathbb{R}$ are arbitrary constants. Then, the cost-to-go $J_d(m(t), n(t), t)$ that satisfies the HJB equation, defined by Eq. 9, is given by:

$$J_d(m(t), n(t), t) = \sqrt{q} \left(c_1 \log(n(t)) + c_2 \frac{\log(d)}{-m(t)^2 + 2m(t) - d} \right) \tag{14}$$

and the optimal intensity is given by:

$$u_d^*(t) = q^{-1/2} [c_2 \frac{\log(d)}{-m(t)^2 + 2m(t) - d} - c_2 \frac{\log(d)}{1-d} + c_1 m(t) \log\left(\frac{1+\beta}{1-\alpha}\right) - c_1 \log(1+\beta)]_+.$$

Proof *Consider the family of losses defined by Eq. 13 and the functional form for the cost-to-go defined by Eq. 14. Then, for any parameter value $d > 0$, the optimal intensity $u_d^*(t)$ is given by*

$$\begin{aligned}
u_d^*(t) &= q^{-1} \left[J_d(m(t), n(t), t) - J_d(1, (1-\alpha) n(t), t) m(t) - J_d(1, (1+\beta) n(t), t)(1 - m(t)) \right]_+ \\
&= q^{-1/2} \left[c_2 \frac{\log(d)}{-m^2 + 2m - d} - c_2 \frac{\log(d)}{1-d} + c_1 m(t) \log\left(\frac{1+\beta}{1-\alpha}\right) - c_1 \log(1+\beta) \right]_+,
\end{aligned}$$

and the HJB equation is satisfied:

$$\begin{aligned}
&\frac{\partial J_d(m,n,t)}{\partial t} - mn \frac{\partial J_d(m,n,t)}{\partial m} + h_d(m,n) + g_d^2(m,n) - \frac{1}{2} q^{-1} \big(J_d(m,n,t) \\
&- J_d(1, (1-\alpha)n, t) m - J_d(1, (1+\beta)n, t)(1-m) \big)_+^2 \\
&= \sqrt{q} mn c_2 \frac{(-2m+2)\log(d)}{(-m^2+2m-d)^2} + h_d(m,n) + g_d^2(m,n) - \frac{1}{2} \left[c_1 \log(n) + c_2 \frac{\log(d)}{-m^2+2m-d} \right. \\
&\quad \left. - m \left(c_1 \log(n(1-\alpha)) + c_2 \frac{\log(d)}{1-d} \right) - (1-m) \left(c_1 \log(n(1+\beta)) + c_2 \frac{\log(d)}{1-d} \right) \right]_+^2 \\
&= \sqrt{q} mn c_2 \frac{(-2m+2)\log(d)}{(-m^2+2m-d)^2} \underbrace{- \sqrt{q} mn c_2 \frac{(-2m+2)\log(d)}{(-m^2+2m-d)^2}}_{h_d(m,n)} \\
&\quad - \frac{1}{2} \big[c_2 \frac{\log(d)}{-m^2+2m-d} - c_2 \frac{\log(d)}{1-d} + c_1 m \log(\frac{1+\beta}{1-\alpha}) - c_1 \log(1+\beta) \big]_+^2 \\
&\quad + \frac{1}{2} \big[c_2 \frac{\log(d)}{-m^2+2m-d} - c_2 \frac{\log(d)}{1-d} + c_1 m \log(\frac{1+\beta}{1-\alpha}) - c_1 \log(1+\beta) \big]_+^2
\end{aligned}$$

where for notational simplicity $m = m(t)$, $n = n(t)$ and $u = u(t)$. ■

9 Proof of Theorem 4

Consider the family of losses defined by Eq. 13 in Lemma 6 whose optimal intensity is given by:

$$u_d^*(t) = q^{-1/2}\left[c_2\frac{\log(d)}{-m^2+2m-d} - c_2\frac{\log(d)}{1-d} + c_1 m(t)\log\left(\frac{1+\beta}{1-\alpha}\right) - c_1\log(1+\beta)\right]_+.$$

Now, set the constants $c_1, c_2 \in \mathbb{R}$ to the following values:

$$c_1 = \frac{-1}{\log\left(\frac{1+\beta}{1-\alpha}\right)} \quad c_2 = \frac{-\log(1-\alpha)}{\log\left(\frac{1+\beta}{1-\alpha}\right)}.$$

