Article

Homeostatic Reinforcement Theory Accounts for Sodium Appetitive State- and Taste-Dependent Dopamine Responding

Alexia Duriez [1,2], Clémence Bergerot [1,3,4,5], Jackson J. Cone [6], Mitchell F. Roitman [7] and Boris Gutkin [1,*]

[1] Group for Neural Theory, LNC2 DEC ENS, PSL University, 75005 Paris, France
[2] School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
[3] Charité—Universitätsmedizin Berlin, Einstein Center for Neurosciences Berlin, 10117 Berlin, Germany
[4] Institute for Theoretical Biology, Department of Biology, Humboldt-Universität zu Berlin, Philippstraße 13, 10115 Berlin, Germany
[5] Bernstein Center for Computational Neuroscience Berlin, Philippstr. 13, 10115 Berlin, Germany
[6] Hotchkiss Brain Institute, Department of Psychology, University of Calgary, Calgary, AB T2N 1N4, Canada
[7] Department of Psychology, University of Illinois Chicago, Chicago, IL 60607, USA
[*] Correspondence: boris.gutkin@ens.fr; Tel.: +33-(0)6-8631-6231

Abstract: Seeking and consuming nutrients is essential to survival and the maintenance of life. Dynamic and volatile environments require that animals learn complex behavioral strategies to obtain the necessary nutritive substances. While this has been classically viewed in terms of homeostatic regulation, recent theoretical work proposed that such strategies result from reinforcement learning processes. This theory proposed that phasic dopamine (DA) signals play a key role in signaling potentially need-fulfilling outcomes. To examine links between homeostatic and reinforcement learning processes, we focus on sodium appetite as sodium depletion triggers state- and taste-dependent changes in behavior and DA signaling evoked by sodium-related stimuli. We find that both the behavior and the dynamics of DA signaling underlying sodium appetite can be accounted for by a homeostatically regulated reinforcement learning framework (HRRL). We first optimized HRRL-based agents to sodium-seeking behavior measured in rodents. Agents successfully reproduced the state and the taste dependence of behavioral responding for sodium as well as for lithium and potassium salts. We then showed that these same agents account for the regulation of DA signals evoked by sodium tastants in a taste- and state-dependent manner. Our models quantitatively describe how DA signals evoked by sodium decrease with satiety and increase with deprivation. Lastly, our HRRL agents assigned equal preference for sodium versus the lithium containing salts, accounting for similar behavioral and neurophysiological observations in rodents. We propose that animals use orosensory signals as predictors of the internal impact of the consumed good and our results pose clear targets for future experiments. In sum, this work suggests that appetite-driven behavior may be driven by reinforcement learning mechanisms that are dynamically tuned by homeostatic need.

Keywords: sodium appetite; dopamine; homeostasis; reinforcement learning

1. Introduction

Seeking and consuming nutrients is essential to survival and the maintenance of life. Animals living in dynamic and volatile environments must develop complex behavioral strategies to obtain the necessary nutritive substances. This has been classically viewed in terms of homeostatic regulation, where complex nutrient-seeking behaviors are triggered by physiological need. Animals also seek nutrients in advance of acute need. How animals acquire nutrient-directed behaviors has most often been examined through the lens of reinforcement learning (RL) theories. In RL, subjects acquire information about signals from the environment that are associated with the receipt of reward [1]. Importantly, RL signals are distributed throughout the brain [2,3]. Similarly, physiological need impacts a

wide array of brain circuits that regulate behaviors motivated by nutrient rewards [4–6]. Intriguingly, the vast majority of RL theories do not treat the physiological origins of primary reward seeking, nor do they speak to how nutrients and their associated values are modulated by internal state. To maximize survival, physiological needs should augment signals that drive RL to promote learning in environments that offer access to essential nutrients. Thus, the reinforcing value of a nutrient, and consequently the degree to which an RL-based agent can learn from actions that acquire said nutrient, should be modulated in an appetite-dependent manner.

An essential area for exploration is thus the degree to which homeostatic and reinforcement learning processes are coupled in the central nervous system. RL processes have been most closely associated with mesolimbic circuitry, namely the midbrain DA neurons and their major target, the striatum [7–9]. While debate remains as to the role of DA in RL [10], it is increasingly clear that midbrain DA neurons and their responses to essential nutrients are modulated by physiological state, through direct hormonal influence [11–14] or via interactions with homeostatic and/or related circuits [15–20]. A particularly powerful example of the impact of physiological need on motivated behavior and DA signaling is sodium appetite. Sodium appetite is a natural behavior [21] whereby a sodium deficit generates sodium-seeking behaviors and selective consumption of sodium over other nutrients. Under homeostatic conditions, rodents avoid consumption of hypertonic sodium solutions. However, sodium depletion (via injection of a diuretic/natriuretic, e.g., furosemide) or removal of the adrenal glands [22] induces avid consumption of hypertonic sodium solutions and appetitive taste reactivity [23]. Importantly, phasic DA responses to the taste of a hypertonic sodium solution are dynamically sensitive to sodium balance [16,17]. As with behavior, the DA response in sodium-depleted rats is blocked by lingual application of the epithelial sodium channel blocker amiloride [16] and is selective for sodium solutions [17]. Lithium chloride, a notable exception, is equally preferred [17,24], likely due to sodium taste fibers responding to lithium as well (but not potassium) [25]. These data argue strongly that information related to the current state of sodium balance is communicated to midbrain DA neurons to regulate brain signals thought to drive RL. Taken together, these data pose a major challenge to current state of the art RL theories, and novel RL models need to be developed that account for the impact of physiological need and the role of gustatory information on reward learning. Sodium appetite is an ideal paradigm to address this issue.

