A multiplicative reinforcement learning model capturing learning dynamics and interindividual variability in mice

Brice Bathellier\textsuperscript{a,b,1}, Sui Poh Tee\textsuperscript{a}, Christina Hrovat\textsuperscript{a}, and Simon Rumpel\textsuperscript{a}

\textsuperscript{a}Research Institute of Molecular Pathology, 1030 Vienna, Austria; \textsuperscript{b}Unité de Neurosciences, Information et Complexité, Unité Propre de Recherche 3293, Centre National de la Recherche Scientifique, 9198 Gif sur Yvette, France

Both in humans and in animals, different individuals may learn the same task with strikingly different speeds; however, the sources of this variability remain elusive. In standard learning models, interindividual variability is often explained by variations of the learning rate, a parameter indicating how much synapses are updated on each learning event. Here, we theoretically show that the initial connectivity between the neurons involved in learning a task is also a strong determinant of how quickly the task is learned, provided that connections are updated in a multiplicative manner. To experimentally test this idea, we trained mice to perform an auditory Go/NoGo discrimination task followed by a reversal to compare learning speed when starting from naive or already trained synaptic connections. All mice learned the initial task, but often displayed sigmoid-like learning curves, with a variable delay period followed by a steep increase in performance, as often observed in operant conditioning. For all mice, learning was much faster in the subsequent reversal training. An accurate fit of all learning curves could be obtained with a reinforcement learning model endowed with a multiplicative learning rule, but not with an additive rule. Surprisingly, the multiplicative model could explain a large fraction of the interindividual variability by variations in the initial synaptic weights. Altogether, these results demonstrate the power of multiplicative learning rules to account for the full dynamics of biological learning and suggest an important role of initial wiring in the brain for predispositions to different tasks.alks.

Behavior | Memory | Cue competition | Savings

It is commonly observed in animal behavior experiments, as much as in the classroom, that different individuals eventually learn the same task with strikingly different dynamics. Many factors were shown to influence learning speed and/or performance at the group level, including genetic background (1), early experience (2), or contextual factors such as stress (3). In all these cases, the underlying idea is that these factors act on synaptic plasticity mechanisms to change different parameters that modulate or gate synaptic updates, thereby modifying learning dynamics at the system scale (e.g., ref. 4). These ideas are in line with most theoretical models of biological learning (5), such as reinforcement learning (6, 7), which are based on the trial-by-trial update of mathematical variables that can be directly or indirectly related to updates of synaptic weights between neurons. In these models, one essential parameter controlling the speed of learning is the learning rate that scales the learning rule (i.e., the rule according to which synapses are updated). Other factors are also known to have an influence on learning speed such as the level of noise and the particular learning rule (8) or metaparameters that dynamically control the balance between different aspects of learning behavior like the exploration vs. exploitation or memory storage vs. renewal trade-offs (3, 9). However, importantly, all these potential variability factors represent core parameters of the system that directly influence the dynamics throughout the entire learning process.

Whereas most theoretical models efficiently capture the dynamics of group learning curves, it is known that in many basic operant conditioning protocols, individual learning curves strongly deviate from the gradually increasing and negatively accelerated learning curves resulting from group averaging (10). Individual learning curves in fact often display a step-like increase from the untrained to the trained performance level after a delay, sometimes termed the “presolution period” (11, 12), whose duration strongly varies from one animal to another. Little is known about the biological underpinning of the learning delays and of their interindividual variations. Delays were tentatively explained by a threshold between the experience accumulated by the animal and its behavioral response (12); however, the biological nature of such a threshold remains elusive.

Here, we combine theoretical modeling and learning experiments in mice and show that both the presence and the variability of learning delays across individuals can be quantitatively explained by variations of the initial connectivity between the representation of the relevant stimuli and the action selection network in a model using a multiplicative learning rule. Hence, we propose that initial connectivity could be a crucial, yet unidentified factor of variability in learning.

Results

Learning Dynamics Strongly Vary Across Individual Mice. To assess sensory learning dynamics in mice, we trained 15 inbred male mice to perform an auditory Go/NoGo discrimination task (Fig. L4). In the task, water-deprived mice had to sustain licking at a spout during a delay period following an S+ sound to obtain a water drop and to refrain from licking after an S− sound to avoid a timeout before they can initiate the next trial. The two specific sounds chosen for the task were short broadband sounds that could be detected accurately.

