nature neuroscience

Stress, genotype and norepinephrine in the prediction of mouse behavior using reinforcement learning

Gediminas Luksys[1,2], Wulfram Gerstner[1,3] & Carmen Sandi[2,3]

Individual behavioral performance during learning is known to be affected by modulatory factors, such as stress and motivation, and by genetic predispositions that influence sensitivity to these factors. Despite numerous studies, no integrative framework is available that could predict how a given animal would perform a certain learning task in a realistic situation. We found that a simple reinforcement learning model can predict mouse behavior in a hole-box conditioning task if model metaparameters are dynamically controlled on the basis of the mouse's genotype and phenotype, stress conditions, recent performance feedback and pharmacological manipulations of adrenergic alpha-2 receptors. We find that stress and motivation affect behavioral performance by altering the exploration-exploitation balance in a genotype-dependent manner. Our results also provide computational insights into how an inverted U–shape relation between stress/arousal/norepinephrine levels and behavioral performance could be explained through changes in task performance accuracy and future reward discounting.

Animal behavior is guided by rewards that can be received in different situations and by modulatory factors such as stress and motivation. Acute stress can have positive or negative effects on learning and memory that depend on stressor properties (timing, duration and relation with the task) and on the predispositions of stressed individuals[1–3]. These effects are thought to be mediated through the modulation of synaptic plasticity by stress hormones and neuromodulators, such as glucocorticoids and norepinephrine[4–7]. However, their role in high-level processes such as learning, action selection and future reward discounting is not well understood. In addition to stress, genotype[8], affective traits[9], motivation[10] and recent performance feedback[11] also influence individual performance, but it may be difficult and inefficient to explicitly model each factor to accurately predict animal behavior.

A number of models have related neuromodulatory systems to cognitive processes and to statistical quantities characterizing the environment[12,13]. Although such models provide insights into potential mechanisms, alone they are often unable to accurately predict animal behavior in a realistic situation as a result of the diversity of the modulatory factors affecting it. Here, we propose a method that can, in principle, quantify the influence of arbitrary modulatory factors on behavior as control parameters of a general behavioral model and that is exemplified here for the case of stress and genetic strain in mice.

In modeling reward-based behavioral learning, approaches based on the theory of reinforcement learning[14] have been the most successful. They have been applied to explain experimental data in animal conditioning[10], human decision-making[15] and addiction[16]. However, the modulatory role of stress and affective traits has not yet been considered. In reinforcement learning, modeled animals occupy different states corresponding to their spatial location or presence of relevant stimuli. From each state they select actions (for example, making a physical movement) on the basis of expected values of future rewards that could be acquired by taking these actions (Q values). The Q values are learned by comparing observed rewards with predicted ones and, in case of mismatch, updating the latter[14]. Such reward prediction errors have been related to the activity of dopaminergic neurons in the ventral tegmental area and the substantia nigra[17]. Dopamine is known to modulate synaptic plasticity in striatum[18], where state and action values are presumably stored[19].

Learning and action selection in reinforcement-learning models depend directly on model parameters (that is, the present Q values), whose update and use are controlled by so-called metaparameters such as the learning rate, the exploration-exploitation balance and the future reward discounting. Most behavioral modeling studies have considered the metaparameters values to be constant. It has also been proposed that they are related to specific neuromodulators, such as norepinephrine, serotonin and acetylcholine[20], and to neural activity in amygdala, striatum and anterior cingulate[21], brain areas with prominent heterosynaptic plasticity[22]. Stress, motivation and other modulatory factors act through the same brain systems as mentioned above[2,23], suggesting that, in reinforcement learning, their effects could correspond to changes in metaparameter values. In contrast with earlier studies that focused on a single metaparameter (for example, the learning rate[24,25] or the future reward discounting[26]), which they related to task uncertainty[24], genetic polymorphisms[25] or serotonin levels[26], we considered all three metaparameters and a wide range of modulatory factors in parallel.

[1]Laboratory of Computational Neuroscience, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [2]Laboratory of Behavioral Genetics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [3]These authors contributed equally to this work. Correspondence should be addressed to G.L. (eurogedi@hotmail.com).

Received 11 March; accepted 22 June; published online 16 August 2009; corrected online 26 August 2009; doi:10.1038/nn.2374

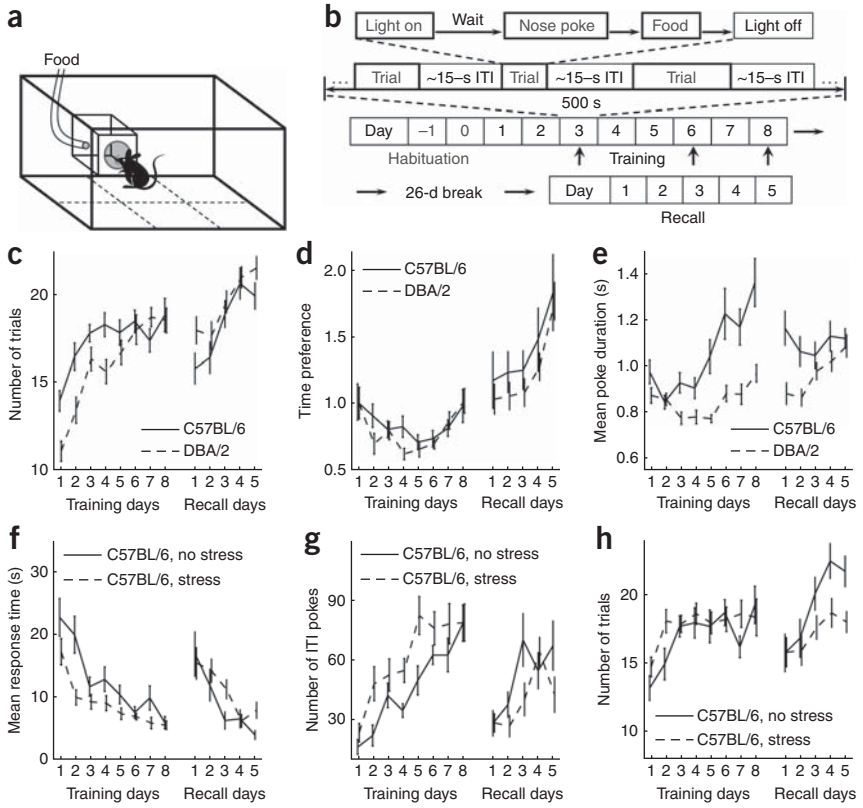

Figure 1 Experiment and behavioral results. (**a**) Hole-box setup: a mouse can move in a rectangular box. In one of the walls there is a hole in which the mouse can poke its nose and in which food is delivered if a nose poke occurs after the light in the hole has been switched on. (**b**) Experimental protocol. Following a 2-d habituation, mice were trained for 8 d; after a 26-d break they were retrained for 5 more days. Each daily session consisted of a 500-s sequence of trials and ITIs. The short vertical arrows indicate the days of pharmacological manipulations. (**c**–**h**) Dynamics of behavioral performance measures. (**c**) The number of trials performed during a session of 500 s increased during training ($R = 0.39$, $P < 0.001$). (**d**) Time preference. Mice responded to light increasingly faster, compared with the average time interval between nose pokes, during the second half of training and during recall ($R = 0.38$, $P < 0.001$). (**e**) Mean nose poke durations were higher for C57BL/6 mice than for DBA/2 mice ($F_{1,804} = 73.9$, $P < 0.001$). (**f**) C57BL/6 mice, stressed during training, responded faster than the nonstressed mice ($F_{1,238} = 17.4$, $P < 0.001$). (**g**) Stressed C57BL/6 mice also made more ITI pokes during training than those that were not exposed to stress ($F_{1,238} = 16.9$, $P < 0.001$). (**h**) The number of trials was not significantly different during training between stressed and nonstressed C57BL/6 groups ($F_{1,238} = 2.61$, $P = 0.11$). Error bars denote s.e.m.