We recently put forth a homeostatic reinforcement learning (HRRL) framework that was developed to study how animals learn need-based adaptive behavioral strategies in their environment to obtain rewarding outcomes [26,27]. The HRRL agent learns to maximize the total cumulative reward by performing actions and predicting the impact of their outcome on its internal state. This framework relies on a new definition of rewards: the rewarding value of an action is a function of the predicted impact on the difference between the current internal state and the ideal one (i.e., "setpoint"). The function that links the internal states to rewards is called the drive function. In other words, the reinforcing value of a stimulus is modulated by the degree to which it alleviates or exacerbates a physiological need. In this way, HRRL joins the predictive homeostatic regulation and reinforcement learning theories by positing that minimizing deviations from a homeostatic setpoint and maximizing reward are equivalent. In other words HRRL synthesizes RL algorithms with the drive reduction theories of motivation [28]. HRRL has been used to simulate the consumption of various resources and reproduce experimental data. It can also be used to represent complex behavior such as anticipatory responding, binge eating [26] and cocaine addiction [29]. Interestingly, it can be shown mathematically that HRRL agents show predictive allostatic behavior and HRRL accounts for the incentive salience proposals: the internal state of the HRRL agents is dynamically changed according to upcoming challenges and the action values (incentives) are modulated dynamically by the internal state of the animal.

Here, we show that the HRRL model can account for sodium-seeking behavior and DA signaling in rats. We first required the HRRL models to reproduce behavioral data showing that sodium-deprived rats preferred sodium and lithium over potassium solutions. We then showed that such HRRL agents also reproduce the dynamics of DA signals. We then used the models to make several predictions about satiety-dependent modulation of behavior and how exposure to lithium may only modulate the behavior and the reinforcing value sodium.

2. Methods

2.1. HRRL Theory for Sodium Consumption

2.1.1. State Space Representation

The internal state is considered a continuous variable that can be represented at each time t by a point in a homeostatic state space. As theorized by Keramati and Gutkin [26], this state space has one dimension per homeostatic variable. The ideal internal state is the equilibrium point of the homeostatic state space. This point we call the setpoint represents the internal state that maximizes the chances of survival (satiety). It is denoted by $H^* = (h1^*, h2^*, \ldots, hN^*)$. In this study, the state space has only one dimension, corresponding to the internal sodium level.

2.1.2. Reward Calculation Mechanism

The HRRL theory provides a function called the drive, which takes as its argument the degree of departure from a "satiety" point and has a unique minimum at that setpoint. The drive is a function of the deviation of the animal's internal state H_t from its homeostatic setpoint H^* (Figure 1). In a homeostatic state space with one dimension, the drive is given by the following expression [26]:

$$D(H_t) = \sqrt[m]{|h^* - h_t|^n} \tag{1}$$

where t represents the time, m and n are free parameters that influence, non-linearly, the mapping between homeostatic deviations and the rewarding value of their reduction.

As an animal performs an action, its internal state is modified by the outcome K_t of the action. The homeostatic reward is defined in a non-circular way, as the reduction in the homeostatic distance from the setpoint caused by the outcome K_t.

$$r, \, K_t = D(H_t) - D(H_{t+1}) \tag{2}$$

$$r(H_t, K_t) = D(H_t) - D(H_t + K_t) \tag{3}$$

The reward associated with taking an action from state H_t resulting in an outcome K_t that transitions the internal state to H_{t+1} is positive if the subsequent internal state (H_{t+1}) remains below or equal to the setpoint However, if the animal is currently at its setpoint (i.e., $H_t = H^*$), the reward value obtained with the outcome K_t is negative: the outcome is negatively reinforcing.
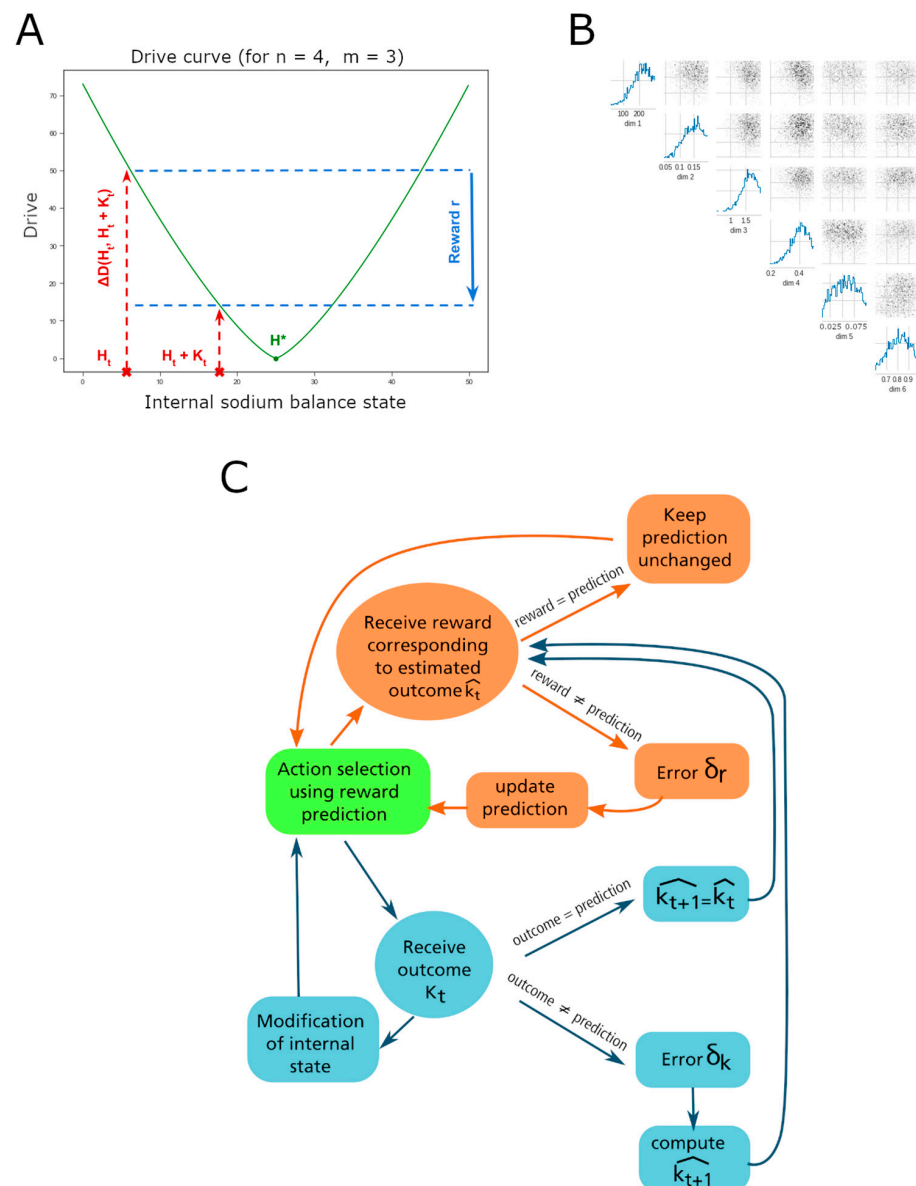
Figure 1. (**A**) Drive as a function of the internal sodium level. The optimal sodium level is denoted by H^*, indicated by a dot. The current sodium state of the agent at time t is H_t. An action apports an outcome that impacts the sodium level denoted by K_t, transitioning the internal state to $H_t + 1 = H_t + K_t$. The change in the drive function is defined as the reward $r = \Delta D (H_t, H_t + K_t)$. (**B**) Probability distributions for parameter values yielding model results consistent with the experimental observations. Parameters are, from top to bottom: the setpoint H^*, the learning rate ϵ, the rate of exploration β, the fixed outcome K, the loss of sodium after each trial and the energy cost of drinking. If each parameter takes a value with a high probability, the simulation results are consistent with the experimental observation. For each pair of parameters, the joint probability distribution that the two parameters fall in their respective range of possible values is also represented. Distributions obtained following Gonçalvez et al. [30], see Methods for details. (**C**) Diagram of the agent computations during a single trial (adapted from Schultz [31]). After performing an action, the agent receives a reward computed with the drive function and the estimated outcome of the action. The reward prediction error will be used to update the predicted value of this action and the probability of the agent to choose it at the next trial. In the meantime, the agent receives the outcome of the action, which modifies its internal state and its prediction of this outcome.