Significance

Learning speed can strongly differ across individuals. This is seen in humans and animals. Here, we measured learning speed in mice performing a discrimination task and developed a theoretical model based on the reinforcement learning framework to account for differences between individual mice. We found that, when using a multiplicative learning rule, the starting connectivity values of the model strongly determine the shape of learning curves. This is in contrast to current learning models based on additive rules, in which the learning speed is typically determined by a single parameter describing the ability of connections to change their strength. Our findings suggest that the particular wiring architecture in the brain strongly conditions our ability to rapidly learn a new task.

Author contributions: B.B. and S.R. designed research; B.B. and C.H. performed research; B.B. and S.P.T. analyzed data; and B.B. and S.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

\textsuperscript{1}To whom correspondence should be addressed. E-mail: brice.bathellier@unic.cnrs-gif.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1312125110/-/DCSupplemental.
and the rise periods were variable across mice, and in more than half of them the delay was even longer than the rise period. More importantly, delay and rise durations were not correlated (Fig. 2C, Pearson’s correlation coefficient = 0.13, P = 0.63), suggesting the presence of at least two independent factors determining the shape of an individual learning curve. We therefore looked for a mechanism that could explain a delay period during the initial part of the training independent from parameters that determine the speed of learning during the rise period.

**A Reinforcement Learning Model of the Behavioral Task.** To do so, we designed a minimal reinforcement learning model of the auditory discrimination task, using a formalism that eases the biological interpretation of different parameters. The model consists of three sensory units projecting onto a simple decision circuit (Fig. 3A). The activity of sensory units is described by a 3D binary vector $\mathbf{x}$. The first dimension represents the port entry (trial initiation) and captures all associations between nonspecific stimuli and the reward that may occur in the initial uncued operant conditioning. The two other dimensions reflect the presence of the S+ and the S− sounds. The decision circuit consists of a unit that linearly sums the sensory inputs and responds in an all-or-none fashion to signal the decision to lick or not to lick ($y = 0$ or 1). In addition, it receives graded feed-forward inhibition from an inhibitory unit. Units of the circuit can be thought of as populations of functionally similar neurons that we model with a single activity variable (e.g., average population activity). Similarly, the connection between units can be thought of as large populations

(Fig. 1A) that we expect mice to readily discriminate perceptually, because we observed in a previous study that learning curves for this pair of sounds were the fastest among all sound pairs tested and because the sounds could be readily decoded from activity patterns in the auditory cortex even in naive mice (13). To dissociate motor learning from sensory–motor associative learning, mice were first trained on an uncued operant conditioning task in which they learned to visit the water delivery port and to sustain licking during the delay period to obtain the reward in the absence of any auditory stimuli. The discrimination task was started when all mice obtained rewards in at least 90% of the operant trials. From this point on, mice slowly learned to differentially adjust their behavior in response to each of the two sounds (Fig. 1B), until they reached a maximal correct overall performance (i.e., the mean fraction of correct choices for S+ and S− trials) of $95.4 \pm 2.7\%$ (mean ± SD). Interestingly, the number of trials necessary to learn the task was very variable. Across the 15 mice, between 800 and 2,160 trials were needed to reach 80% performance (Fig. 1C). Whereas the mean group learning curve suggested a smooth and steady increase in performance from the start of the training, individual learning curves deviated clearly from that average as previously observed in a variety of tasks (10). Although some mice improved their performance very early in the training, we also observed several sigmoid-like learning curves displaying a delay period during which performance was at chance level (Fig. 2A), which was very similar to the presession periods described in other studies, also in symmetrical “two alternative forced-choice” tasks (12, 14). We evaluated the duration of the delay period (number of trials until the mouse reaches 20% of its maximum performance increase) and of the subsequent rise period (number of trials for the mouse to go from 20% to 80% of its maximum performance increase), based on sigmoid functions fitted to individual learning curves (Fig. 2B). The durations of both the delay

![Fig. 1. Interindividual variability in auditory discrimination learning. (A) Sketch of the Go/NoGo behavioral task. For the sound spectrograms, the time and frequency axes range from 0 ms to 70 ms and from 1 kHz to 100 kHz (logarithmic scale), respectively. (B) Population learning curve for the overall performance of 15 mice. Binning: 100 trials. (C) Cumulative distribution of the number of trials necessary for each mouse to reach 80% correct performance measured from sigmoid functions fitted to the individual learning curves.](image)

