A Reciprocal-Practice-Success (RPS) Model of Free Practice

Pritam Laskar (p.laskar@uva.nl)
Department of Psychology
University of Amsterdam

H.L.J. van der Maas (h.l.j.vandermaas@uva.nl)
Department of Psychology
University of Amsterdam

Abstract

Understanding how humans learn by themselves is crucial to develop interventions to prevent dropout and improve learner engagement. Classical learning curves were proposed to fit and describe experimental data involving enforced learning. However in real-world learning contexts such as MOOCs and hobbies, learners may quit - and often do. Even in formal settings such as college, success typically requires intensive self-study outside lectures. Previous research in educational psychology supports a positive reciprocal relationship between motivation and achievement. Integrating insights from learning curves, forgetting curves and motivation-achievement cycles, we propose a formal Reciprocal-Practice-Success (RPS) model of learning 'in the wild'. First, we describe the different components of the basic RPS model. Using simulations, we then show how long term learning outcomes critically depend on the shape of the learning curve. Concave curves lead to more consistent learning outcomes whereas S-shaped curves lead to either expertise or dropout. We also provide a dynamical systems version for the RPS model which shows similar qualitative behaviour. Through a bifurcation analysis of two controllable learning parameters - minimum practice rate and success sensitivity, we show which learner-specific interventions may be effective to preventing dropout. We also discuss theorized mechanisms which affect the inflection point of S-shaped learning curves such as task-complexity and relative feedback from failures vs. successes. These provide more task-specific interventions to lower quitting rates. Finally, we discuss possible extensions to the basic RPS model which will allow capturing spacing effects and insights from other motivation theories.

Keywords: Self-Regulated Learning, Motivation-Achievement Cycles, Learning Curves

Introduction

Reciprocal relationships involve two variables which are mutually causes and effects of each other. In learning processes such reciprocal processes may reinforce either positive or negative outcomes. The literature on this type of developmental processes is extensive and diverse. Examples include studies of the Matthew effect (Perc, 2014), academic motivation (Guay et al., 2010; Schunk, 1991; Marsh and Craven, 2006), drop-out (De Witte et al., 2013), and, on the positive side, expertise learning (Anders Ericsson and Towne, 2010). An important reciprocal relationship is between practice and success or performance (Vu et al., 2022, van Bergen et al., 2021). When practice and success reinforce each other, it may lead to positive reinforcing cycles as learners become highly engaged, practice more and achieve higher skill. However there is also a danger of learners performing poorly and losing interest, leading to even less practice and in extreme cases, dropout. Indeed, in the increasingly popular Massive Open Online Courses (MOOCs), combating high dropout rates are one the biggest challenges (Hone and El Said, 2016; Goopio and Cheung, 2021).

In studying the dynamics of practice and success there is an important difference between free and forced practice. Our proposed model is applicable in *free practice* scenarios such as taking up learning chess, guitar or running as a hobby, where practice needs to be regulated by the learners themselves. In contrast, *forced practice* include scenarios such as learning math in school or ceramic pottery lessons forced by your partner. These are scenarios where the learner has no choice (or rather, choices with severe consequences) but to practice.

We build upon work on learning curves for forced practice which have been extensively studied in mathematical psychology. However, free practice is very different from the experimental conditions of traditional learning experiments. Participants in such experiments could drop out, but normally their data were excluded. As an example, Thurstone (1919) in his seminal paper on the learning curve eliminated 20 out of 81 participants in his sample because of irregular attendance.

The Reciprocal-Practice-Success Model Framework

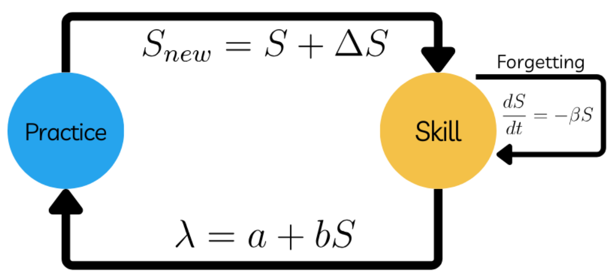
Our central idea is that learning a skill, operationalized as improvement in some metric of performance, can only occur through practice which in turn is (partially) dependent upon current skill. How often an agent practices is related to the psychological construct of *motivation*. In our model motivation is not an observable but it impacts learning by changing the frequency of practice, which can be measured. A higher practice rate means that learners practice more often or, equivalently, less time is spent between practice sessions. When learners aren't practicing, their skills decay due to forgetting. Practice sessions are modelled to be points in time when learning takes place. Crucially, we claim that the practice rate is influenced by the skill of the learner in a positive reciprocal relationship. (Vu et al., 2022; Marsh and Craven, 2006; van Bergen et al., 2021). So while learning the guitar, the better you are, the more often you will practice *all else remaining equal*. Depending on the task, there are always other factors which influence the practice rate. For instance, even