We analyzed mouse conditioning experiments using a reinforcement-learning model. Our aim was to understand how stress, individual differences ('calm' C57BL/6 versus 'anxious-like' DBA/2 mouse strains[8]), motivation (food deprivation level) and other factors control mouse learning and memory, exploration-exploitation balance and future reward discounting. Furthermore, to study the link between reinforcement-learning metaparameters and neuromodulation, we pharmacologically manipulated the norepinephrinergic system, which is important in stress response[2] and is thought to be related to task performance accuracy[27]. We found that our model, whose metaparameters were controlled on the basis of information about modulatory factors, could predict mouse behavior in a simple conditioning task. The observed dynamics of model metaparameters provided insights into how mice adjust their performance throughout the course of a learning experience and how they respond to stressors, motivational demands and norepinephrinergic manipulations.

RESULTS

We studied mouse behavior using a simple conditioning task, the hole box (**Fig. 1a**), in which mice had to learn to make a nose poke into the hole on the onset of light and to avoid making nose pokes under no light. Correct responses were rewarded with a food pellet, whereas incorrect ones were not. Mice performed the task for 8 consecutive days and, after a 26-d break, were given the same task for another 5 d (**Fig. 1b**). Each 500-s-long daily session consisted of trials separated by intertrial intervals (ITIs) of variable duration, averaging 15 s. During the 8 d of training, selected groups of mice were exposed to stress before each daily session. On training days 3, 6 and 8, different groups of mice were treated with adrenergic alpha-2 receptor drugs clonidine and yohimbine, which affect brain norepinephrine levels[28]. During the 5 d of recall, we asked whether mouse performance was most influenced by current stress or by memories of stress experienced during training.

Modulatory factors influence task performance of the mice

We first characterized mouse performance in the hole-box task using seven different performance measures (listed in **Supplementary Fig. 1**), calculated for each daily session (over its duration of 500 s). They indicate how quickly mice learned to respond, how accurately they associated responses with light and how well they coped with anxiety to make sufficiently long nose pokes (to pick up the delivered food pellets). Thus, different performance measures reflected different aspects of behavior and thus showed different dynamics with learning (**Fig. 1c–h**). For both strains of mice, the number of trials (**Fig. 1c**) increased and the mean response times (**Supplementary Fig. 2**) decreased with training, indicating that mice had already learned to respond faster to light and thus obtain more food pellets by the first days of training. However, increases in the number of ITI pokes during the training period (**Supplementary Fig. 2**) implied that the mice also increased the number of incorrect responses during the early phase of learning. The dynamics of time preference (**Fig. 1d**) and duration preference (**Supplementary Fig. 2**) suggested that most mice started to preferentially respond to light only during the recall phase. Such relatively slow learning might be a result of uneaten pellets, left during earlier trials and picked up during subsequent ITI pokes. This aspect was explicitly included in our computational model and became useful for revealing differences between experimental groups.

We observed the differences between the C57BL/6 and DBA/2 mouse strains in mean nose-poke durations (**Fig. 1e**) and food pellets left uneaten after each session (**Supplementary Fig. 2**); DBA/2 mice made shorter nose pokes and, as a result, left more pellets. These differences might be related to the different anxiety levels of the two strains (**Supplementary Fig. 3**). For C57BL/6 mice, stress exposure before each training session led to reduced mean response times (**Fig. 1f**) and increased numbers of ITI pokes (**Fig. 1g**), whereas numbers of trials were not affected during training (**Fig. 1h**).

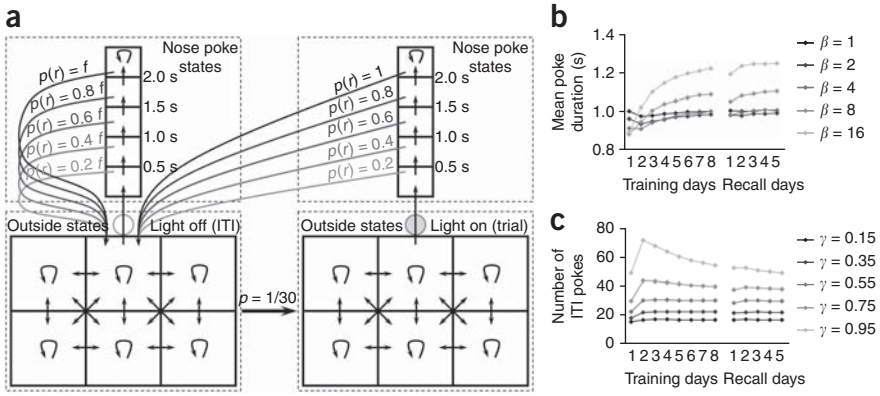

Figure 2 The hole-box model and its simulations. (**a**) State-action chart of the model. Rectangles are states and arrows are actions. $p(r)$ denotes the probability of receiving a reward and f indicates whether a food pellet is available during an ITI nose poke ($f = 1$) or not ($f = 0$). Rewards could be acquired with ITI pokes if the food pellets delivered in the previous trials were not eaten. The thick arrow denotes a probabilistic transition between ITI and trial (with 1/30 probability). (**b**,**c**) Effects of reinforcement-learning metaparameter values on model performance over the course of training and recall. (**b**) The influence of β on mean nose-poke durations ($\alpha = 0.4$, $\gamma = 0.6$). (**c**) The effect of γ on the number of ITI pokes ($\alpha = 0.4$, $\beta = 8$).

The above analyses indicate that, although conventional behavioral measures provide information about dynamics of learning and differences between experimental groups, this information is often hard to interpret, as each measure describes an unknown mixture of cognitive processes involved in learning and memory. Sometimes performing a principal component analysis (PCA) may reduce the behavioral measures to a few main components that could be easily interpreted[29]. In our case, however, PCA was not very informative (**Supplementary Fig. 1**). As an alternative to conventional behavioral analyses, we reasoned that a reinforcement-learning model could be sufficiently flexible to fit a wide range of behavioral effects and, in contrast with the performance measures, reinforcement-learning metaparameters could be easily interpreted in cognitive terms.

A reinforcement-learning model can fit individual behavior

We formalized mouse behavior using a simple model (**Fig. 2a**) in which states corresponded to the light condition (on or off) and to the mouse's position (outside referring to in the box and nose poke referring to in the nose-poke hole), whereas actions represented movements in the box as well as making nose pokes. The values for different actions were updated using the temporal difference error (equation 2 in Online Methods), with the learning rate α determining the speed of their update and the future reward discount factor γ controlling the balance between immediate and delayed rewards. Actions were selected on the basis of learned Q values and the exploitation-exploration factor β, which determined the extent to which decisions were biased toward actions with higher values. To simulate forgetting during the 26-d break, we adjusted the Q values on the first recall day on the basis of a memory decay factor, ε. Unlike conventional performance measures, such metaparameters could be interpreted in cognitive terms: α relating to acquisition intensity, β to immediate performance accuracy, γ to impulsivity and ε to long-term memory of reward predictions.