![Fig. 2. Discrimination performance increases following variable delays. (A) Individual learning curves for four mice. The red and blue lines represent the probability of correct performance for the rewarded sounds (S+) and the nonrewarded sound (S−), respectively. The black line is the overall performance (average of the red and blue lines). Binning: 180 trials. (B) Overall performance for a fifth mouse (black line) fitted with a sigmoid function (red dashed line; the black dashed lines indicate the asymptotic values). The sigmoid function is used to evaluate the delay until behavioral performance reaches 20% of the asymptotic performance increase (delay = $t_{20\%}$) and the additional number of trials needed to reach 80% of the asymptotic performance increase (rise = $t_{80\%} - t_{20\%}$). (C) Plot of $t_{20\%}$ against $t_{80\%}$ for 15 mice with the best linear fit to the data (black line). The magenta circle stands for the measurements in B.](image)
in which $\theta$ is the Heaviside step function. $\tilde{w}_E$ and $\tilde{w}_I$ are 3D positive vectors describing the excitatory synaptic weights from the sensory units to the decision and inhibitory units, respectively. The variable $\xi$ is a Gaussian noise process of unit variance that models the stochasticity of behavioral choices.

For learning, the synaptic weights are updated trial-wise according to the stimulus received ($\tilde{x}_{E,S} = [1 \ 0 \ 1]$ or $\tilde{x}_{E,S} = [1 \ 1 \ 0]$) and the result of the model’s decision ($R = 1$ for a reward, $R = -1$ for no reward). To follow common reinforcement learning models (7), we first chose to implement the following additive learning rule,

$$\delta \tilde{w}_E = a f[R - \sigma^{-1}(\tilde{w}_E - \tilde{w}_I) \cdot \tilde{x}] y \tilde{x},$$

in which $a$ is the learning rate and $y \tilde{x}$ is a Hebbian term that conditions updates to coactivation of pre- and postsynaptic units. In this implementation of the model the learning rule for the inhibitory unit is nonlocal as its update depends on the activity of $y$. Note that these equations are equivalent to a model with excitatory and inhibitory inputs from a population of sensory neurons directly impinging on the decision neuron because the inhibitory unit reverses only the sign of its input. The central term corresponds to an expectation error as used in the Resclosa–Wagner model (15) or in temporal difference learning (7). This term is the difference between the reward $R$ and the prediction $Q = \sigma^{-1}(\tilde{w}_E - \tilde{w}_I) \cdot \tilde{x}$ that corresponds to the $Q$ value of canonical reinforcement learning models (SI Methods). $\sigma^{-1}$ is a parameter that sets the asymptotic performance of the model. In contrast to canonical reinforcement learning models, we suppose that positive expectation errors are more strongly weighted than negative ones as expressed by the asymmetric function $f[u] = u$ if $u \leq 0$ and $f[u] = \nu u$ if $u > 0$ (the parameter $\nu$ is typically larger than 1). We introduced this function to account for the fast improvement of performance for rewarded trials and the slower improvement for negative trials observed in all mice (Fig. 2A). However, it is noteworthy that such an asymmetry is typically observed in mice (16) and monkeys (6) in the activity of dopaminergic neurons of the basal ganglia that code for reward expectation errors (Fig. S1). The model has three core parameters ($a$, $\sigma^{-1}$, $\nu$) to which we added three parameters describing the initial connectivity of the model at the beginning of the training (initial conditions). The first two parameters are $\tilde{w}_{E,\text{port}}$ and $\tilde{w}_{I,\text{port}}$, which are the initial weight values of connections from the “port” unit to the excitatory and inhibitory decision neuron, respectively. The third parameter is $\tilde{w}_{E,\text{sound}}$, the initial value of the four connections from the sound units to the decision circuit, which we supposed to be the same at the beginning of the training.

In this form, the model produces stochastic responses on a trial to trial basis, similar to mouse behavior. However, for simplicity and computational speed, it is more advantageous to directly compute the average response probabilities than single instantiations of the response sequence. This is done by transforming the above-described stochastic equations into deterministic probability equations as described in SI Methods.