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

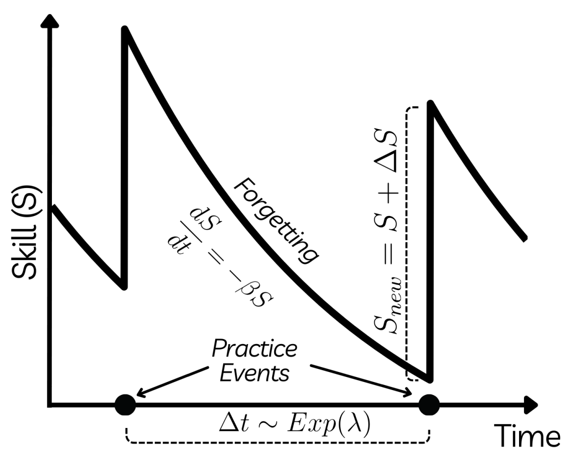
if you are a good guitarist you may stop practicing when you no longer notice improvements. Such details could also be incorporated into the RPS model, although the basic model doesn't consider them.

A Simple Formalization - The Basic RPS Model

We now describe one possible way to formalize the ideas of RPS framework into a minimal model. Depending on the specific context we are modelling - be it running, learning the guitar or a second language - the details of the model will change. Certain forced-practice scenarios can also be incorporated within the RPS framework. However, our aim is to have a minimal model incorporating the ideas presented above. Figure 1 shows a block diagram of the model along with a possible learning trajectory. We briefly describe each component of the model.



(a) Model Schematic



(b) Sample Learning Trajectory

Figure 1: Learner skill S ($0 \leq S \leq 1$) jumps at practice-events. Waiting-times between practice events are assumed to follow an exponential distribution with practice rate $\lambda = a + bS$, which is an increasing linear function of skill S in the basic model. $a > 0$ is the minimum practice rate while $b > 0$ is the success-sensitivity of a learner. Between practice events an exponential decay of skill occurs, representing forgetting.

1. **Measuring Success (or Skill):** The state of a learner is presumed to be captured by the skill S which is an increasing function of how 'good' the learner is. This may be ELO ratings in chess or average 1 km running speed in case of running. In the model we assume $0 \leq S \leq 1$ and interpret it as the probability of success on the task.

2. **Forgetting:** As time passes continuously, the learner's skill decays exponentially with a forgetting rate β (Anderson and Tweney, 1997; Averell and Heathcote, 2011), except at practice events. So without practice, the dynamics of skill is given by:

$$\frac{dS}{dt} = -\beta S \tag{1}$$

Here we do not consider lowered forgetting rates with subsequent practice or spacing effects (Hintzman, 1974; Pavlik Jr and Anderson, 2005a).

3. **Practice Events:** At practice events, the learner practices and their skill S increases according to the update rule $S_{new} = S_{old} + \Delta S$. The impact of practice is captured by the change in skill ΔS at a practice event. We take it to be a function of the skill S just before the practice event, so that:

$$\Delta S = f(S) \tag{2}$$

Different functional forms of $f(S)$ give us different shapes of the learning curve. If $f(S)$ decreases with increasing skill S, a concave learning curve is obtained where practice gives diminishing returns - the higher your skill, the less it improves from practicing once. If $f(S)$ first increases to a maximum and then decreases, we have an S-shaped learning curve.

4. **Waiting Times:** The waiting-time Δt between two consecutive practice-events are assumed to be an exponential distribution with rate λ, so that $\Delta t \sim Exp(\lambda)$. The expected waiting time is then $\mathbb{E}[\Delta t] = \frac{1}{\lambda}$. A higher practice rate indicates shorter waiting times and thus less forgetting of skill S before the next practice-event increases S again.