2.1.3. Taste Value Estimation Mechanism

We hypothesize that animals sense the rewarding value of a tastant through the gustatory information they receive before experiencing its post-ingestive qualities [26]. The reward is thus computed by the animals' orosensory approximation of the nutritional value K_t given the amount of solution they consumed. This estimate of the outcome, based on the orosensory properties of the stimulus, is denoted by $\hat{K}_t$. The reward therefore becomes:

$$r\left(H_t\,,\,\hat{K}_t\right) = D(H_t) - D\left(H_t + \hat{K}_t\right) \tag{4}$$

We further hypothesize that $\hat{K}_t$ is not constant. In our model, $\hat{K}_t$ is learned with a learning rate ϵ through tasting the corresponding solution and experiencing its impact on the internal sodium level H_t. We introduced this aspect since it has been suggested that assigning a reinforcing value to a taste of food requires that the animal experiences its nutritional impact [32]. According to this study, hungry animals learn that a taste stimulus is predictive of a need-reducing reward by experiencing the association between these two properties of food. We consider that $\hat{K}_t$ also has an innate non-zero initial value, which is supported by recording of DA transients: the first NaCl infusion already elicits a DA response [16].

$$\hat{K}_t = \hat{K_{t-1}} + \epsilon \delta_k \tag{5}$$

$$\delta_k = K_t - \hat{K}_t \tag{6}$$

It has been shown that in the absence of prior experience, sodium-depleted rats cannot discriminate between sodium and lithium chloride [17]. We therefore hypothesize that taste information alone is insufficient to distinguish between sodium and lithium in the absence of any post-ingestive impact on D. In our model, this means that, for naïve animals, sodium and lithium have the same $\hat{K}_t$, which represents the estimated impact of a sodium-like tasting solution on the internal state.

2.1.4. Action Value Estimation Mechanism

With the Q-learning method, rats estimate the value v of each choice as they discover which actions are more rewarding than the others [33] Once a rat executes an action a and the homeostatic reward r is computed, the value $v(a)$ of this action is updated using the reward prediction error (RPE), δ_r, with the learning rate ϵ [26].

$$v(a) \; \leftarrow \; v(a) + \epsilon \delta_r \tag{7}$$

$$\delta_r = r(H_t,\,\hat{K}_t) - v(a) - cost \tag{8}$$

The cost is a penalty we introduced in this model associated with the energy cost of approaching the sipper tubes and consuming any of the solutions. By decreasing the reward prediction error term (RPE), it reduces the reinforcing value of an action, and thus the motivation of the agent to pursue that action in the future. We assume that the cost of approaching and drinking from a sipper tube is a priori encoded in the rats' representation of their environment.

δ_r is the RPE signal that is purportedly encoded by midbrain dopaminergic neurons (e.g., see [34]). We therefore monitor the RPE by sampling DA fluctuations in the Nucleus Accumbens (NAc). A positive RPE in our model corresponds to a phasic DA response in the NAc. The RPE signal is negative when the predicted reward is superior to the actual one. The RPE can also be negative for actions that yield no reward due to the energy cost associated with performing said action.