When simulating the model (Fig. 3B), we observed no delay in the learning curves if $Q$ is initially comprised between 1 and −1, which are the asymptotic values toward which it converges when learning goes on (SI Methods). All learning curves for these initial conditions resembled inverted exponentials. The only condition in which the model produced a delay in overall performance learning curves was for much larger (or smaller) initial values of $Q$. In this situation the model initially operates in a saturation regime and

![Figure 3](image-url). Additive vs. multiplicative rule in a reinforcement learning model. (A) Sketch of the reinforcement learning (RL) model. (B) Learning curves for a model based on an additive learning rule for three different initial conditions sketched in Insets: (Top) balanced start, all initial weights are equal; (Middle) slightly unbalanced start, for the “port” unit the difference between the synaptic weights to the decision and the inhibitory unit (\(\Delta w_{\text{port,0}} = w_{\text{port}} - w_{\text{port}}\)) is initially equal to 0; and (Bottom) strongly unbalanced start, \(\Delta w_{\text{port,0}} = 5\). In the latter situation, the model initially responds only with lick decisions (arrow) until \(\Delta w_{\text{port,0}}\) decreases. (C) Learning curves for a model based on a multiplicative learning curve for three different initial conditions sketched in Insets: (Top) all initial weights initially large; (Middle) the synaptic weights between sound units and the decision circuit (\(w_{E,\text{sound}_0}\)) are 10-fold smaller than those in the Top; the low synaptic weights initially slow down discrimination learning; and (Bottom) \(w_{E,\text{sound}_0}\) is 100-fold smaller. Red and blue lines: probability of correct performance for the rewarded and the nonrewarded sound, respectively. Black line: overall performance.

Bathellier et al.
systematically chooses one of the two possible responses irrespective of the stimulus until \( Q \) has decreased enough (Fig. 3B). However, it was clear that this phenomenon could not explain the delay observed in the mouse learning curves, because none of the 15 mice was observed to systematically choose only one of the two responses at the beginning of the discrimination training (e.g., Fig. 2C). On the contrary, response probabilities in mice in the first 100 trials were often close to 50% for both S+ and S− trials (Fig. 2A and Fig. 4G). These qualitative observations indicated that the model as such could not explain the learning dynamics observed in mice.

A Multiplicative Learning Rule Can Explain Delayed Learning. We reasoned that the exponential-like learning curves in the unsaturated conditions were resulting from the near linearity of the time evolution equations of the synaptic weights (Eqs. 2 and 3), which results from the additive learning rule. In contrast, a multiplicative learning rule as used in some neural network models (17–19) or machine learning applications (20) is expected to render Eqs. 2 and 3 strongly nonlinear. Interestingly, it was recently shown that ongoing dynamics of synaptic spines (21) and boutons (22) in the mouse cortex are multiplicative rather than additive. Therefore, we decided to test whether a multiplicative learning rule would better account for the behavioral learning curves.

We changed the learning rate \( \alpha \) for each excitatory synaptic connection \( w_i \) into a weight-dependent parameter,

\[
\alpha_t = \alpha w_i,
\]

whereas the rest of the model was unchanged. With this modification, the effective learning rate depends on the current synaptic weight. Hence, if learning starts with small synaptic weights, learning is initially slow but accelerates when synaptic weights become larger and start to influence the output decision. In line with this qualitative idea, our simulations showed that the multiplicative learning rule gave rise to delays in the learning curve when initial synaptic weights are small, whereas the delay vanishes if the initial weights are large (Fig. 3C).

Reversal Learning Is Faster Than Initial Learning in Mice. We next wanted to test more rigorously the idea that multiplicative learning rules might mediate the acquisition of the auditory discrimination task. A prediction of the multiplicative learning rule is that learning is fast in situations where synaptic weights are high. For example, if learning is initially slow because of weak initial synaptic weights, a task involving the same synapses, but starting from a state where weights are high, should be learned much faster and without a delay. Such a learning situation could be achieved in mice during so-called reversal training, consisting of switching the rewarded and nonrewarded sounds (Fig. 4A). Hence, the 15 mice initially trained in the sound discrimination task were submitted to a reversal. We observed that all mice were faster in learning the reversal training than the initial training (Fig. 4B and C). Interestingly, we did not observe a significant correlation between reversal and initial learning speeds (Fig. 4C; correlation coefficient = 0.39, \( P = 0.15 \)), which is consistent with the idea that the factor causing a delay in initial training contributes much less during reversal training.