We observed that the choice of metaparameter values strongly influenced model performance (**Fig. 2b**,**c** and **Supplementary Fig. 4**). The relationship between metaparameters and performance measures was typically nonlinear, but smooth, implying that the estimation of metaparameters from observed data was unlikely to get stuck in local optima. Moreover, fixed metaparameter values often yielded performance measure dynamics that were incompatible with those observed in mouse data, suggesting that metaparameter values should be dynamically adapted with learning.

To evaluate how well mouse behavior in the hole box could be explained by our model, we estimated a set of metaparameters for each day and each mouse (or a subgroup of 2–4 mice in each experimental group) with which the model could best fit mouse behavior (**Fig. 3** and **Supplementary Fig. 5**; for the dynamics of goodness-of-fit across days, see **Supplementary Fig. 6**). Of the individually estimated metaparameter sets (**Fig. 3a**), 95% passed the χ^2 test of goodness-of-fit (satisfying $P(\chi_{\text{indiv}}^2, \nu) > 0.01$, mean $\chi_{\text{indiv}}^2 = 4.4$), meaning that our reinforcement-learning model was sufficiently flexible to reproduce mouse behavioral dynamics. The estimations for the subgroups of mice were also good (mean $\chi_{\text{pairs}}^2 = 5.5$, $\chi_{\text{triplets}}^2 = 6.8$ and $\chi_{\text{groups}}^2 = 7.7$, with 82–93% of estimated sets passing the χ^2 test). The s.d. among the five best metaparameter sets (from different optimization runs for a given day and mouse or subgroup) indicated that estimated exploration-exploitation and reward discount factors were more reliable than learning rates ($\sigma_\beta = 4.6\%$, $\sigma_\gamma = 5.2\%$ and $\sigma_\alpha = 9.5\%$ of the range, respectively).

Strain, stress and norepinephrine affect metaparameters

The estimated reinforcement-learning metaparameters showed a dependence on time, genetic strain, stress and norepinephrine manipulation groups and were consistent across different subgroup divisions (**Supplementary Table 1**). γ and β showed progressive increase (**Fig. 4a–c**), meaning that, during the course of learning, the importance of delayed rewards and the accurate use of knowledge for selecting actions increased. α decreased with training for C57BL/6 mice (**Fig. 4d**), indicating that the mice learned most intensively during the early days of the experiment.

α was slightly smaller for the stressed mice of both strains (**Supplementary Fig. 7**), suggesting that extrinsic stress could impair the acquisition process. γ was much lower for DBA/2 than for C57BL/6 mice, in particular for the stressed DBA/2 group (**Fig. 4a**), indicating that the anxious-like DBA/2 mice, especially under stress, are more impulsive than the calm C57BL/6 mice. Stress during training increased β for C57BL/6 mice (**Fig. 4b**), but not for DBA/2 mice (**Supplementary Fig. 7**). However, during recall the memories of training stress had the opposite effect: the previously stressed C57BL/6 mice now had lower exploitation factors (**Fig. 4b**). In addition, β values were higher for mice that were stressed during recall (**Fig. 4c**). This indicates that immediate stress and memories of previous stress act oppositely on the performance accuracy of C57BL/6 mice.

The results of the norepinephrine manipulations revealed the relationship of norepinephrine to several reinforcement-learning metaparameters and its interactions with genetic strain and stress. For DBA/2 mice, pharmacologically decreasing norepinephrine levels led to lower β, whereas increasing them had little effect (**Fig. 4e**). For both strains of mice (except stressed DBA/2), norepinephrine had a negative relation with γ; injections that increased norepinephrine levels reduced γ values, whereas those decreasing norepinephrine levels led to elevated γ values (**Fig. 4f** and **Supplementary Fig. 7**). This suggests that larger

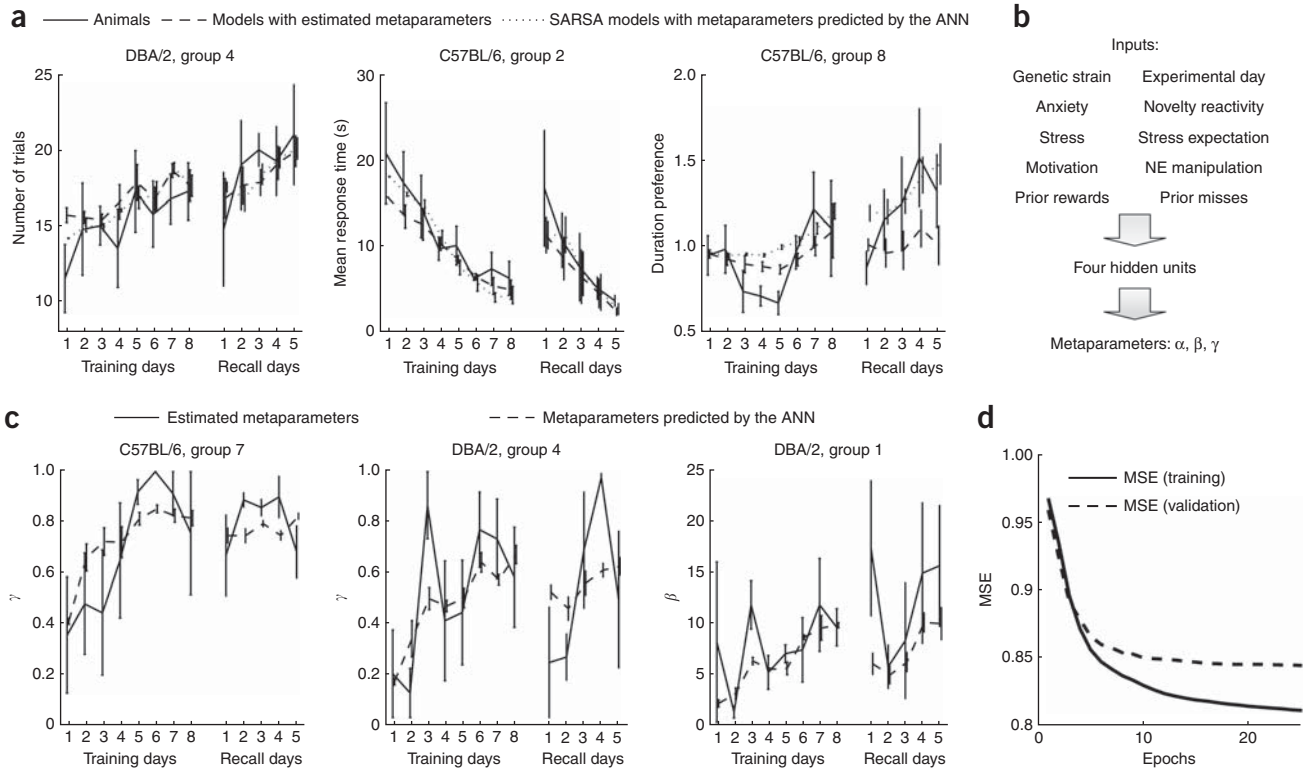

Figure 3 Behavioral performance measures can be reproduced and de-noised by the model. (**a**) Performance comparison of mice (solid lines), models with daily estimated metaparameters (dashed lines) and SARSA models with metaparameters predicted by the trained ANN (dotted lines) for selected groups and performance measures. (**b**) Scheme of the ANN for metaparameter prediction. Ten inputs describe the mice's genotype, affective phenotype, experimental condition and recent performance and three outputs are predicted metaparameters. (**c**) Comparing daily estimated metaparameters (solid lines) with predictions of the trained ANN (dashed lines) for selected groups and metaparameters (group numbers are given in **Supplementary Table 3** for **a** and **c**). Error bars denote s.e.m. (**d**) Cross-validation results indicated that there was virtually no over-fitting.

amounts of norepinephrine generally lead to higher impulsivity, consistent with previous animal and human studies[30].