Exponential waiting times are arise naturally as the distance between two consecutive events in a Poisson processes and are used to model phenomenon - from time spent in a queue to distance between mutations in DNA. Barabasi (2005) showed through simulations that waiting times for a task are heavy tailed when agents decide to perform the highest-priority task from a set of to-do tasks. Exponential waiting times result when users pick a random task to complete, regardless of priority.

5. **Reciprocity:** The reciprocal relationship between practice and success is captured by assuming a simple linear dependence of practice rate upon skill for the basic RPS model:

$$\lambda(S) = a + bS = g(S) \tag{3}$$

In Equation 3, a is the minimum practice rate and b the sensitivity to success. When $b = 0$, practice occurs with a fixed rate independent of success rate. This can be used to model forced-practice scenarios, such as attending school or doing a fixed number of trials in a learning task in a behavioural experiment.

The Shape of Learning

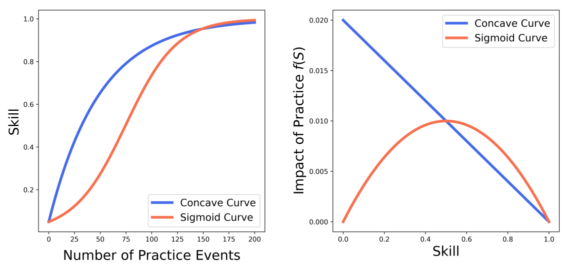


Figure 2: Learning curves without forgetting (left) and the corresponding impact function (right)

The effect of practice on success has been studied extensively since the start of psychology as a scientific discipline (Ebbinghaus, 1885; Thurstone, 1919). Illustrative of the initial interest in this issue is Gulliksen's (1934) insightful review of mathematical equations for the learning curve in the context of forced practice. There are two main forms of the learning curve reported in the literature 2 (left). The first form is concave, characterized by deceleration of success with increasing practice. The second form is S-shaped characterized by first an acceleration and then a deceleration. Note that these equations have been developed mainly to fit data from experiments where the practice periods are not under the participant's control. This is also why learning curves are sometimes described as success as a function of practice trials $S(P)$ and not time $S(t)$. When practice trials are evenly spaced in time (say every 10 seconds), P also counts time t . The crucial point we make is that in free-practice scenarios, time between consecutive practice sessions are not evenly spaced. We thus describe the forced-practice learning curves as functions of time $S(t)$.

In our notation, $f(S) = \Delta S$ is the marginal increase in skill from the next practice trial. In other words, $f(S)$ captures the *impact of practice* 2 (right). In our basic model this is taken to be the rate of learning (the time derivative) in continuous-time learning curves $S(t)$:

$$f(S) = \frac{dS}{dt} \tag{4}$$

Concave Learning Curve When $f(S)$ is a decreasing function of S, we have a concave learning curve. In such a scenario, impact of practice is highest when the learner knows nothing and gets progressively lower as learning progresses 2. Many possible functional forms exist but two popular choices are the (negative) exponential and power laws (Estes, 1950; Newell and Rosenbloom, 2013; Rescorla, 1972). We will use the exponential learning curve where the rate of learning $\frac{dS}{dt}$ is proportional to how much skill can yet be improved:

$$\frac{dS}{dt} = \alpha(S_{max} - S)$$

Here α is the learning rate [1] which we assume is a constant throughout the learning process for a given individual on a given task. Also in our model, $S_{max} = 1$. So the above equation translates into the following marginal impact of practice:

$$\Delta S_{con} = f_{con}(S) = \alpha(1 - S) \tag{5}$$

Sigmoid Learning Curve: The learning curve with a S-shaped form starts slow, then accelerates and finally decelerates, as is typical in logistic growth. A number of empirical examples of such curves are depicted in Culler and Girden, 1951. This form is often called S-shaped. S-shape functions are consistent with theories that divide the learning process in phases of acceleration and then a deceleration; for instance, into algorithmic and retrieval based (Logan, 1988; Rickard, 2004).