Since the HJB equation is satisfied for any value of $d > 0$, we can recover the quadratic loss $l(m, n, u)$ and derive its corresponding optimal intensity $u^*(t)$ using point wise convergence:

$$l(m(t), n(t), u(t)) = \lim_{d\to 1} l_d(m(t), n(t), u(t)) = \frac{1}{2}(1-m(t))^2 + \frac{1}{2}qu^2(t),$$
$$u^*(t) = \lim_{d\to 1} u_d^*(t) = q^{-1/2}(1-m(t)),$$

where we used that $\lim_{d\to 1}\frac{\log(d)}{1-d} = -1$ (L'Hospital's rule).

This concludes the proof.

10 The MEMORIZE Algorithm under the Power-Law Forgetting Curve Model

In this section, we first derive the optimal reviewing schedule under the power-law forgetting curve model and then validate such schedule using the same Duolingo dataset as in the main section of the paper.

Problem formulation and algorithm. Under the power-law forgetting curve model, the probability of recalling an item i at time t is given by [24]:

$$m_i(t) := \mathbb{P}(r_i(t)) = (1+\omega(t-t_r))^{-n_i(t)}, \tag{15}$$

where t_r is the time of the last review, $n_i(t) \in RR^+$ is the forgetting rate and ω is a time scale parameter.

Similarly as in Proposition 1 for the exponential forgetting curve model, we can express the dynamics of the recall probability $m_i(t)$ by means of a SDE with jumps:

$$dm_i(t) = -\frac{n_i(t)m_i(t)\omega dt}{(1+\omega D_i(t))} + (1-m_i(t))dN_i(t) \tag{16}$$

where $D_i(t) := t - t_r$ and thus the differential of $D_i(t)$ is readily given by $dD_i(t) = dt - D_i(t)dN_i(t)$.

Next, similarly as in the case of the exponential forgetting curve model in the main paper, we consider a single item with $n_i(t) = n(t)$, $m_i(t) = m(t)$, $D_i(t) = D(t)$ and $r_i(t) = r(t)$, and adapt Lemma 3 to the power-law forgetting curve model as follows:

Lemma 7 *Let $x(t)$ and $y(t)$, $k(t)$ be three jump-diffusion processes defined by the following jump SDEs:*

$$\begin{aligned}
dx(t) =& f(x(t), y(t), t)dt + g(x(t), y(t), t)z(t)dN(t) + h(x(t), y(t), t)(1-z(t))dN(t)\\
dy(t) =& p(x(t), y(t), t)dt + q(x(t), y(t), t)dN(t)\\
dk(t) =& s(x(t), y(t), k(t), t)dt + v(x(t), y(t), k(t), t)dN(t)
\end{aligned}$$

where $N(t)$ is a jump process and $z(t) \in \{0, 1\}$. If function $F(x(t), y(t), k(t), t)$ is once continuously differentiable in $x(t)$, $y(t)$, $z(t)$ and t, then,

$$\begin{aligned} dF(x, y, k, t) = (F_t + fF_x + pF_y + sF_s)(x, y, k, t)dt + [F(x + g, y + q, k + v, t)z(t) \\ + F(x + h, y + q, k + v, t)(1 - z(t)) - F(x, y, t)]dN(t), \end{aligned}$$

where for notational simplicity we dropped the arguments of the functions f, g, h, p, q, s, v and argument of state variables.

Then, if we consider $x(t) = n(t)$, $y(t) = m(t)$, $k(t) = D(t)$, $z(t) = r(t)$ and $J = F$ in the above Lemma, the differential of the optimal cost-to-go is readily given by

$$\begin{aligned} dJ(m, n, t) = J_t(m, n, t) - \frac{\omega nm}{1 + \omega D} J_m(m, n, D, t) + J_D(m, n, D, t) + [J(1, (1 - \alpha)n, 0, t)r(t) \\ + J(1, (1 + \omega)n, 0, t)(1 - r) - J(m, n, D, t)]dN(t). \end{aligned}$$