Finally, ε did not show any substantial differences between strain, stress or pharmacological treatment groups. This does not exclude the possibility that our different experimental conditions affect long-term memory of learned reward values, but suggests that their effects on immediate performance accuracy and future reward discounting are more pronounced.

Modulatory factor–driven model can predict mouse behavior

In the previous sections, we examined how the estimated metaparameters differed between experimental groups. For predicting the behavior of a given mouse on a given day, information about motivation, anxiety and the previous performance of that mouse was also important. For this purpose, we constructed a multilinear regression with the mouse's strain, affective phenotype (see **Supplementary Methods** for details on phenotype characterization), stress and its expectation, motivation, pharmacological manipulations, experimental day and previous task performance as independent variables and estimated reinforcement-learning metaparameters as dependent ones. We found that, for individually estimated metaparameter sets, $R^2_{\mathrm{indiv}} = 0.15$ of the variance could be explained by known factors. For metaparameter sets, estimated from subgroup averages, this number increased to $R^2_{\mathrm{triplets}} = 0.22$, indicating that intra-group variability accounted for a substantial part of unexplained variance. The partial results for each metaparameter ($R^2_{\mathrm{indiv}}(\alpha) = 0.02$, $R^2_{\mathrm{indiv}}(\beta) = 0.21$ and $R^2_{\mathrm{indiv}}(\gamma) = 0.21$) indicated that predictions of α were the least accurate, whereas predictions of β and γ were much more accurate.

The main limitation of the multilinear regression was that it could not account for nonlinear interactions between modulatory factors. Therefore, we also predicted metaparameters using an artificial neural network (ANN) with different modulatory factors as inputs (**Fig. 3b**). For individually estimated metaparameters, 18.2% of variance could be explained (mean square error of the individually estimated metaparameters, $\mathrm{MSE}_{\mathrm{indiv}} = 0.818$), whereas the explained variance increased to nearly 30% ($\mathrm{MSE}_{\mathrm{triplets}} = 0.706$) for subgroup-based metaparameters. The resulting group averages fit the daily estimated values quite well (**Fig. 3c** and **Supplementary Fig. 5**). Nearly all of the effects observed from the comparisons of estimated metaparameters between experimental groups and partial multilinear regressions were also predicted by the ANN. This suggests that our metaparameter prediction method preserves useful information, such as how metaparameters change with learning and how they differ between experimental groups, and eliminates arbitrary variation, which cannot be explained by any input factor. To ensure that our ANN does not over-fit the training data, we trained it using random subsets consisting of 80% of the data and tested it on the remaining 20%. The results (**Fig. 3d**) indicated that there was virtually no over-fitting, and our model could therefore produce good predictions given new (unseen) data.

The final test of our model was to simulate the behavior of individual mice with the SARSA algorithm[14] (equation 6 in Online Methods) using the outputs of the trained ANN as metaparameters, instead of the

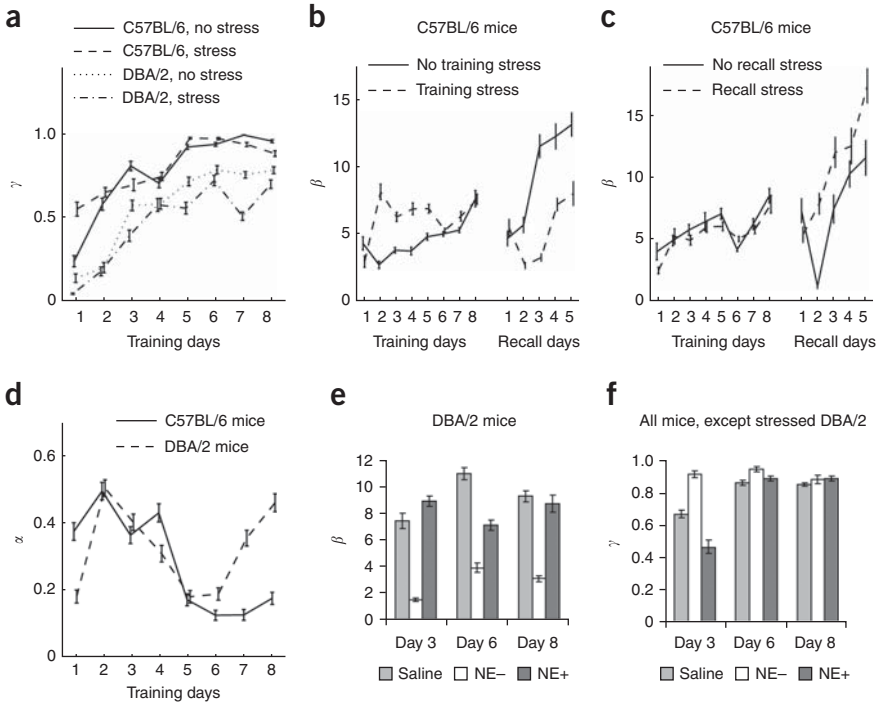

Figure 4 Genetic strain, stress and norepinephrine manipulations influence daily estimated metaparameters. (**a**) γ increased with training ($R = 0.49$, $P < 0.001$) and was higher for C57BL/6 mice than for DBA/2 mice ($F_{1,6184} = 1370$, $P < 0.001$) and lower for the stressed than for the nonstressed DBA/2 mice ($F_{1,1904} = 58.6$, $P < 0.001$). (**b**) β increased during training and recall ($R = 0.26$, $P < 0.001$). C57BL/6 mice, stressed during training, had higher β values than the nonstressed C57BL/6 mice ($F_{1,1874} = 62.0$, $P < 0.001$). However, previously stressed C57BL/6 mice had lower β values during recall than previously nonstressed mice ($F_{1,1190} = 100.1$, $P < 0.001$). (**c**) C57BL/6 mice, stressed during recall, had higher β than the nonstressed mice ($F_{1,472} = 25.9$, $P < 0.001$). (**d**) The α values of C57BL/6, but not of DBA/2, mice decreased with training ($R = -0.33$, $P < 0.001$). (**e**) Pharmacologically reducing the amount of norepinephrine (NE–) led to lower β values for DBA/2 mice ($F_{1,534} = 237.0$, $P < 0.001$). (**f**) For all mice, except the stressed DBA/2 mice, reducing norepinephrine levels led to higher γ ($F_{1,804} = 48.3$, $P < 0.001$). All plots correspond to the case of the five best metaparameter sets, estimated from subgroups containing pairs of mice. Error bars denote s.e.m.

'averaged' method used for metaparameter estimation. We found that a reasonably good fit between the model and mouse performance was preserved (**Fig. 3a** and **Supplementary Fig. 5**); 82% of metaparameter sets passed the χ^2 square test of goodness-of-fit (mean $\chi^2 = 8.8$). This indicates that our reinforcement-learning model, combined with metaparameter prediction by the ANN, can predict the performance of most mice using only previously available information about the mouse, its experimental condition and recent performance in the experiment.