The Multiplicative Learning Rule Can Explain Fast Reversal Learning. When trying to fit the learning curves during initial and reversal training, we observed that both the additive and the multiplicative model failed to capture the learning dynamics (Fig. S2A). The reason for the failure lies in the learning rules expressed in Eqs. 2 and 3, which leads to the potentiation of only those synapses that drive correct behavior. In the case of the S+ unit this means potentiation of the excitatory connection whereas the connection to the inhibitory unit is weakened to very low levels. Hence, the connections that become relevant during the reversal learning start again from very low weights (Fig. S2B). However, we observed that processes that lead to even just a slight reduction in the specificity of the potentiation of a given connection will

![Fig. 4](https://www.pnas.org/cgi/doi/10.1073/pnas.1312125110)
endow the model with the ability to capture the full dynamics of initial and reversal learning. In a biological sense such processes can be reflected by heterosynaptic plasticity (23, 24) in which potentiation of a subset of synapses induced by simultaneous pre- and postsynaptic activity leads to potentiation of also neighboring synapses in a non-Hebbian manner (Fig. S2E). In addition, non-specificity in potentiation could be caused by incorrect targeting or residual turnover of axonal boutons (25) or dendritic spines (26) (Fig. S2C). We modeled the latter process by a synaptic diffusion term added to the learning rules,

\[
\delta w_{E} = \alpha f \left[ R - \sigma^{-1} \left( \bar{w}_{E} - \bar{w}_{I} \right) \right] x + c \left( \bar{w}_{I} - \bar{w}_{E} \right) \tag{5}
\]

\[
\delta w_{I} = - \alpha f \left[ R - \sigma^{-1} \left( \bar{w}_{E} - \bar{w}_{I} \right) \right] x y + c \left( \bar{w}_{E} - \bar{w}_{I} \right), \tag{6}
\]

where \( c \), the fraction of “diffusing” synapses, is typically more than 100 times smaller than \( \tilde{\alpha} \). With this term, even connections that are not driving correct behavior have a residual increase allowing for a fast reversal.

We tested quantitatively the ability of our multiplicative model to capture the dynamics of individual learning curves of mice during initial and reversal training. We observed a good match of the average of the individual fits and the average overall behavioral performance (Fig. 4D). Furthermore, the learning curves for S+ and S− trials were also reproduced with high precision (Fig. 4G). In contrast, the synaptic diffusion term was not sufficient in the additive model to produce a good fit when initialized with nonsaturated connectivity (Fig. 4E). The additive model always produced slower learning in the reversal (Fig. 4F), which was observed in none of the 15 mice (Fig. 4C). To precisely quantify the contribution of each feature of the full model to the goodness of fit, we measured the fraction of unexplained variance between the fitted learning curves and the behavioral measurements. This measure clearly indicated that the full multiplicative model with seven unconstrained parameters explains much more of the observed learning dynamics than either the additive model or the multiplicative model lacking the synaptic diffusion term or the asymmetric expectation error function (Fig. 4H), which also suggests that parameters \( c \) and \( \tilde{\alpha} \) do not overfit the data.

**Initial Connectivity Can Explain a Large Fraction of the Interindividual Variability.** An interesting aspect of the multiplicative learning rule is its sensitivity to initial conditions. So far we have shown fits of the model where all parameters were allowed to vary to find the optimum match to an individual learning curve. However, we noted that very good fits could also be obtained when only the three initial connectivity parameters were allowed to vary across animals while the four core parameters, including the learning rate \( \tilde{\alpha} \), were optimized with the constraint that they should be the same for all 15 mice (Fig. S4). In this setting, the learning curves of individual mice during both the initial and the reversal training could be well fitted by adjusting only the initial connectivity values at the beginning of the initial training. Initial conditions of the model could explain large differences in learning dynamics between two individuals that showed similar performance during initial training but striking differences during reversal training (Fig. 5B and C). Further quantification (Fig. S3) indicated that initial conditions are sufficient to account for a large fraction of the observed variance and can explain the lack of correlation between delay and rise durations as seen in Fig. 2C. This suggested that initial connectivity could be an important determinant of interindividual variability that can even explain nontrivial aspects of learning dynamics.

**Discussion**

We have designed a reinforcement learning model that can reproduce individual learning dynamics in a cohort of mice involved in an auditory Go/NoGo discrimination task. The model includes two crucial features: (i) a multiplicative learning rule that allows changing the learning speed with increasing experience and (ii) a process that introduces imprecision in the potentiation of synaptic connections to accelerate behavior switching after reversal (Fig. S2). The multiplicative rule creates a nonlinearity that allows fitting the sigmoid-like learning dynamics observed during initial training to the task. It is not fully excluded that another type of nonlinearity (such as a threshold on learned associations to gate their impact on behavior) would lead to similar dynamics. However, when we modeled such a nonlinearity, although sigmoid-like curves could be generated in certain conditions, the model was not able to reproduce faster learning in the reversal, and other shortcomings were observed (Fig. S4). Hence, the nonlinearity alone does not seem to be a sufficient alternative to our model.