If $f(S)$ first increases and then decreases with increasing S, we have a scenario where rate of learning $\frac{dS}{dt}$ is initially slow but accelerates, reaching a maximum before slowing down again as the maximum limit of skill S_{max} is reached. A popular example is *logistic growth*, where the rate of learning is a product of how much is learned and how much is yet to be learned so that $\frac{dS}{dt} = \alpha S(S_{max} - S)$. As before, α is the learning rate which is assumed to be constant for an individual on a given task. We use this as the marginal impact of practice for the sigmoidal case:

$$\Delta S_{sig} = f_{sig}(S) = \alpha S(1 - S) \tag{6}$$

A point to note is that the inflection-point in logistic-growth, where the marginal impact of practice ΔS_{sig} is maximum, occurs at $S_{inf} = \frac{1}{2}$ which is the midpoint between the minimum ($S_{min} = 0$) and maximum ($S_{max} = 1$) skill values. Indeed, two proposed mechanisms through which sigmoid learning curves may arise allow for inflection points to be located at points other than 0.5 (Leibowitz et al., 2010; Murre, 2014). Richard's curves (Richards, 1959) provide another general family of sigmoid curves where the inflection point can be anywhere between $1/e \approx 0.37$ and 1. S-shaped curves with low inflection points are more similar to concave curves as the slowly-growing initial phase is over soon. The exponential learning curve can be thought of as having $S = 0$ as its inflection point where impact of practice is highest.

Learning Mechanisms Shape the Curve

Here we present three mechanisms which can lead to differently shaped learning curves. One straightforward way individual S-shaped curves may remain undetected is by averaging across participants to get a mean learning curve (Gallistel et al., 2004). Higher inflection points which visually look

[1] Not to be confused with the rate of learning $\frac{dS}{dt}$

more S-shaped also depends on how exactly performance is measured. Murre (2014) showed that when trials are scored as 1 (correct) or 0 (incorrect), increasing task complexity - the number of subtasks one needs to correctly perform on one trial - leads to more of an S-shape, even when learning on the subtasks increase concavely. For example, while learning a new language the test questions (trials) to measure success may be either to translate words or entire sentences of 4-5 words. If a point is only awarded if the whole word/sentence is correctly translated, we will see more of an S-shaped learning while translating sentences. Finally, Leibowitz et al. (2010) showed how feedback plays an important role - if both failures and successes provide the same amount of information a concave shape is seen, while if failures provide no or very low information we get an S-shaped learning curve.

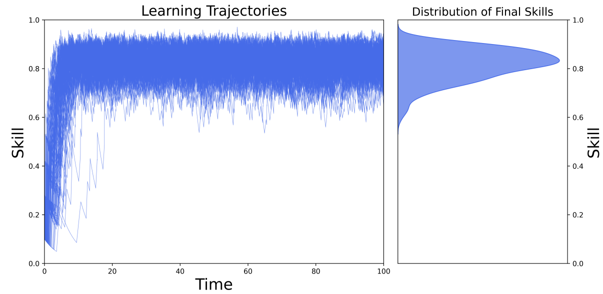
Fate of Free-Learners - A Simulation Study

We examined how identical learners under the Reciprocal-Practice-Success model formalization stated earlier would fare. Specifically we were interested in how $f(S)$ affects long-term skill S and practice rate λ. We simulated $N = 1000$ agents over a time period of $t_{max} = 100$ units of time. Each agent starts with a skill of $S_0 = 0.1$ and practice rate $\lambda_0 = 1$. The minimum practice rate, sensitivity and forgetting rates are fixed at $a = 0.2, b = 5$ and $\beta = 0.2$ respectively.

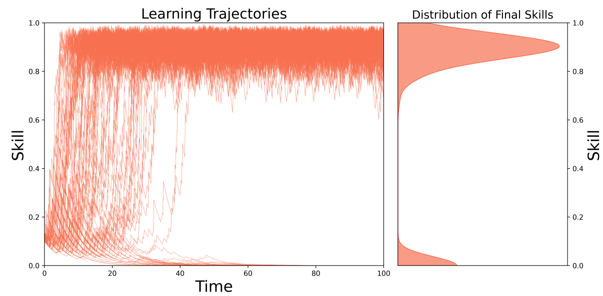
Simulation Results

Concave Learning Curve For a concave learning curve, we used Equation 5 with concave learning rate $\alpha_{con} = 0.2$. The learning trajectories of 50 random agents are given in Figure 3 (a), along with the histogram of final skills at time $t = 100$. All the learners eventually hover around mean skill value (around $S = 0.8$ in the figure). Learners improve their skill in large jumps early on in the learning process. This is due to ΔS_{con} being larger the lower the skill is. Forgetting occurs with rate β but is not enough to take away all gains from practice. As skill improves, the benefits due to practice decrease and eventually a kind of *equilibrium* is obtained where learners forget between sessions as much as they gain from one practice-event. Importantly no learners have dropped-out.