Moreover, under the same loss function $\ell(m(t), n(t), u(t))$ as in Eq. 10, it is easy to show that the optimal cost-to-go J needs to satisfy the following nonlinear partial differential equation:

$$\begin{aligned} 0 = J_t(m(t), n(t), t) - \frac{\omega n(t)m(t)}{1 + \omega D} J_m(m(t), n(t), t) + \quad J_D(m(t), n(t), t) + \frac{1}{2}(1 - m(t))^2 \\ - \frac{1}{2} q^{-1} \big(J(m(t), n(t), t) - J(1, (1 - \alpha)n(t), t)m(t) - J(1, (1 + \beta)n(t), t)(1 - m(t))\big)_+^2. \end{aligned} \tag{17}$$

Then, we can adapt Lemma 6 to derive the optimal scheduling policy for a single item under the power-law forgetting curve model:

Lemma 8 *Consider the following family of losses with parameter $d > 0$,*

$$\begin{aligned} \ell_d(m(t), n(t), D(t), u(t)) &= h_d(m(t), n(t), D(t)) + g_d^2(m(t), n(t)) + \frac{1}{2} q u(t)^2, \\ g_d(m(t), n(t)) &= 2^{-1/2} \left[c_2 \frac{\log(d)}{-m(t)^2 + 2m(t) - d} - c_2 \frac{\log(d)}{1 - d} + c_1 m(t) \log\left(\frac{1 + \beta}{1 - \alpha}\right) - c_1 \log(1 + \beta) \right]_+, \\ h_d(m(t), n(t)) &= -\sqrt{q} \frac{\omega n(t)m(t)}{1 + \omega D(t)} c_2 \frac{(-2m(t) + 2)\log(d)}{(-m(t)^2 + 2m(t) - d)^2}. \end{aligned} \tag{18}$$

where $c_1, c_2 \in \mathbb{R}$ are arbitrary constants. Then, the cost-to-go $J_d(m(t), n(t), t)$ that satisfies the HJB equation, defined by Eq. 17, is given by:

$$J_d(m(t), n(t), D(t), t) = \sqrt{q} \left(c_1 \log(n(t)) + c_2 \frac{\log(d)}{-m(t)^2 + 2m(t) - d} \right) \tag{19}$$

which is independent of $D(t)$, and the optimal intensity is given by:

$$u_d^*(t) = q^{-1/2} [c_2 \frac{\log(d)}{-m(t)^2 + 2m(t) - d} - c_2 \frac{\log(d)}{1 - d} + c_1 m(t) \log\left(\frac{1 + \beta}{1 - \alpha}\right) - c_1 \log(1 + \beta)]_+.$$

Proof *Consider the family of losses defined by Eq. 18 and the functional form for the cost-to-go defined by Eq. 19. Then, for any parameter value $d > 0$, the optimal intensity $u_d^*(t)$ is given by*

$$\begin{aligned} u_d^*(t) &= q^{-1} \left[J_d(m(t), n(t), t) - J_d(1, (1 - \alpha)n(t), t)m(t) - J_d(1, (1 + \beta)n(t), t)(1 - m(t)) \right]_+ \\ &= q^{-1/2} \left[c_2 \frac{\log(d)}{-m^2 + 2m - d} - c_2 \frac{\log(d)}{1 - d} + c_1 m(t) \log\left(\frac{1 + \beta}{1 - \alpha}\right) - c_1 \log(1 + \beta) \right]_+, \end{aligned}$$

	Effort, $\hat{e}$		Emp. forgetting, $\hat{n}$	
Pattern	$\frac{\hat{e}_M}{\hat{e}_T}$	$\frac{\hat{e}_M}{\hat{e}_U}$	$\frac{\hat{n}_M}{\hat{n}_T}$	$\frac{\hat{n}_M}{\hat{n}_U}$
●●●	0.00*	0.02*	0.11*	0.11*
●●●●	0.01*	0.05*	0.14*	0.15*
●●●●●	0.02*	0.07*	0.20*	0.20*
×●●	0.01*	0.03*	0.10*	0.10*
×●●●	0.04*	0.07*	0.13*	0.13*
××●●	0.03*	0.06*	0.11*	0.11*
×××●●	0.04*	0.06*	0.13∗	0.13*

Figure 6: Performance of MEMORIZE (M) in comparison with uniform (U) and threshold (T) based reviewing schedule for the power-law forgetting curve model. Each row of each table corresponds to a different recall pattern, depicted in the first column, where markers denote ordered recalls, green circles indicate successful recall and red crosses indicate unsuccessful ones. In the second column, each cell value corresponds to the ratio between the average effort for the top 25% pairs in terms of likelihood for the uniform (or threshold based) schedule and MEMORIZE. In the right column, each cell value corresponds to the ratio between the median empirical forgetting rate. The symbol $*$ indicates whether the change is significant with p-value < 0.01 using the Kolmogrov-Smirnov 2-sample test.