Our model includes an asymmetry of the reward expectation error signal (27), which was essential to capture the large differences in performance for S+ and S− trials, specific to the Go/NoGo compared with two alternative forced-choice paradigms. To account for observed behavior, the error signal must be much stronger (e.g., \( \tilde{\sigma} = 6 \), Fig. 5B and C) for the presence of an unexpected reward (positive error) than for the absence of a reward (negative error), similar to the activity of the neurons suspected to signal expectation errors (6, 16). By accelerating learning for positive outcomes, the asymmetry allows the animal to collect a maximum of rewards even if the outcome is uncertain. The asymmetric rule produces two learning speeds, as in models designed to adapt to different timescales of external fluctuations (28) or models that change state according to

---

**Fig. 5.** Initial connectivity can explain a large fraction of the interindividual variability. (A) Behavioral learning curves (dashed lines) and the fit obtained with the multiplicative model (solid lines) when only the initial connectivity parameters are allowed to vary across the 15 animals. Binning: 180 trials. (B and C) Two examples of single-mouse learning curves and of the fit obtained when only the initial connectivity parameters are allowed to vary across the 15 animals. Strikingly, strong discrepancies in the reversal learning curves (arrow) can be accounted for by different initial conditions.
contextual inferences (29, 30), except that the speed is adjusted as a function of the relevance for getting rewards in our model. It is noteworthy that the asymmetry of the task and of the learning rule is not a condition for obtaining delayed learning curves. Large delays are also behaviorally observed in two alternative forced-choice tasks (12, 14) and a multiplicative learning rule is also sufficient to model delayed learning in a symmetric task (Fig. S5).

Our model can precisely account for the learning dynamics in an auditory version of the Go/NoGo task. To what extent can the model account for learning dynamics in other contexts? The model is a dynamical extension of the Rescorla–Wagner rule (15) and therefore is expected to reproduce a wide range of effects observed in classical conditioning. As an illustration, we demonstrated that the model can reproduce cue competition effects (Fig. S6) as efficiently as the original rule. Going beyond the Rescorla–Wagner model, specific dynamic effects can be modeled. One example is savings effects in relearning (31). We observed that if forgetting is modeled as a loss of precision in synaptic connections with a limited net loss of synapses, relearning with the multiplicative rule will be faster than initial learning (Fig. S7). In some experimental paradigms (32–34), reversal learning is actually slower than initial learning. As a second example, we show that slower reversal learning can be modeled by either low synaptic diffusion or large initial synaptic weights (Fig. S8). When studying learning delays, Heinemann (12) observed that the delays increase with the similarity of the two stimuli trained to be discriminated. As a last example, we show that this effect can be easily reproduced in our model by introducing correlations between the vectors representing the stimuli (Fig. S9).

What could be factors determining variability in initial wiring in a biological learning situation? It could emerge during development due to genetic or even stochastic causes, but also during postnatal learning experiences. Hence, our results suggest that multiplicative learning rules can give rise to a large variety of learning dynamics across individuals for different tasks independent of genetic factors. Intriguingly, our behavioral (Fig. 4C) and theoretical results also suggest that slow learning in a particular task does not necessarily predict slow learning in the future. If a long enough period of training is used to overcome certain weak initial connections, subsequent learning based on these connections could actually occur in a much shorter time period.

Methods
Experiments were performed with male CB57BL/6J mice and complied with the Austrian laboratory animal law guidelines (approval no. M58/001236/2010/08). Mice were trained twice a day in 30-min sessions of ~200 trials. The pretraining lasted exactly six sessions for all mice. The reversal training was initiated independently for each mouse after it had performed three discrimination sessions with more than 90% correct performance. In all graphs, the error bars indicate SEM. All analyses and simulations were performed in Matlab. The additive model is fully described by Eqs. 1, 5, and 6 and the multiplicative model is obtained by making a proportional to the synaptic weight. We fitted the response probabilities for S+ and S− trials, using a brute force approach to minimize the square error between the binned learning curves of each mouse and the output of the model (bins of 180 trials). Extended methods are found in SI Methods.

ACKNOWLEDGMENTS. We thank H. Sprekeler, A. Destexhe, N. Kaouane, and members of the S.R. laboratory for helpful discussions and comments on the manuscript and A. Biehl, M. Ziegler, and M. Colombini for technical assistance. This work was supported by Boehringer Ingelheim GmbH and a postdoctoral fellowship (to B.B.) from the Human Frontier Science Program.