Sigmoid Learnerning Curve In contrast, for a sigmoidal learning curve, we used impact of practice $f_{sig}(S)$ from Equation 6 with learning rate $\alpha_{sig} = 0.4$. The results are given in Figure 3 (b). Now two extreme outcomes are possible - learners either get into a positive feedback loop of more success from a higher skill S leading to a higher practice rate λ or a negative cycle of low skill S leading to a decreasing practice rate until they dropout. This occurs due to the reciprocity built into the RPS model (Equation 3). The final fate of learners is shown in the histogram in Figure 3 (b). Around 20% of the learners have dropped out and stopped practicing, forgetting everything (skill $S \approx 0$). The rest have gone into a positive feedback loop and their skill hovers around a high skill value.



(a) Concave Learning Curve



(b) Sigmoid Learning Curve

Figure 3: Sumarry of Simulation

Intuition

To understand what is going on, it is helpful look again at Figure 2 (right) which plots the impact due to practice in concave f_{con} and sigmoid f_{sig} learning curves. For a sigmoidal learning curve, the impact of practice is very low at low skill levels and forgetting overwhelms any gains from practice. Peak practice impact is at the inflection point of $S_{inf} = 0.5$, after which learning slows down again. In our simulation the 'lucky' learners manage to practice enough early in the learning trajectory to get their skill over the hill at $S_{inf} = 0.5$. Thereafter it is uncommon to drop out despite continuous decay of skill. The practice rate is high enough that learners do not forget enough to go below the inflection point. On the other hand the 'unlucky' learners in our model simply do not practice enough in the beginning and forgetting keeps lowering their skill, reducing the practice rate further. Learners are stuck below the hill at $S_{inf} = 0.5$ and never recover, eventually *'quitting'*. We see a kind of Matthew effect, where the tiny initial gap between the learners in positive and negative feedback loops diverge as time goes on. Depending on the initial learning trajectory learners reach one of the two '*stable states*'.

Dynamical Approximation of the RPS Model

We can form a dynamical approximation of the RPS model which shows similar qualitative behaviour. This allows us to analytically understand the role of control parameters. Depending on the learning context what can be controlled may vary. In this section we take minimum practice rate (a) and sensitivity to success (b) as control parameters - knobs we can turn up or down to intervene in a learning context. To get the dynamical model, we ask what the rate of change of skill $S(t)$ is. Consider the time interval $(t, t+\Delta t]$. The practice rate is $\lambda(S) = a + bS$, so on average there are $\lambda\Delta t$ practice events. We assume that each of these practice events have an impact $f(S)$, which is the impact function. At the same time, forgetting occurs with rate βS. Combining these two effects, the change in skill ΔS in the time interval $(t, t+\Delta t]$ is:

$$\Delta S = (\lambda(S)\Delta t) f(S) - (\beta S)\Delta t$$
$$\implies \frac{\Delta S}{\Delta t} = \lambda(S) f(S) - \beta S$$

In the limit as $\Delta t \to 0$, this becomes the derivative $\frac{dS}{dt}$. So the dynamics of skill S is given by:

$$\frac{dS}{dt} = \lambda(S) f(S) - \beta S \quad (7)$$

For a concave learning curve, where impact is $f_{con}(S) = \alpha(1-S)$ and Equation 7 gives the following differential equation:

$$\frac{dS}{dt} = (-b\alpha)S^2 + ((b-a)\alpha - \beta)S + a\alpha$$

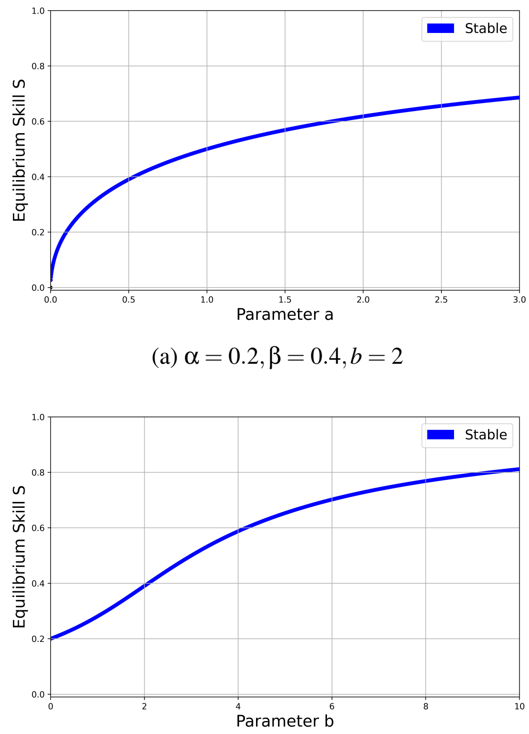
For a sigmoid learning curve the impact function is $f_{sig}(S) = \alpha S(1-S)$ and after some algebra we have:

$$\frac{dS}{dt} = (-b\alpha)S^3 + (b-a)\alpha S^2 + (a\alpha - \beta)S$$

What happens to skill in the long run in the dynamical approximation of the RPS model is qualitatively captured by the location of the stable fixed points of skill S. S_0 is called a fixed point if the derivative is zero at the point, $\frac{dS}{dt}|_{S=S_0} = 0$. This means that in absence of perturbations a learner who starts off at skill S_0 will remain there forever. A fixed point S_0 is stable if $S(t)$ returns back to S_0 when perturbed slightly like a ball placed at the bottom of a valley rolls back to the bottom on being slightly nudged. In contrast, an unstable fixed point S_0 is like placing the ball at the top of a hill. On being slightly disturbed, $S(t)$ will tend to move away from the unstable fixed point. In the real world noise is always present - learners constantly face successes and failures, bursts in motivation along with periods of feeling less competent. This means that only the stable fixed points are feasible.

Bifurcation Diagrams

Concave Curve Figure 4 shows the fixed points as a function of minimum practice rate a and sensitivity to success b. There is always one stable fixed point S_{eq} given by the blue line, which increases on increasing both a and b. All learners eventually reach a skill of S_{eq}. Increasing a or b only increases this equilibrium skill value but otherwise qualitatively nothing changes. Intuitively, as we force the learner to practice more often (increasing a), they will end up with a higher asymptotic skill. Increasing b which measures how sensitive their practice rate is to skill also does the same thing qualitatively.



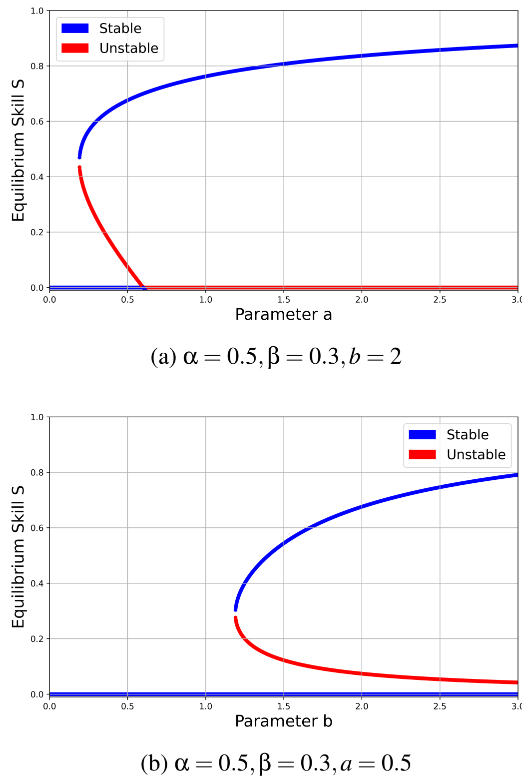
(a) $\alpha = 0.2, \beta = 0.4, b = 2$

(b) $\alpha = 0.2, \beta = 0.4, a = 0.5$

Figure 4: Bifurcation diagrams of minimum practice a and success sensitivity b in Concave Case

Sigmoid Curve: A more interesting behaviour is seen in the sigmoidal case (Figure 5). Now $S = 0$ is the only stable fixed point as as long $\beta > \frac{(a+b)^2\alpha}{4b}$. In this region, quitting is the only possible outcome and skill eventually decays to 0. $S = 0$ is stable when $\beta > a\alpha$ and unstable otherwise.