2.1.5. Action Choice Mechanism

The probability of taking an action a_i is proportional to its estimated value $v(a_i)$, according to the softmax rule [33]:

$$P_t(a_i) = \frac{exp(\beta v(a_i))}{\sum_i^{N_a} exp(\beta v(a_i))} \tag{9}$$

N_a is the number of possible actions, such as drinking sodium chloride or drinking nothing. The probability distribution over the possible actions is the stochastic policy used by the simulated rats to choose their next action. β is a parameter used to modulate the probability of selecting actions with different estimated values. A high beta causes the selection probabilities to diverge faster. For extreme betas, this can make the action choice almost deterministic (i.e., greedy). Conversely, a low beta makes the probabilities change more slowly, and actions are therefore selected more randomly. In general, β controls exploration versus exploitation behaviors.

2.1.6. Trial Schedule

A new trial begins every 15 s. At the beginning of a trial, the rats choose an action based on the learned stochastic policy. The outcome of the action is added to the internal state.

$$H_t = H_{t-1} + K_t \tag{10}$$

The new drive, then the homeostatic reward, are calculated. The RPE is computed, then the value of the action performed is updated. Meanwhile, the values of the other actions remain unchanged. The error δ_k in the approximation of K_t is also computed, allowing $\hat{K}_t$ to be updated. The policy is then updated using the new action values. Finally, the internal state loses a small quantity of sodium, to simulate grossly the dynamics of sodium in a real organism:

$$H_t \leftarrow H_t - loss \tag{11}$$

2.1.7. Optimal Parameters

We first used an ad hoc procedure to adjust the free parameters of the model to yield a qualitative agreement with the data and to determine physiological ranges that would indeed produce such appropriate model behavior. The "Simulation-Based Inference" approach was then employed to find optimal parameter distributions, within a range of possible biological values estimated with the search by hand, for the algorithm to return a desired output [30]. The parameters used in the simulations in the manuscript were then sampled from the modes of those distributions to produce individual simulated animals (agents).

For clarity we briefly review the SBI methodology. The Simulation-Based Inference method uses an algorithm called "Sequential Neural Posterior Estimation" (SNPE). SNPE requires three inputs [30]:

- A simulator, which is the learning algorithm (an HRRL agent), returning one output of interest. The preference score of potassium chloride when sodium and potassium are available was chosen as the output of the simulator. This preference score for a solution was defined throughout this study as the volume of this solution that was consumed during the experiment, divided by the total volume of solutions consumed [17].
- Prior knowledge on the free parameters, in the form of a uniform distribution over the possible values each can take, within a defined interval. The maximum and minimum values each parameter can take are chosen based on biological plausibility, previous models and the set of values found by hand.
- The observations of the experimental data, which we want to reproduce as closely as possible with our model, is the preference score for potassium relative to sodium at the end of a 10 min two-bottle intake test that should match the data to be 0.1.

SNPE returns for each parameter a probability distribution over its range of possible values. The values with a high probability are consistent with the observation. Simulation-Based Inference was used to determine the values of the setpoint H^*, the learning rate ϵ, the rate of exploration β, the fixed outcome K of every drink taken by the rats, the loss and the cost. In total, 2000 samples were drawn from 1000 batches of 1000 simulations via the SNPE method to obtain the distributions (Figure 1B). Each parameter was assigned the average value of the distribution, as an approximation for the value with the highest probability.

For each simulated agent, we recorded the evolution of multiple parameters or values: (1) the probability of consuming one solution over another, (2) the amount of ingested sodium chloride, (3) the predicted nutritive value associated with the taste of sodium, (4) the reward prediction error signal and (5) the choices made throughout the simulated experiments. Individual agents were endowed with parameters that were sampled from the peaks of the posterior distributions (see above), hence, representative instantiations of optimal parameter sets. The agents were then used to show how the studied variables evolve depending on the solutions available and the initial internal state. Interagent variability was introduced to calculate the statistics (e.g., averages) of the relevant variables over several different individual agents (to match the experimental data). To generate individual agents, we sampled the free parameter distributions within the ranges giving the highest probability of the agreement with the data. This way we obtained a simulated cohort of animals, model agents, for which we could compile response statistics. The free parameters of the HRRL model are the setpoint, the learning rate, the exploration rate and the loss of sodium between each trial. For each of these parameters and individual agents, values were randomly sampled within a range in which the probability of the learning algorithm output, to be consistent with the behavioral data from Fortin and Roitman [17], was around its maximum (approx. 1 std of the peak, see Figure 1B).

2.2. Statistical Analysis

In the simulated experiments, several variables were calculated in 14 simulated rats having access to potassium and 15 simulated rats having access to lithium: the probability of choosing each of the available solutions at the end of the experiment, the average reward prediction error throughout the experiment, the predicted nutritive value of one lick of NaCl or LiCl (K_t) and the number of trials necessary to reach the optimal sodium level. For each of these variables, the averages of the two groups were compared with a two-tailed Welch's *t*-test. The preference scores for potassium group (n = 14) and lithium group (n = 15) were compared using a two-tailed Student's *t*-test. The cumulative consumption between the two groups was compared with a two-tailed Student's *t*-test. This analysis aims to reproduce analysis methods used by Fortin and Roitman [17].

3. Results

We first optimized the free parameters of the model (see Methods) to generate a cohort of individual agents whose simulated behavior reproduces the results of the 10 min two-bottle intake test conducted by Fortin and Roitman [17]. The parameters sampled and optimized were the setpoint, the learning rate, the exploration rate and the quantity of sodium lost between each trial. They were tuned for the potassium preference score of the simulated sodium-deprived agents to be as close as possible to the experimentally observed values. Our goal was to capture the preference scores for the different salt solutions as a function of the animal's internal state and choices in the experiment. To illustrate the results, we picked a parameter set (see Appendix A Table A1) with nearly optimal parameter values producing a simulated individual "rat" (N.B. for the rest of the manuscript we will denote such simulated animal by agent). Figure 2 shows the behavior of such an individual agent throughout the experiments, and indicates that the model accounts for the qualitative observations made by Fortin and Roitman [17] on the preference for sodium and lithium over potassium. In Figure 2A, the evolution of the probability for an agent with optimized free parameters to consume either NaCl, KCl or

nothing (left column), and the evolution of the probability for the same agent to consume either NaCl, LiCl or nothing (right column) are shown. In both simulations, the agent was initialized with a simulated depleted sodium internal state variable. As expected, when sodium and potassium are available, the probability of consuming NaCl increases greatly, while the probability of KCl consumption decreases and becomes lower than the chances of drinking nothing. This suggests that sodium-depleted agents selectively target the taste of sodium in their strategies to regulate their sodium level (as seen in the experiment). On the other hand, the probability of consuming NaCl or LiCl remains around 0.4, higher than the probability of drinking nothing. This indicates that the taste of sodium and lithium are both expected to restore sodium balance.