and the HJB equation is satisfied:

$$
\begin{aligned}
&\frac{\partial J_d(m,n,t)}{\partial t} - \frac{\omega n m}{1+\omega D}\frac{\partial J_d(m,n,t)}{\partial m} + h_d(m,n) + g_d^2(m,n) - \frac{1}{2}q^{-1}\big(J_d(m,n,t) \\
&- J_d(1,(1-\alpha)n,t)m - J_d(1,(1+\beta)n,t)(1-m)\big)_+^2 \\
&= \sqrt{q}\frac{\beta c_2 n m}{1+\omega D}\frac{(-2m+2)\log(d)}{(-m^2+2m-d)^2} + h_d(m,n) + g_d^2(m,n) - \frac{1}{2}\left[c_1\log(n) + c_2\frac{\log(d)}{-m^2+2m-d}\right. \\
&\quad \left. -m\left(c_1\log(n(1-\alpha)) + c_2\frac{\log(d)}{1-d}\right) - (1-m)\left(c_1\log(n(1+\beta)) + c_2\frac{\log(d)}{1-d}\right)\right]_+^2 \\
&= \sqrt{q}\frac{\omega n m c_2}{1+\omega D(t)}\frac{(-2m+2)\log(d)}{(-m^2+2m-d)^2} \underbrace{-\sqrt{q}\frac{\omega n m c_2}{1+\omega D}\frac{(-2m+2)\log(d)}{(-m^2+2m-d)^2}}_{h_d(m,n)} \\
&\quad - \frac{1}{2}\big[c_2\frac{\log(d)}{-m^2+2m-d} - c_2\frac{\log(d)}{1-d} + c_1 m\log(\frac{1+\beta}{1-\alpha}) - c_1\log(1+\beta)\big]_+^2 \\
&\quad + \frac{1}{2}\big[c_2\frac{\log(d)}{-m^2+2m-d} - c_2\frac{\log(d)}{1-d} + c_1 m\log(\frac{1+\beta}{1-\alpha}) - c_1\log(1+\beta)\big]_+^2
\end{aligned}
$$

where for notational simplicity $m = m(t)$, $n = n(t)$, $D = D(t)$ and $u = u(t)$. ■

Finally, reusing Theorem 4, the optimal reviewing intensity for a single item under the power-law forgetting curve model is given by

$$u^*(t) = \lim_{d\to 1} u_d^*(t) = q^{-1/2}(1-m(t)).$$

It is then straightforward to derive the optimal reviewing intensity for a set of items, which adopts the same form as in Theorem 5.

Experimental evaluation. In this section, the dataset, baselines, experimental setup and quality metrics are the same as in the main paper.

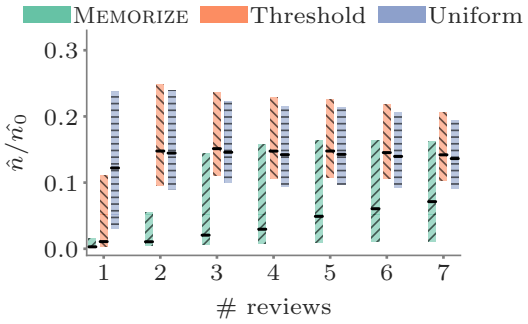
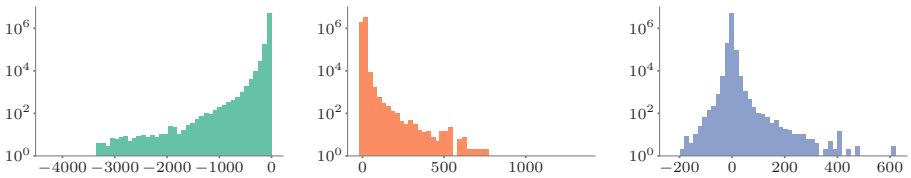


Figure 7: Performance of MEMORIZE with the power-law forgetting curve model compared to the uniform and threshold schedules. Boxes indicate 25% and 75% quantiles and solid lines indicate median value. Lower values indicate better performance. In all cases, the competitive advantage MEMORIZE achieves is statistically significant (p-value < 0.01 using the Kolmogrov-Smirnov 2-sample test).