Graded modulatory factors correlate with metaparameters

Individual mice in each experimental group can differ in anxiety, novelty reactivity and motivation. We studied how such graded modulatory factors influence reinforcement-learning metaparameters and how they interact with strain and stress groups. Multilinear regressions revealed correlations between modulatory factors and metaparameters β and γ, some of which occurred across all experimental groups and others that were dependent on genetic strain and/or stress condition (**Supplementary Table 2**).

As in the group-based analysis, norepinephrine changes correlated positively with β for DBA/2 mice and negatively with γ for C57BL/6 mice (**Fig. 5a**). Furthermore, the effects of norepinephrine were differentially pronounced depending on the anxiety trait (**Fig. 5b**). We also observed that the novel object–exploration trait correlated with β for DBA/2 mice: positively under stress and negatively under no stress (**Fig. 5c**). This suggests that stress affects the balance of exploration and exploitation more strongly for mice with high novelty seeking.

Although food deprivation was controlled in the experiment, small arbitrary variation was unavoidable; during the experiment mice weighed 87.7 ± 1.3% (mean ± s.d.) of their initial weight. As a result, we could study the effects of food deprivation (as indexed by body

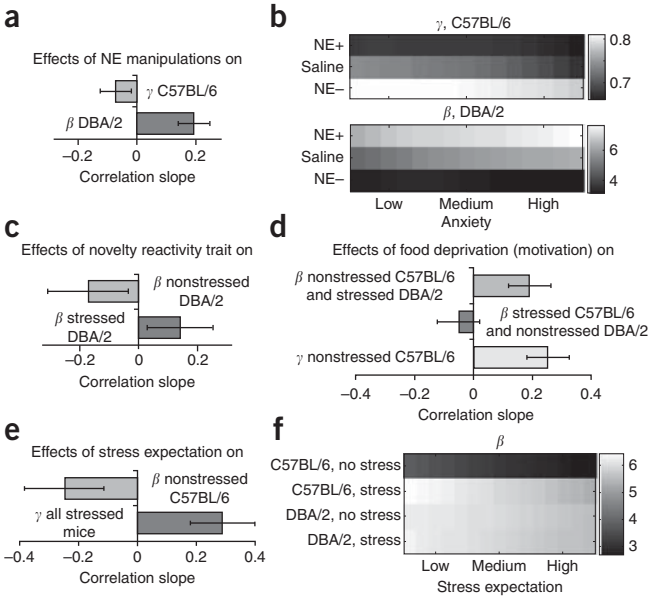

Figure 5 Multilinear regression analyses and simulations of the trained ANN reveal interactions between modulatory factors and their effects on model metaparameters. (**a**) Norepinephrine manipulations correlated negatively with γ for C57BL/6 mice and positively with β for DBA/2 mice. (**b**) More anxious mice of both strains had lower γ and higher β than less anxious mice. Increasing norepinephrine (NE+) had a stronger effect on γ for less anxious C57BL/6 mice, whereas decreasing norepinephrine (NE–) had a stronger effect on β for more anxious DBA/2 mice. (**c**) The novelty reactivity trait correlated with β for DBA/2 mice in a stress-dependent manner; mice that explored the object in the center of the open field more intensively had lower β values under no stress and higher β values under stress. (**d**) For the nonstressed C57BL/6 mice and the stressed DBA/2 mice, food deprivation (motivation) correlated positively with β, whereas there was no correlation for the remaining two groups of mice. For the nonstressed C57BL/6 mice, motivation also correlated positively with γ. (**e**) Stress expectation correlated negatively with β for the nonstressed C57BL/6 mice and positively with γ for the stressed mice of both strains. (**f**) ANN simulations indicated the relative importance of stress expectation and immediate stress in controlling β. Error bars in **a** and **c**–**e** denote 95% confidence intervals.

weight loss and related to motivation) on reinforcement-learning metaparameters. We observed that generally higher deprivation led to higher values of both γ and β (**Fig. 5d**). However, its effect differed between strain and stress groups. For nonstressed C57BL/6 mice, motivation substantially increased β, whereas the effect was smaller for mice under stress. For DBA/2 mice, it was the opposite; stress made their β values more sensitive to motivation. These findings indicate that, although C57BL/6 mice respond to extrinsic stress and increased motivation in an additive manner, the motivation of DBA/2 mice is an indicator of how they should respond to stress; the hungriest mice improved their performance accuracy, whereas less hungry mice engaged in more exploration.

We also studied how sensitively mice respond to immediate stress and to the expectation of previous stress. For nonstressed C57BL/6 mice, stress expectation correlated negatively with β (**Fig. 5e**). However, ANN simulations (**Fig. 5f**) revealed that the effect of immediate stress was relatively stronger for C57BL/6 mice. γ correlated positively with stress expectation for the stressed mice of both strains (**Fig. 5e**), indicating that mice who previously experienced stress would behave less impulsively when they experience it again.

Analyzing ANN simulations and comparing regression slopes for different modulatory factors could reveal their relative importance in controlling certain metaparameters. For example, prior rewards and misses, which are respectively the positive and the negative feedback to the mouse about how well it has been performing, both substantially increased β (**Supplementary Fig. 8**). However, γ was predominantly influenced by rewards and much less by misses.

DISCUSSION

In this study, we found that a simple reinforcement-learning model with metaparameters that are controlled by various modulatory factors could predict mouse behavior in the hole-box task. The predicted model metaparameters showed a variety of relations with stress, genetic strain, motivation and other factors, allowing us to make specific inferences about their role in different cognitive processes. Moreover, the results of pharmacological manipulations provided supporting evidence that reinforcement-learning metaparameters are indeed related to norepinephrine and brought insights into how the norepinephrinergic system affects behavioral performance.

The increase of β and the decrease of α with learning are consistent with how the metaparameters of artificial agents should presumably be controlled to achieve optimal performance[20]. One should learn more rapidly in the beginning when little information is available and more slowly later, to preserve the acquired knowledge. Similarly, a strategy involving exploration during early learning and switching to exploitation once a sufficient amount of knowledge has been acquired seems to be the best. The increase of γ with learning may be a result of several reasons. Initially, low γ values could appear because hungry mice, exposed to a new environment, are anxious about their future (for example, whether they could find food to survive), and receiving immediate rewards therefore becomes of vital importance for them. The later increase of reward discount factors may occur because high γ values are necessary for making sufficiently long trial responses and avoiding ITI pokes, thereby maximizing the acquired rewards (see **Supplementary Discussion** for mathematical justification).

Computational explanation of the inverted U–shape relation

Extrinsic stress slightly decreases the learning rate in the subsequent task, which is consistent with scarce evidence in the literature suggesting that it has impairing effects on learning[3]. More importantly, stress leads to increased exploitation of the current knowledge for C57BL/6 mice and to increased impulsivity for DBA/2 mice. The anxiety trait has similar effects: more anxious mice have higher exploitation factors and lower future reward discount factors. Motivation and stress act additively in increasing exploitation factors for C57BL/6 mice, whereas the interaction between them is more complex in DBA/2 mice. Finally, stress expectation and immediate stress affect the metaparameters in opposite ways.