As minimum practice a is increased from a low starting value, a new pair fixed-points appear at $a = a_1$, but quitting remains a stable outcome, if participants start off with low skill. On further increasing minimum practice a, zero eventually becomes unstable at $a = a_2 = \frac{\beta}{\alpha}$. Now only one stable fixed point remains which is positive. Thus when $a > a_2$, all learners eventually reach a high skill and quitting never happens. Between a_1 and a_2 two stable states are possible - one at $S = 0$ and another at a high skill $S > 0$. Hysteresis is seen in this region. When $a_1 \leq a \leq a_2$, learners starting off below the unstable fixed point (red curve in Figure 5 (a)) eventually quit while those starting off above the red line reach expertise. The bifurcation diagram of a informs us about possible interventions to prevent quitting. If a learner has quit, a possible intervention is to increase $a \geq a_2$. Now $S = 0$ loses stability and the learner's skill will increase to the high stable point (blue curve in Figure 5 (a)). This gets the learner into a positive feedback loop of practice and success and a can be reduced again (though not below a_1) and the learner will maintain their high skill. When $a < a_1$, again the high stable



(a) $\alpha = 0.5, \beta = 0.3, b = 2$

(b) $\alpha = 0.5, \beta = 0.3, a = 0.5$

Figure 5: Bifurcation diagrams of minimum practice a and success sensitivity b in Sigmoid Case

point disappears and all learners eventually quit.

The control parameter b (sensitivity to success) has no effect on stability of 0. So if a learner has already quit increasing b is not a feasible intervention to get them into a positive practice-success feedback cycle. However if b is too low, 0 remains the only stable point and quitting occurs.

Interventions

The basic RPS simulations and bifurcation analysis of the dynamical model inform us about what interventions would work to prevent quitting and nudge learners towards expertise. We describe both task and learner specific interventions which follow from our work below.

1. Tasks should be structured so that the forced-practice learning curves are concave or otherwise have low inflection points. This may be done through providing adequate feedback on failures (Leibowitz et al., 2010) or breaking down complicated tasks (Murre, 2014). These are task-specific interventions, but may be non-trivial to carry out in reality. Unlike in our simulations, learners in the real-world are not identical with the same starting skills, learning and forgetting rates. The shape of learning curves really depends on the interaction between the task at hand and present state of the learner.

2. The minimum practice rate a can be increased through changing the environment (placing books in sight if one wants to read more), connecting socially with other learners or even forced-practice interventions (Franken et al., 2023; Hone and El Said, 2016). To maintain expertise, success should be motivating. In our basic model this translates into requiring success sensitivity b to be high enough.

Limitations, Extensions and Future Work

In this paper we provided a formal framework to model free-practice learning by combining results from learning and forgetting curves, modelling practice-events as a point-process with varying rate (λ) which reciprocally depends on current skill. We showed the shape of learning curves play a critical role in long-term learning outcomes. However the map is not the territory and the basic RPS model ignores a lot of complexity. First, we described a learner's state by a single variable S for simplicity. Depending on context, multidimentional states may be more appropriate. Similarly, we took re-learning forgotten skills to have the same learning rate as learning for the first time. Overall skill need not determine or even be the major motivating factor - the rate of increase of skill may also play a role. We tried out different setting practice rate to $\lambda = a + bS + c(\Delta S)$, where ΔS is the change in skill from the last practice event. This leads to a positive impact on practice rate during progress and a negative impact otherwise. These changes did not affect the qualitative results of our simulations.

The basic RPS model also does not consider spacing effects which are robust findings showing that forgetting rates decrease when practice is spaced out (Delaney et al., 2010). Within the RPS framework this can be incorporated by allowing forgetting rates β to depend on the entire history of wait-times $(\Delta t_1, \Delta t_2, \ldots \Delta t_n)$ between practice events (Pavlik Jr and Anderson, 2005b):

$$\beta(\Delta t_1, \ldots, \Delta t_n) = \beta_{min} + (\beta_{max} - \beta_{min}) \cdot e^{-\varepsilon \cdot \sum_{i=1}^{n} (\Delta t_i)^s} \quad (8)$$

Here $\varepsilon > 0$ controls the amount of spacing effect, s the degree of non-linearity (Cepeda et al., 2008). β_{min} and β_{max} are the min. and max. forgetting rates. It is also possible to change Eqn (3) to model higher practice rates as assignment deadlines and exams approach closer in more formal learning settings. Below we give a simple extension inspired from temporal motivation theory (Steel and König, 2006). Here, Γ and D_t are respectively impulsivity and time to next deadline for the learner.

$$\lambda = \frac{a + bS}{1 + \Gamma D_t} \quad (9)$$

Presenting detailed simulation results of such extensions are beyond the scope of this paper. Importantly, we also left out the social aspect of free-learning. We learn effective strategies from our peers and may get (de)motivated from comparing ourselves to them. Future work will involve incorporating more realistic assumptions in model dynamics (inevitably making it more complicated) and testing the different components of our model to real-world data.