(a) MEMORIZE (b) threshold (c) uniform

Figure 8: Histogram of log-likelihood of (user, item) pairs following different reviewing schedules. By looking at the mode of the distribution of MEMORIZE log-likelihoods in Panel (a), most (user, item) pairs appear to follow MEMORIZE closely.

Figure 6 summarizes the results for the average effort and empirical forgetting rate for the most common recall patterns. Here, we report the ratio between the effort and empirical forgetting rate values achieved by the uniform and threshold schedules and the values achieved by MEMORIZEwith the power-law forgetting curve model. We find that both in terms of effort and empirical forgetting rate, MEMORIZE outperforms the baselines for all recall patterns. For example, for the recall pattern consisting of two unsuccessful recalls followed by two successful recalls (red-red-green-green), MEMORIZE achieves 0.06× lower effort and 0.11× lower empirical forgetting rate than the second competitor. Figure 7 summarizes the results for the average effort for sequences with different number of reviews, which show that MEMORIZE with the power-law forgetting curve model also outperforms the baselines.

Overall, we would like to highlight that we did not find a clear winner in performance between MEMORIZE with the exponential forgetting curve model and MEMORIZE with the power-law forgetting curve model.

11 Distribution of likelihood values for different reviewing schedules

We compute likelihood of each sequence of review events in our dataset under different reviewing schedules and evaluate the metrics on top 25% of sequences under each competing review schedules. Figure 8 shows the distribution of estimated likelihood values for MEMORIZE, threshold and uniform schedules. The histogram for MEMORIZE shows that the peak of the distribution coincides with the highest likelihood values under MEMORIZE. This observation is in agreement with the fact that Duolingo already uses a near-optimal

	HLR Exponential	**Our Model** Exponential	**Our Model** Power-law
MAE↓	0.128	0.129	**0.105**
AUC↑	0.538	**0.542**	0.533
$\text{COR}_h \uparrow$	**0.201**	0.165	0.123

Table 1: Predictive performance of the exponential and power-law forgetting curve models in comparison with the results reported by Settles et al. [23]. The arrows indicate whether a higher value of the metric is better (↑) or a lower value (↓).

hand-tuned spacing algorithm for scheduling reviews.

12 Predictive performance of the memory model

Before we evaluate the predictive performance of the exponential and power-law forgetting curve models, whose forgetting rates we estimated using a variant of Half-life regression (HLR) [23], we highlight the differences between the original HLR and the variant we used.

The original HLR and the variant we used differ in the way successful and unsuccessful recalls change the forgetting rate. In our work, the forgetting rate at time t depends on $n_{\checkmark}(t) = \int_0^t r(\tau) dN(\tau)$ and $n_{\times}(t) = \int_0^t (1 - r(\tau)) dN(\tau)$. In contrast, in the original HLR, the forgetting rate at time t depends on $\sqrt{n_{\checkmark}(t) + 1}$ and $\sqrt{n_{\times}(t) + 1}$. The rationale behind our modeling choice is to be able to express the dynamics of the forgetting rate using a linear stochastic differential equation with jumps. Moreover, Settles *et al.* consider each session to contain multiple review events for each item. Hence, within a session, the $n_{\times}(t)$ and $n_{\checkmark}(t)$ may increase by more than one. In contrast, we consider each session to contain a single review event for each item because the reviews in each session take place in a very short time and it is likely that after the first review, the user will recall the item correctly during that session. Hence, we only increase one of $n_{\checkmark}(t)$ or $n_{\times}(t)$ by exactly 1 after each session and consider an item has been successfully recalled during a session if all reviews were successful, *i.e.*, $p_{recall} = 1$. Noticeably, ∼83% of the items were successfully recalled during a session.

Table 1 summarizes our results on the Duolingo dataset in terms of mean absolute error (MAE), area under curve (AUC) and correlation (COR_h), which show that the performance of both the exponential and power-law forgetting curve models with forgetting rates estimated using the variant of HLR is comparable to the performance of the exponential forgetting curve model with forgetting rates estimated using the original HLR.