The individual choices made by the agent are also shown in Figure 2A(a,e). This allows us to understand how the preference for sodium relative to a non-sodium salt evolves depending on the learned subjective value of each solution and the deviation from the setpoint (Figure 2A(b,f)). We can make an inference about the dynamics of DA signaling in the NAc by studying the reward prediction errors (RPEs) of the agents during the simulated experiments (Figure 2A(c,g)). Notice that NaCl and LiCl consumption gives rise to positive RPE signals, which correspond in our model to the experimentally observed positive DA responses in the NAc. Note that in the NaCl/LiCl task, the RPE signals are rapidly quenched, as the agent rapidly learns that both of these outcomes have an equal orosensory quality. We also tracked how the approximation of sodium nutritive value may evolve throughout the experiments (Figure 2A(d,h)). We can see that when NaCl and KCl are present, the value of NaCl choice increases (the value of KCl decreases, result not shown), whereas in the NaCl/LiCl task, the value of NaCl stays relatively constant since their orosensory quality is equal. Indeed, our HRRL agent simulations show how DA dynamics are linked to internal state fluctuations, and also how the animals' learned estimation of the nutritional value is based on gustatory information.

The model quantitatively reproduced the experimental results [17] for the amount of ingested liquid and the individual preference for sodium, lithium or potassium. Figure 2B(a,b) show that the average amount of liquid ingested during the experiment and the preference scores for the non-sodium solutions obtained with the simulations are consistent with the experimental values. We further confirmed qualitative observations for simulated individual agents (shown in Figure 2A): at the end of the 10 min experiment, rats (as well as the simulated agents) with access to NaCl and LiCl (N = 15) are, approximately, equally likely to drink either of the two solutions. Rats with access to NaCl and KCl (N = 14), however, are unlikely to drink KCl, with a choice probability close to zero (see Figure 2B(c)).

Based on the behaviorally validated model, we asked what would be predicted for the dopaminergic signal (in our model this corresponds to the PRE) and the predicted value associated with the sodium gustatory cue during the final part of the experiment. Figure 2B(d) shows that the inter-individual-average phasic DA signal strength (as measured experimentally by the DA concentration) in the NAc during the 10 min experiment is higher when KCl is available than when LiCl is present as a choice, while in Figure 2B(e) we show the average predicted value of a lick of sodium (or lithium) associated with the corresponding taste for the model agents at the end of the 10 min simulated experiment. The true nutritive value of one lick of sodium is indicated with a dotted line. Our simulations predict that rats learn the true value of sodium associated with its taste when NaCl and KCl are available. However, their estimate is equal to about half of the true value when NaCl and LiCl are available. This higher value of the rats' estimated value of sodium and lithium when KCl is available may explain why the average RPE is also higher in this condition as compared to the LiCl-available condition (Figure 2B(d)). On the other hand, when NaCl and LiCl are both available, the taste-dependent rewards are equally distributed between the two solutions, and hence the RPE for NaCl is reduced and a positive RPE is misattributed to the LiCl choice (in the sense that it is based purely on the taste information and not on the internal impact). Hence LiCl acquires a positive predicted value at the expense of the NaCl choice.
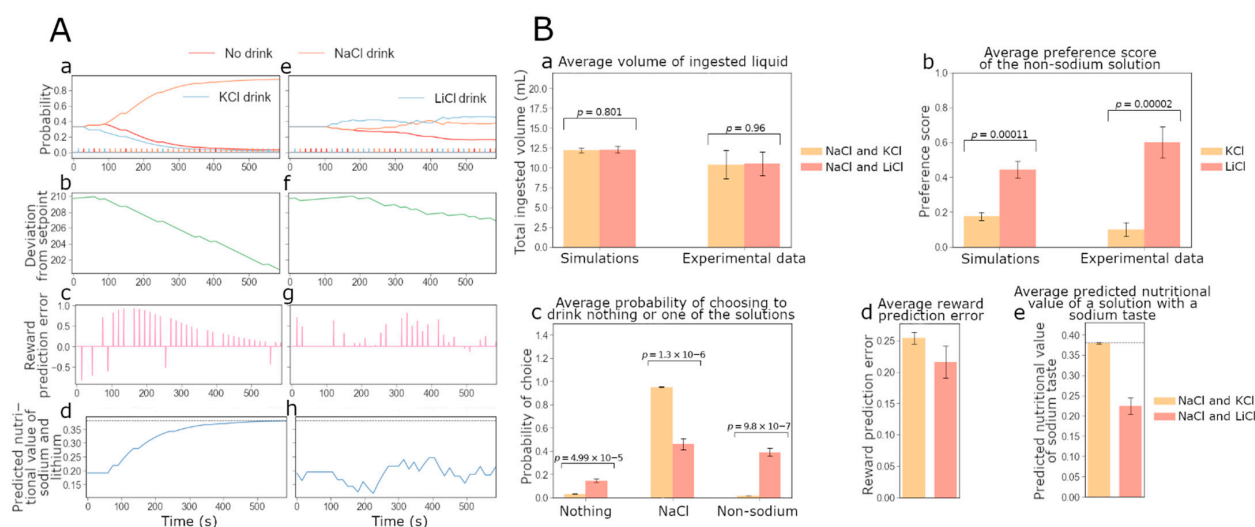
Figure 2. Simulations for the two-bottle intake test [17] are consistent with experimental observations on sodium appetite. Simulation results are shown for experiments where sodium-depleted rats have access to NaCl and KCl (left column, (**A**(**a–d**))), and in the other version, NaCl and LiCl are the solutions available (right column, (**A**(**e–h**))). Both versions of the experiment last 10 min (40 trials). (**A**) (left column): (**Aa**) Evolution of the choice probability for a simulated agent with optimized free parameters to consume NaCl, KCl or nothing over time (left). Evolution of the probability for an agent with optimized parameters to consume NaCl, LiCl or nothing over time (right). Note that the two simulated individuals are identical in the values of the free parameters: the discrepancy between their behaviors is due to the different solutions available to them. Note that the bars at the bottom indicate the individual choices (color codes the choice). (**Ab**) Amount of sodium ingested over time. Note that for these simulations, the optimal level of internal sodium for survival is set at 212 in arbitrary units. (**Ac**) Reward prediction error signals over time. Note that in our simulations the RPE is a proxy for the phasic dopamine signal. (**Ad**) Evolution of the approximated nutritional value $\hat{K}_t$ for sodium and lithium. (**A**(**e–h**)) panels show results corresponding to the left column but with NaCl and LiCl as choices. (**B**) Statistics of group simulation data (N = 28) under the two-bottle intake test (Fortin and Roitman [17]) accounts for quantitative experimental results on sodium appetite. As in the original study, 14 simulated rat agents take the version with KCl and 15 take the version with LiCl. The free parameter values for individual agents (N = 28) are all different and are sampled from the distributions near that optimized values used for example simulations in panels (**A**). (**Ba**) Average volumes of ingested liquid by rats (N = 28) during 10 min, in the simulated experiments and data (modified from Fortin and Roitman [17]). Intake volumes for rats that have access to NaCl and KCl (N = 14; yellow), and intake volumes for rats that have access to NaCl and LiCl (N = 15; red). (**Bb**) Average preference scores of rats (N = 28), in the model simulations and in experiments (modified from Fortin and Roitman [17]). Preference scores for KCl versus NaCl (N = 14; yellow), and preference scores for LiCl versus NaCl (N = 15; red). (**Bc**) Average probabilities (N = 28) to drink NaCl, the available non-sodium solution (KCl or LiCl) or nothing at the end of the 10 min experiment. Probabilities under access to NaCl and KCl (N = 14) are shown in yellow, and under access to NaCl and LiCl (N = 15) are shown in red. (**Bd**) Average reward prediction error (N = 28) during 10 min. The RPE under access to NaCl and KCl (N = 14) is shown in yellow, and the reward prediction error for access to NaCl and LiCl (N = 15) is shown in red. (**Be**) Average predicted value of a lick of sodium (or lithium) based on its taste (N = 28) at the end of the 10 min simulated experiment. This value for access to NaCl and KCl (N = 14) is shown in yellow, and it is shown in red for rats that have access to NaCl and LiCl (N = 15) ($p = 1.16 \times 10^{-5}$). The true nutritive value of one lick of sodium is indicated with a dotted line.