The alterations of norepinephrine levels resembled the effects of stress; increasing norepinephrine levels led to lower reward discount factors, whereas decreasing norepinephrine levels (for DBA/2 mice) led to lower exploitation factors. As anxious-like DBA/2 mice have higher brain norepinephrine levels than calm C57BL/6 mice[31,32], our results imply that performance accuracy is decreased only in the case of norepinephrine reduction for mice with usually high norepinephrine levels. Our observations of the effects of norepinephrine and stress are consistent with the inverted U–shape relation theory[12], according to which increasing arousal/stress/norepinephrine up to moderate levels facilitates behavioral performance accuracy. The same theory suggests that attentional switches become frequent at very high norepinephrine levels, decreasing concentration ability and impairing focused performance. Our results provide a computational explanation for this phenomenon (**Supplementary Fig. 9**). In our experiment, pharmacologically increasing norepinephrine levels did not lead to lower β values for any experimental group. Instead, it reduced γ, impairing mice's ability to take future rewards into account during learning. Combined with high exploitation factors, this would lead to impulsive pursuit of choices that are associated with short-term rewards, which is similar to behavioral observations of animals with very high norepinephrine levels[12,27] (labile attention and excessive strategy switches).

Neural correlates and model generalization

Although our results suggest that norepinephrine influences performance accuracy and future reward discounting, it certainly cannot be excluded that other neuromodulators are important in these processes. For example, serotonin[26,33] and dopamine D2 receptors[34] have been implicated in impulsivity and control of reward discounting. Notably, DBA/2 mice, whose reward discount factors were lower than those of C57BL/6 mice, are known to have reduced expression of D2 receptors in nucleus accumbens[35]. As neuromodulatory systems are known to interact with each other[36], the effect of norepinephrine increase on future reward discounting may not be direct, but may instead be mediated through interactions with other neuromodulators, such as serotonin and dopamine.

The basis of our hole-box behavioral prediction is a simple reinforcement-learning model with discrete states and actions. The described effects of modulatory factors on model metaparameters are not only useful for the hole-box experiment, but are likely to generalize; in a different experiment in which we studied the role of stress timing (unpublished observations), most effects were qualitatively the same. Although it is not obvious that such a model could predict animal behavior in substantially more complex tasks, inferences from metaparameter estimation are probably more general; when we performed metaparameter estimation for different models of five hole-box and Morris water maze tasks, most of the resulting dynamics were similar across the two experiments[37]. Even more complex behavioral and decision-making models are likely to have a reinforcement learning-like module; therefore, a similar method could be applied for controlling its metaparameters on the basis of numerous modulatory influences. Further studies relating such metaparameters to other

neuromodulatory systems and activation patterns of specific brain areas could provide interesting insights and would be an ultimate test box for the biological relevance of such an approach.

Finally, if we are capable of predicting most important aspects of individual behavior, we can ask a reverse question: given knowledge about the individual phenotype, under which environmental conditions (such as stress, motivation and uncertainty) would it be possible to achieve the desired behavior? If a method answering such question were applied to humans in practical situations, this could be of great use for everyone. We believe that our study is a first important step toward this goal.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We would like to thank C. Rossetti and E. Gantelet for their help with experiments and E. Vasilaki for her useful comments on the manuscript. This work was supported by a grant from the Swiss National Science Foundation to C.S. by funds from École Polytechnique Fédérale de Lausanne to W.G. and by a collaborative Swiss National Science Foundation Sinergia Project to W.G. and C.S.

AUTHOR CONTRIBUTIONS

All three authors were involved in designing the study and writing the manuscript. G.L. performed the experiments, model simulations and data analyses. C.S. and W.G. supervised the project.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Kabbaj, M., Devine, D.P., Savage, V.R. & Akil, H. Neurobiological correlates of individual differences in novelty-seeking behavior in the rat: differential expression of stress-related molecules. *J. Neurosci.* **20**, 6983–6988 (2000).
2. Joëls, M., Pu, Z., Wiegert, O., Oitzl, M.S. & Krugers, H.J. Learning under stress: how does it work? *Trends Cogn. Sci.* **10**, 152–158 (2006).
3. Sandi, C. & Pinelo-Nava, M.T. Stress and memory: behavioral effects and neurobiological mechanisms. *Neural Plast.* doi:10.1155/2007/78970 (15 April 2007).
4. Sandi, C., Loscertales, M. & Guaza, C. Experience-dependent facilitating effect of corticosterone on spatial memory formation in the water maze. *Eur. J. Neurosci.* **9**, 637–642 (1997).
5. Kim, J.J. & Yoon, K.S. Stress: metaplastic effects in the hippocampus. *Trends Neurosci.* **21**, 505–509 (1998).
6. Kim, J.J. & Diamond, D.M. The stressed hippocampus, synaptic plasticity and lost memories. *Nat. Rev. Neurosci.* **3**, 453–462 (2002).
7. Sara, S.J The locus coeruleus and noradrenergic modulation of cognition. *Nat. Rev. Neurosci.* **10**, 211–223 (2009).
8. Holmes, A., Wrenn, C.C., Harris, A.P., Thayer, K.E. & Crawley, J.N. Behavioral profiles of inbred strains on novel olfactory, spatial and emotional tests for reference memory in mice. *Genes Brain Behav.* **1**, 55–69 (2002).
9. Herrero, A.I., Sandi, C. & Venero, C. Individual differences in anxiety trait are related to spatial learning abilities and hippocampal expression of mineralocorticoid receptors. *Neurobiol. Learn. Mem.* **86**, 150–159 (2006).
10. Dayan, P. & Balleine, B.W. Reward, motivation and reinforcement learning. *Neuron* **36**, 285–298 (2002).
11. Botvinick, M.M., Braver, T.S., Carter, C.S., Barch, D.M. & Cohen, J.D. Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 624–652 (2001).
12. Aston-Jones, G. & Cohen, J.D. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* **28**, 403–450 (2005).
13. Yu, A.J. & Dayan, P. Uncertainty, neuromodulation and attention. *Neuron* **46**, 681–692 (2005).
14. Sutton, R.S. & Barto, A.G. *Reinforcement Learning: an Introduction* (MIT Press, Cambridge, Massachusetts, USA, 1998).
15. Tanaka, S.C. *et al.* Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat. Neurosci.* **7**, 887–893 (2004).
16. Redish, A.D. Addiction as a computational process gone awry. *Science* **306**, 1944–1947 (2004).
17. Schultz, W., Dayan, P. & Montague, P.R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
18. Wickens, J.R., Begg, A.J. & Arbuthnott, G.W. Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex *in vitro*. *Neuroscience* **70**, 1–5 (1996).
19. Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of action-specific reward values in the striatum. *Science* **310**, 1337–1340 (2005).
20. Doya, K. Metalearning and neuromodulation. *Neural Netw.* **15**, 495–506 (2002).
21. Doya, K. Modulators of decision making. *Nat. Neurosci.* **11**, 410–416 (2008).
22. Bailey, C.H., Giustetto, M., Huang, Y.Y., Hawkins, R.D. & Kandel, E.R. Is heterosynaptic modulation essential for stabilizing Hebbian plasticity and memory? *Nat. Rev. Neurosci.* **1**, 11–20 (2000).
23. Moscarello, J.M., Ben-Shahar, O. & Ettenberg, A. Dynamic interaction between medial prefrontal cortex and nucleus accumbens as a function of both motivational state and reinforcer magnitude: a c-Fos immunocytochemistry study. *Brain Res.* **1169**, 69–76 (2007).
24. Behrens, T.E.J., Woolrich, M.W., Walton, M.E. & Rushworth, M.F.S. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
25. Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T. & Hutchison, K.E. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc. Natl. Acad. Sci. USA* **104**, 16311–16316 (2007).
26. Schweighofer, N. *et al.* Low serotonin levels increase delayed reward discounting in humans. *J. Neurosci.* **28**, 4528–4532 (2008).
27. Aston-Jones, G., Rajkowski, J. & Cohen, J. Locus coeruleus and regulation of behavioral flexibility and attention. *Prog. Brain Res.* **126**, 165–182 (2000).
28. Homayoun, H., Khavandgar, S. & Zarrindast, M.R. Morphine state-dependent learning: interactions with alpha2-adrenoceptors and acute stress. *Behav. Pharmacol.* **14**, 41–48 (2003).
29. Clement, Y. *et al.* Anxiety in mice: a principal component analysis study. *Neural Plast*, doi 10.1155/2007/35457 (21 March 2007).
30. Pattij, T. & Vanderschuren, L.J. The neuropharmacology of impulsive behaviour. *Trends Pharmacol. Sci.* **29**, 192–199 (2008).
31. Ciaranello, R.D., Barchas, R., Kessler, S. & Barchas, J.D. Catecholamines: strain differences in biosynthetic enzyme activity in mice. *Life Sci. I.* **11**, 565–572 (1972).
32. David, D.J., Renard, C.E., Jolliet, P., Hascoet, M. & Bourin, M. Antidepressant-like effects in various mice strains in the forced swimming test. *Psychopharmacology (Berl.)* **166**, 373–382 (2003).
33. Amat, J. *et al.* Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nat. Neurosci.* **8**, 365–371 (2005).
34. Dalley, J.W. *et al.* Nucleus accumbens D2/3 receptors predict trait impulsivity and cocaine reinforcement. *Science* **315**, 1267–1270 (2007).
35. Cabib, S., Puglisi-Allegra, S. & Ventura, R. The contribution of comparative studies in inbred strains of mice to the understanding of the hyperactive phenotype. *Behav. Brain Res.* **130**, 103–109 (2002).
36. Briand, L.A., Gritton, H., Howe, W.M., Young, D.A. & Sarter, M. Modulators in concert for cognition: modulator interactions in the prefrontal cortex. *Prog. Neurobiol.* **83**, 69–91 (2007).
37. Luksys, G., Knuesel, J., Sheynikhovich, D., Sandi, C. & Gerstner, W. Effects of stress and genotype on metaparameter dynamics in reinforcement learning. *Adv. Neural Inf. Process. Syst.* **19**, 937–944 (2007).