References

Anders Ericsson, K., & Towne, T. J. (2010). Expertise. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(3), 404–416.

Anderson, R. B., & Tweney, R. D. (1997). Artifactual power curves in forgetting. *Memory & Cognition*, *25*, 724–730.

Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of mathematical psychology*, *55*(1), 25–35.

Barabasi, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, *435*(7039), 207–211.

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

Culler, E., & Girden, E. (1951). The learning curve in relation to other psychometric functions. *The American journal of psychology*, *64*(3), 327–349.

De Witte, K., Cabus, S., Thyssen, G., Groot, W., & van Den Brink, H. M. (2013). A critical review of the literature on school dropout. *Educational Research Review*, *10*, 13–28.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of learning and motivation*, *53*, 63–147.

Ebbinghaus, H. (1885). *Üabout memoryämemory: Investigations into experimental psychology*. Duncker & Humblot.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological review*, *57*(2), 94.

Franken, R., Bekhuis, H., & Tolsma, J. (2023). Kudos make you run! how runners influence each other on the online social network strava. *Social Networks*, *72*, 151–164.

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences*, *101*(36), 13124–13131.

Goopio, J., & Cheung, C. (2021). The mooc dropout phenomenon and retention strategies. *Journal of Teaching in Travel & Tourism*, *21*(2), 177–197.

Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and individual differences*, *20*(6), 644–653.

Gulliksen, H. (1934). A rational equation of the learning curve based on thorndike's law of effect. *The Journal of General Psychology*, *11*(2), 395–434.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect.

Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting mooc retention: A survey study. *Computers & Education*, *98*, 157–168.

Leibowitz, N., Baum, B., Enden, G., & Karniel, A. (2010). The exponential learning equation as a function of successful trials results in sigmoid performance. *Journal of Mathematical Psychology*, *54*(3), 338–340.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological review*, *95*(4), 492.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on psychological science*, *1*(2).

Murre, J. M. (2014). S-shaped learning curves. *Psychonomic bulletin & review*, *21*, 344–356.

Newell, A., & Rosenbloom, P. S. (2013). Mechanisms of skill acquisition and the law of practice. In *Cognitive skills and their acquisition* (pp. 1–55). Psychology Press.

Pavlik Jr, P. I., & Anderson, J. R. (2005a). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive science*, *29*(4), 559–586.

Pavlik Jr, P. I., & Anderson, J. R. (2005b). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive science*, *29*(4), 559–586.

Perc, M. (2014). The matthew effect in empirical data. *Journal of The Royal Society Interface*, *11*(98), 20140378.

Rescorla, R. A. (1972). Informational variables in pavlovian conditioning. In *Psychology of learning and motivation* (pp. 1–46, Vol. 6). Elsevier.

Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of experimental Botany*, *10*(2), 290–301.

Rickard, T. C. (2004). Strategy execution in cognitive skill learning: An item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 65.

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational psychologist*, *26*(3-4), 207–231.

Steel, P., & König, C. J. (2006). Integrating theories of motivation. *Academy of management review*, *31*(4), 889–913.

Thurstone, L. L. (1919). The learning curve equation. *Psychological Monographs*, *26*(3), i.

van Bergen, E., Vasalampi, K., & Torppa, M. (2021). How are practice and performance related? development of reading from age 5 to 15. *Reading Research Quarterly*, *56*(3), 415–434.

Vu, T., Magis-Weinberg, L., Jansen, B. R., van Atteveldt, N., Janssen, T. W., Lee, N. C., van der Maas, H. L., Raijmakers, M. E., Sachisthal, M. S., & Meeter, M. (2022). Motivation-achievement cycles in learning: A literature review and research agenda. *Educational Psychology Review*, *34*(1), 39–71.