We then set out to see if the behavior-optimized model could also account for experimentally observed NAc DA responses to intraoral NaCl delivery in sodium-depleted rats and how this response develops as the animal continues to consume NaCl. Figure 3 shows

a simulated agent with parameters optimized to account for the experimental behavioral preference scores, as in the previous simulations. We can see that the simulated agents' RPE signals are qualitatively consistent with the experimentally observed DA responses. Cone et al. [16] measured the average DA response evoked by ten intraoral infusions of NaCl in sodium-depleted rats (N = 5). They observed that the first five infusions evoke larger DA responses than the last five. Figure 3A shows the simulation results for the same experiment alongside the summary data from Cone et al. In order to simulate this experiment, we first measured the RPE for the naïve depleted model during the first five simulated trials of exposure to NaCl stimulus, then during the last five trials. The model was "depleted" by initializing the internal state significantly below the optimal value (variable H_t = 2). NaCl delivery was simulated as a series of 10 trials where the agent receives the palatable sodium reward. In the model, the impact of NaCl on the internal state was simulated by a dose-dependent shift in the internal state variable (here of $K_t = 0.3819$ per injection) and increase in the taste signal by $\epsilon \delta_k$, with $\delta_k = K_t - \hat{K}_t$ (see methods). The results were averaged over five agents whose free parameter values were randomly sampled as previously presented. To compare the RPE with biological results, we referred to Cone et al.'s experiments [16]. With their data, we calculated the baseline DA concentration for each trial and each rat as the average DA concentration during the 4 s preceding the NaCl infusion period. We subtracted this baseline from the DA concentration values during the 4 s infusion period. Then, for each rat, we computed the average peak of the DA concentration increase over the first and the last five trials. These two variables were finally averaged over the five rats. In order to compare the DA signal strength to the RPE, we normalized the DA signal from the first five trials and the RPE from the first five trials to unity. Consistent with the experimental results, we observed a decrease in the response between the first group of trials and the last. Indeed, Welch's *t*-test showed that the difference between the average responses to the first five trials and to the last ones was significant ($p = 6.93 \times 10^{-7}$). Moreover, the differences between the simulation and the experimental results were not statistically significant (Welch's *t*-test; see p-values on the figure). We then asked if the model can also account for the differential taste dependence of the phasic DA response. Fortin and Roitman (2018) tracked the average DA response of sodium-depleted rats during 10 intraoral infusions of NaCl, of KCl, of LiCl or of water and found that NaCl and LiCl evoke a strong DA response, while KCl does not. We simulated an analogue experiment. Figure 3B compares the simulated and experimental results. Note that in our simulations we left out the response to water since that would, in the simulations, lead to a null result. With the experimental data from Fortin and Roitman [17], we computed the peak of the DA concentration during the 4 s infusion period and subtracted the average DA concentration measured during the 4 s before the infusion onset as a baseline. This baseline was computed for each rat. We then averaged the DA increase over the rats. For the simulated data, we computed the average RPE signal evoked by 10 infusions of each solution. The results were averaged over six or five agents whose free parameter values were randomly sampled as previously. The baseline is null in our model. Then, in order to compare the DA data to the RPE, we normalized the response to NaCl to unity. We showed that in our simulations NaCl and LiCl are also the only solutions triggering a DA signal. Welch's *t*-test showed that the differences between the simulation and experimental results are not statistically significant.
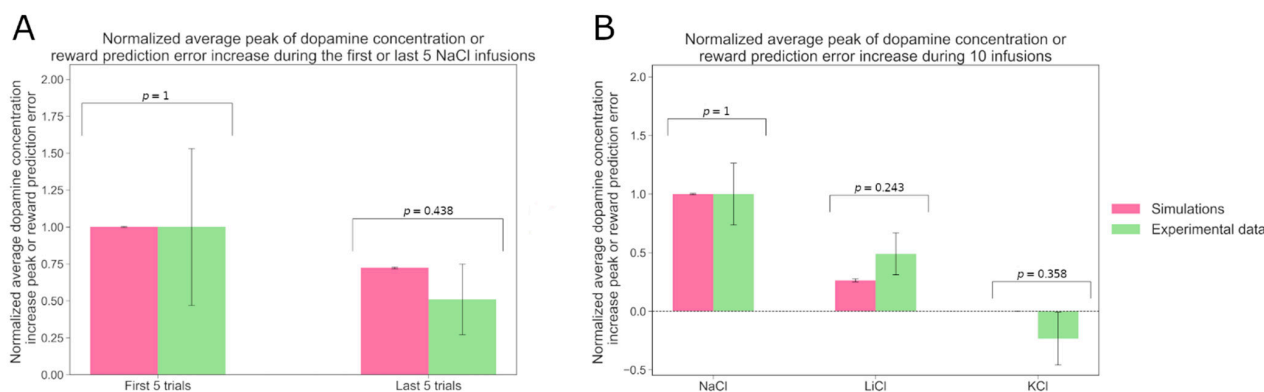
Figure 3. The model accounts for experimental data on dopamine dynamics. (**A**) Comparison of the sum of the dopamine concentration increase during the infusion period, averaged over 6 or 5 rats, and averaged over the first 5 or last 5 of a series of 10 intraoral NaCl infusions, as reported by Cone et al. in 2016 (in green), with the average RPE during the first 5 or last 5 of 10 successive simulated infusion trials, averaged over 6 or 5 rats (in pink). (**B**) Comparison of the sum of the dopamine concentration increase during the infusion period, averaged over 6 or 5 rats, as reported by Fortin and Roitman in 2018 (in green), with the sum of the RPE during ten simulated infusion trials, averaged over 6 or 5 rats (in pink).