ONLINE METHODS

Animals, materials and licenses. All experiments were approved by the veterinary commission of Canton de Vaud and carried out in accordance with Swiss animal care regulations. Mice were provided by Charles River Laboratories and pharmacological agents by Sigma Aldrich. Hole-box equipment and software came from TSE Systems. Model simulations were performed using C/C++ (GNU Compiler Collection) and statistical analyses using MATLAB (MathWorks).

Experimental protocol The experimental subjects were 64 male mice (32 of C57BL/6 strain and 32 of DBA/2 strain) that were 10-weeks-old at the beginning of the experiment. During each 500-s-long session, each mouse was placed into the hole box. The mice had to learn to make a nose poke into the hole on the onset of lights (and not under the condition of no light). After a response to light, the mice received a reward in the form of a food pellet and the light was switched off, marking the end of a trial. The ITI duration was varying; the probability of starting a new trial during each 0.5-s-long time step was 1/30, resulting in the average ITI of 15 s.

During 2 d of habituation, food delivery was not paired with light (mice were given ten pellets per session at arbitrary times). After habituation, they were trained for 8 consecutive days. Half of the mice were exposed to extrinsic stress (30 min on 11 cm × 11 cm platform, elevated 90 cm above the ground) before each training session. On training days 3, 6 and 8, we injected mice intraperitoneally (5 ml per kg of body weight) 30 min before the experimental session with either saline (16 mice of each strain), the adrenergic alpha-2 receptor agonist clonidine (eight mice of each strain, 0.05 mg per kg), which reduces brain norepinephrine levels[28], or the adrenergic alpha-2 receptor antagonist yohimbine (eight mice of each strain, 1 mg per kg), which increases brain norepinephrine levels[28]. The drug doses were selected on the basis of previous studies[28,38].

After a 26-d break, the mice performed the task for another 5 d. To evaluate the effects of stress and strain on long-term memories, groups 1–6 (**Supplementary Table 3**) were not exposed to stress on recall days 1–3. On recall day 4, stress was applied to groups that received it during training to check whether performance of trained mice was sensitive to a sudden change in stress. To study the role of stress memories in modulating recall performance, we stressed groups 7 and 8 during recall days 1–3 and 5. In this manner, we could compare four different conditions, [training stress versus no stress] × [recall stress versus no stress].

Statistical procedures. To compare the values of a selected performance measure or a metaparameter between two groups, we used two-way ANOVA with repeated measures, where the within subjects factor was experimental day and the between subjects factor was experimental group. If the within subjects factor yielded significance, we tested whether the values increased or decreased with time using Pearson correlation analysis between the variable of interest and the corresponding days. To determine whether a performance measure changed significantly after the break, we compared group values between the last day of training and the first day of recall using the paired Student's t test. Significance was accepted at $P < 0.05$.

To evaluate the influence of graded modulatory factors, we performed multilinear regressions with the following inputs: strain (0 for C57BL/6, 1 for DBA/2), anxiety (fraction of time outside the center of the open field; see **Supplementary Methods**), novelty reactivity (fraction of time in the zone near a novel object), prior stress (0 or 1), stress expectation (average stress experienced during all previous experimental days, 0 for the first day), food deprivation (percentage loss of mouse weight), norepinephrinergic manipulation (–1 for norepinephrine reduction, 1 for norepinephrine increase, 0 for control), experimental day and prior rewards (the number of food pellets eaten on the previous day) and misses (the number of nose pokes during which no food was consumed). The dependent variables were metaparameters α, β and γ. All variables were normalized to zero mean and unit variance before performing the regression.

We also used partial regression models that only included data of a certain strain and/or stress group. This was the best way to study the interactions between graded factors and strain or stress. Using higher-order interaction terms in the full regression was not practical because the results would depend on which terms were included, whereas including all second-order terms would cause the model to have too many parameters, leading to poor generalization.

Implementation of the reinforcement learning model. We used a simple temporal difference model to formalize mouse behavior. Conceptually, the model consisted of four states, [ITI, trial] × [outside, nose poke], and two actions, move (in or out) and stay. To make the model's performance realistic, we introduced several extensions (see **Supplementary Discussion** and **Supplementary Fig. 10**). First, the outside state was divided into six states corresponding to different places in the box that the mouse could occupy, adding actions for the transitions between these new states. Second, when mice made trial responses that were too short, they often could not pick up the delivered food. Conversely, when the nose pokes were longer than 1.5 s, mice usually picked up the delivered pellet. To account for this, we divided the nose poke state into five states, representing different nose poke durations, with increasing probability of picking up the reward (for simplicity, we chose a linear increase, from $P = 0.2$ for the first state to $P = 1.0$ for the fifth). A food pellet was delivered at the start of each trial response, irrespective of whether the mouse picked it up subsequently or not. Unconsumed pellets could be eaten during later (sufficiently long) ITI nose pokes.

$$Q(s_t, a_t) = E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots | s_t, a_t] \tag{1}$$