13 Effect of review time on forgetting rate

In previous studies, it has been shown that the interval between reviews has an effect on the forgetting rate, especially at large review/retention time-scales [8]. In this section, we discuss how we tested for such effects in our dataset and justify our decision to employ the simpler model with constant updates to the forgetting rate, independent of the review interval.

Formulation. In Eq. 3, we have considered α and β as constants, *i.e.*, they do not vary with the review interval $t - t_r$, where t_r is the time of last review. We have dropped the subscript i denoting the item for ease of exposition. We can make a zeroth-order approximation to time varying (α, β) by allowing them to be piecewise constant for K mutually exclusive and exhaustive review-time intervals $\left\{B^{(i)}\right\}_{[K]}$. We denote the value that α (β) takes in interval $B^{(i)}$ as $\alpha^{(i)}$ ($\beta^{(i)}$) and modify the forgetting rate update equation to

$$dn(t) = \underbrace{-\alpha^{(i)} n(t) r(t) dN(t) + \beta^{(i)} n(t)(1 - r(t)) dN(t)}_{i \text{ such that } t - t_r \in B^{(i)}}$$

K	Interval boundaries
3	[0, 20 minutes, 2.9 days, ∞]
4	[0, 9 minutes, 21.5 hours, 5.2 days, ∞]
5	[0, 6 minutes, 1.5 hours, 1.8 days, 7.3 days, ∞]

Table 2: The boundaries of intervals used to divide review times into bins based on K-quantiles of inter-review times in the Duolingo dataset.

	2	3	4	5
1	0.317	0.317	0.317	0.317
2		0.312	0.312	0.312
3			0.317	0.317
4				0.318

(a) Comparing $\alpha^{(\text{row})}$ and $\alpha^{(\text{col})}$

	2	3	4	5
1	0.165	0.172	0.165	0.165
2		0.318	0.302	0.302
3			0.318	0.318
4				0.302

(b) Comparing $\beta^{(\text{row})}$ and $\beta^{(\text{col})}$

Table 3: p-values obtained after using Welch's t-test for populations with different variance to reject the null hypothesis that the samples (*i.e.*, 400 samples of $\{\alpha^{(i)}\}$ and $\{\beta^{(i)}\}$ for $i \in [5]$) have the same mean value. In all cases, we find no evidence to reject the null hypothesis. The results for other values of K were qualitatively similar.

If we find that $\exists \{i, j\} \subset [K]$ such that $\alpha^{(i)}$ ($\beta^{(i)}$) is *significantly* different from $\alpha^{(j)}$ ($\beta^{(j)}$), then we would conclude that α (β) vary with review-time.

We obtain repeated estimates of $\left\{\alpha^{(i)}\right\}_{[K]}$ and $\left\{\beta^{(i)}\right\}_{[K]}$ by fitting our model to datasets sampled with replacement from our Duolingo dataset, *i.e.*, via bootstrapping. The Welch's t-test is used to test if the difference in mean values of the parameters in different bins is significant.

Experimental setup. We set the bins boundaries by determining the K-quantiles of the review times in our dataset. Table 2 shows that the bin boundaries for different K are quite varied and adequately cover long time-scales as well as review intervals which are short enough to capture *massed practicing*. This method of binning also ensures that we have sufficient samples for accurate estimation ($\sim 5.2e6/K$) for all parameters. Then we use the variant of HLR described in Appendix 12 to fit the parameters in 400 different datasets using bootstrapping. The regularization parameters are determined via grid-search using a train/test dataset. We thus obtain 400 samples of $\left\{\alpha^{(i)}\right\}_{[K]}$ and $\left\{\beta^{(i)}\right\}_{[K]}$ for $K \in \{3, 4, 5\}$ and $i \in [K]$. Using Welch's t-test for distributions with varying variances, we observe that the mean values of the distributions of $\left\{\alpha^{(i)}\right\}_{[K]}$ ($\left\{\beta^{(i)}\right\}_{[K]}$) and $\left\{\alpha^{(j)}\right\}_{[K]}$ ($\left\{\beta^{(j)}\right\}_{[K]}$) are not significantly different for any $\{i, j\}$. As an example, the p-values obtained for $K = 5$ are shown in Table 3.

As discussed in Section 2, a possible explanation for this is that our model takes the recall of the learners at each review $r(t)$ into account to update the forgetting rate while in [8], the updates do not take the recall into account.