The Q values (equation 1) estimate the total reward that can be gained from state s_t by choosing action a_t (r_t is the reward at time t, γ is the reward discount factor and $E[\ldots]$ denotes the expected value averaged over all possible outcomes). In analogy to offline temporal difference learning algorithms[14], the Q values were updated by

$$\Delta Q(s_t, a_t) = \alpha \cdot p(s_t, a_t) \Big\{ E[r_t] - Q(s_t, a_t) + \gamma \sum_{\forall s_{t+1}, a_{t+1}} Q(s_{t+1}, a_{t+1}) p(s_{t+1}, a_{t+1} | s_t, a_t) \Big\} \tag{2}$$

where α is the learning rate and $E[r_t]$ the expected reward at time t, given state and action probabilities $p(s_t, a_t)$. If $p(s_t, a_t)$, $p(s_{t+1}, a_{t+1} | s_t, a_t)$ and $E[r_t]$ were replaced with stochastic state, action and reward sampling, this rule would become analogous to SARSA, an online temporal difference learning rule (equation 6). The state and action probabilities $p(s, a)$ were determined on the basis of Q values and β.

$$p(s, a) = p(s) p(a | s) = \frac{p(s) \exp(\beta \cdot Q(s, a))}{\sum_{a_k} \exp(\beta \cdot Q(s, a_k))} \tag{3}$$

The sum runs over actions a_k, accessible from state s.

The starting state on each day was outside, near the hole. Q values were initialized at zero before the first training day. To simulate forgetting during the break of 26 d, we updated all of the Q values as follows:

$$Q_{new}(s, a) = Q_{old}(s, a) \cdot (1 - \varepsilon) + \langle Q_{old}(s, a) \rangle_{s,a} \varepsilon \tag{4}$$

where ε is a memory decay factor and $\langle Q_{old}(s, a) \rangle_{s,a}$ is the average of values $Q_{old}(s,a)$ before the break over all states and actions.

A single mouse performs only one action sequence; therefore the results obtained using the update rule in equation 2 indicate how a mean statistical mouse would perform under particular known circumstances. Thus, we could not predict arbitrary individual variability, but we could account for the effects of modulatory factors. We also found that simulating single action sequences over the entire experiment using the SARSA update rule[14] (equation 6) and averaging performance measures obtained from such multiple simulations led to comparable results.

Metaparameter estimation based on fit to behavioral data. In two-choice sequential decision-making tasks[24–26], model performance is typically compared with subject's behavior by directly matching the chosen actions (for example, by using a maximum likelihood criterion). In our study, mouse behavior was modeled using a large number of actions, some of which (for example, making a nose poke) were more behaviorally relevant than others.

This led to a large number of possible action sequences that could not be all explored. Therefore, we used a different approach: we compared model performance with mouse behavior on the basis of seven performance measures that describe mouse behavior during a single session.

To compare our model with mouse behavior we used the following goodness-of-fit function[39]:

$$\chi^2 = \sum_{k=1}^{7} \frac{\left(PM_k^{\exp} - PM_k^{\mathrm{mod}}(\text{metaparameters})\right)^2}{\left(\sigma_k^{\exp}\right)^2} \quad (5)$$

where metaparameters indicate α, β, γ and (for recall day 1) ε; $PM_k^{\exp}$ and PM_k^{mod} are the performance measures calculated for each mouse (or each subgroup) and the model, respectively, and $\left(\sigma_k^{\exp}\right)^2$ is the variance in the experimental data of $PM_k^{\exp}$. PM_k^{mod} was calculated after simulating the model for one session with fixed metaparameters. To evaluate whether our model is sufficiently flexible to fit various mouse behaviors, we performed an estimation procedure of daily metaparameters. For each session, we first generated 50 random sets (α, γ and ε in the range of [0.03, 0.99] and β in the range of $[10^{-1.0}, 10^{1.5}]$), of which the 15 sets with the lowest χ^2-values were selected for stochastic gradient ascent. We then performed steps of different sizes (4% and 20% of the range) along each metaparameter to further decrease the χ^2 values, terminating the procedure when no step resulted in improvement. To evaluate how well the model fits the experimental data, we used the χ^2 test with $\nu = 7–3 = 4$ degrees of freedom, as our model has three metaparameters (except for the first recall day, when it also has ε). For each session, we calculated the $P(\chi^2, \nu)$ value, defined as the probability that a realization of a χ^2-distributed random variable would exceed χ^2. Values of $P(\chi^2, \nu) > 0.01$ were necessary to pass the χ^2 test[39].

In addition to using individual mouse data for metaparameter estimation, averages of subgroups containing pairs of mice in each group (1 and 2, 1 and 3, 1 and 4, 2 and 3, 2 and 4, 3 and 4), triplets of mice (1, 2 and 3, 1, 2 and 4, 1, 3 and 4, 2, 3 and 4), or the whole group were used for reducing intra-group variability and testing the robustness of estimation procedure. To take into account uncertainty in metaparameter estimation, that is, whether only a single set of metaparameters was the best or several very different sets were similarly good, we performed all statistics (**Supplementary Tables 1** and **2**) using either the single set with the best χ^2 value or the end points of the five best gradient ascent runs from different starting points.

ANN for metaparameter prediction. To predict reinforcement-learning metaparameters on the basis of modulatory factors, we trained an ANN, whose inputs included the same information as in multilinear regression and outputs were the predicted values of α, β and γ. To avoid over-fitting, the ANN had only four hidden layer units. Its target outputs were the daily estimated metaparameter sets. After normalizing the inputs and the targets to zero mean and unit variance, the network was trained using the Levenberg-Marquardt method[40]. Because of the normalization, the resulting MSEs indicated how much variance in the metaparameters could not be explained by the ANN. For stability purposes, we trained 20 separate ANNs and used their averaged output for predicting the metaparameters.

After training the network, we performed simulations to analyze how different input factors affect each metaparameter and to discover interactions between factors. We simulated the ANN 1,000,000 times, linearly varying one or two selected inputs in their range, while all other inputs were given random values with the same statistical distributions as in the input data.

Evaluation of SARSA model performance. Our temporal difference model, used for metaparameter estimation, was intended to simulate a mean statistical mouse, as, in updating our Q values, we considered probabilities of states and actions and expected values of rewards instead of individual state-action-reward sequences. To show that simulating individual sequences for the entire experiment and averaging the resulting performance measures produced a similar result, we simulated our model using metaparameters predicted by the ANN and updating the Q values according to the SARSA rule[14].

$$\Delta Q(s_t, a_t) = \alpha[r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \quad (6)$$

On the basis of the resulting state-action-reward sequences, we calculated the performance measures and averaged them across 10,000 runs. We then used the χ^2 test to determine how well the performance of such a model fits individual mouse behavior.

38. Samini, M., Kardan, A. & Mehr, S.E. Alpha-2 agonists decrease expression of morphine-induced conditioned place preference. *Pharmacol. Biochem. Behav.* **88**, 403–406 (2008).
39. Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, UK, 1992).
40. Marquardt, D. An algorithm for least squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* **11**, 431–441 (1963).

Erratum: Stress, genotype and norepinephrine in the prediction of mouse behavior using reinforcement learning

Gediminas Luksys, Wulfram Gerstner & Carmen Sandi
Nat. Neurosci. **12, 1180–1186 (2009); corrected online 26 August 2009**

In the version of this article originally posted online, the range for β in the Online Methods section 'Metaparameter estimation based on fit to behavioral data' was given as $[10^{-1.0}, 10^{-1.5}]$. It should have been $[10^{-1.0}, 10^{1.5}]$. The error has been corrected in the PDF and HTML versions of this article.