We further reasoned that the original 10 min two-bottle intake test used by Fortin and Roitman [17] may not give enough time for the acutely sodium-depleted rats to reach sodium satiety (or in terms of our model—to reach the optimal settling point of its internal state that corresponds to the minimum of the drive function). In order to study rat behavior when their sodium internal state approaches and reaches the optimal balance, we simulated the two-bottle intake test over a period that was sufficiently long for the agent to reach the satiation state. Figure 4A shows simulations for the evolution of behavior and DA dynamics when the agents approach satiety and become replete in sodium. Figure 4A(a,e) show the evolution of the probability for a simulated individual with optimized free parameters to drink either NaCl, KCl or nothing (a), and the evolution of the probability to drink either NaCl, LiCl or nothing (e). Interestingly, in both simulated conditions, the reduction in the consumption of NaCl or LiCl appears to anticipate satiety: the probability to drink starts to decrease before the amount of NaCl ingested (shown in Figure 4A(b,f)) makes the individual reach satiety. We then tracked the RPE particularly when agents approach the replete state. In this case, consuming NaCl deviates the internal state away (beyond) from the setpoint. This deviation translates into a negative RPE. In our model, assuming that the baseline of DA outflow is sufficiently low, this corresponds to no DA being released in the NAc (should the baseline be high, the negative RPE would be interpreted as a phasic decrease in DA outflow). This is consistent with the fact that in sodium-repleted animals, intraoral infusions of hypertonic sodium chloride evoke no DA signal in the NAc: the DA concentration does not change from the baseline [16,17]. We also note the tell-tale oscillatory pattern in the RPE/DA over trials for the replete condition. This may be explained as follows: a small amount of sodium is lost between each trial; therefore, after having reached a replete state for the first time, as the agents have learned not to consume NaCl, the internal state can move below the setpoint again. When this happens, NaCl becomes rewarding again, which corresponds to the positive peaks of the RPE. The dynamics of RPE and the sodium-consumptive choice frequency are then determined by the physiological processes that regulate sodium loss.
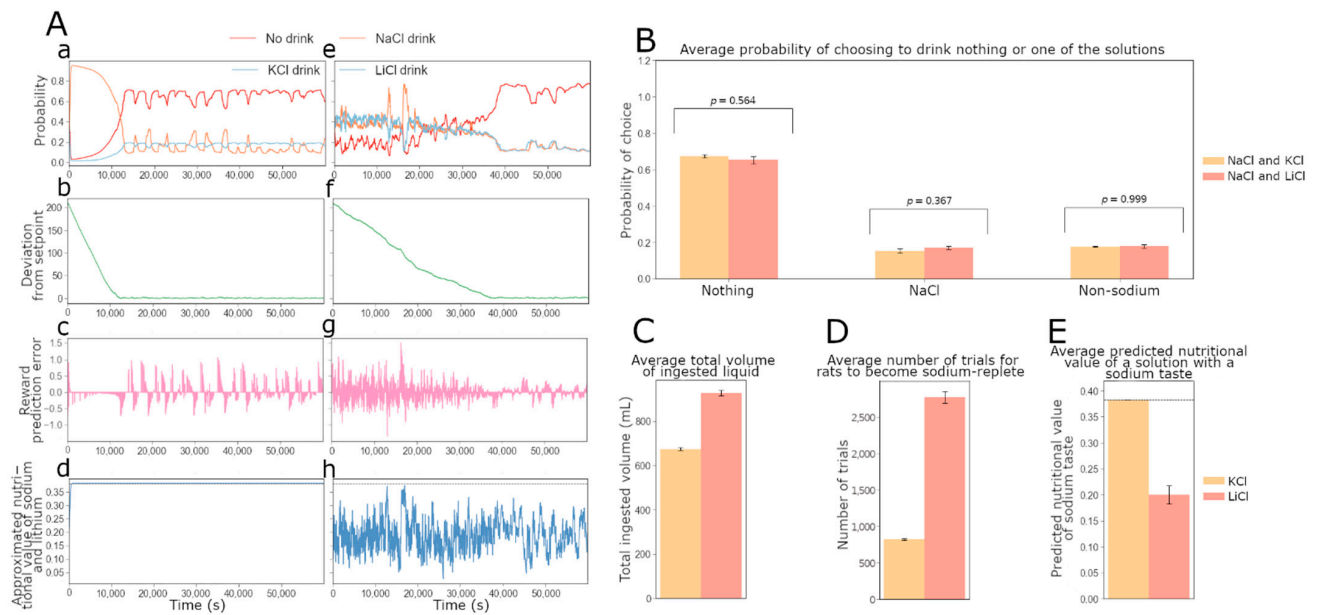
Figure 4. Model predictions for behavior during approach to sodium satiety. Here we produced example simulated data for two rats taking the two-bottle intake test developed by Fortin and Roitman (2018) for an extended period of time. In one version of the experiment, sodium-depleted agents have access to NaCl and KCl, and in the other version, NaCl and LiCl are the solutions available. Both versions of the experiment last 60,000 s (4000 trials). For panels in (**B**): 14 agents perform the version with KCl and 15 others perform the version with LiCl. The free parameter values for the agents in this group (N = 28) are all different and are sampled around the optimized ones used for the example data shown in (**A**). (**A**(**a**,**e**)) Evolution of the probability for an optimized agent to drink NaCl, KCl or nothing over time (left). Evolution of the probability for the same agent to drink NaCl, LiCl or nothing over time (right). (**A**(**b**,**f**)) Deviation from the optimal sodium state. The optimal level of internal sodium for survival is set at 0. (**A**(**c**,**g**)) Reward prediction error signals. (**A**(**d**,**h**)) Evolution of the approximated nutritional value ($\hat{K}_t$) for sodium (**d**) and lithium (**g**). (**B**) Average probabilities for agents (N = 28) to drink NaCl, the available non-sodium solution (KCl or LiCl) or nothing at the end of the 1000 min experiment. Probabilities for agents that have access to NaCl and KCl (N = 14) are shown in yellow, and probabilities for access to NaCl and LiCl (N = 15) are shown in red. (**C**) Average volumes of ingested liquid by agents (N = 28) during the 1000 min simulated experiments. The intake volume for agents that have access to NaCl and KCl (N = 14) is shown in yellow, and the intake volume for access to NaCl and LiCl (N = 15) is shown in red ($p = 1.39 \times 10^{-11}$). (**D**) Average number of trials to reach the setpoint for sodium for the first time during the simulated experiments. The number of trials for agents that have access to NaCl and KCl (N = 14) is shown in yellow, and the number of trials for access to NaCl and LiCl (N = 15) is shown in red ($p = 1.27 \times 10^{-11}$). (**E**), Average predicted nutritional value of sodium or lithium associated to their taste at the end of the 1000 min simulated experiments. This value for access to NaCl and KCl (N = 14) is shown in yellow, and for rats with access to NaCl and LiCl (N = 15) it is shown in red ($p = 4.45 \times 10^{-7}$).

When there is access to lithium chloride in the replete state, consuming this solution does not change the sodium internal state, but it can still be a punishment, similar to drinking sodium, since the two solutions acquire the same predictive value due to taste similarity. As in the 10 min experiment, Figure 4A(d,h) indicates that the value of the predicted nutritional value of sodium associated to its taste, $\hat{K}_t$, is learned when rats have access to NaCl and KCl. On the other hand, when rats consume both NaCl and LiCl, the long-term value of $\hat{K}_t$ oscillates around the value learned in the first 10 min of the experiment (equal to half the real nutritional value of one lick of NaCl).

In Figure 4, the right panels show the average results for this simulated experiment, which could be tested with new experiments. First, Figure 4B confirms the qualitative observation from Figure 4A that in the replete state, the probability to drink nothing

is much higher than the probability to drink one of the two solutions (NaCl or LiCl). Interestingly, agents with access to NaCl and LiCl (N = 15) appear to be slightly more likely to consume NaCl than agents in the other condition (N = 14). This is likely because the agent's estimation of the nutritive value of sodium (under LiCl-NaCl access) is lower than under NaCl-KCl access (Panel 4e). Thus, consuming NaCl in a replete state under LiCl-NaCl access is a weaker punishment than for agents with access to NaCl and KCl. We are also interested in the amount of liquid agents consume during the experiment, in each condition (Figure 4C). It appears that more liquid is consumed when NaCl and LiCl are the solutions available. This can be linked to the fact that more trials are necessary for the agent to reach satiety in this condition (Figure 4D). Finally, in Figure 4E, the agents appear to learn the true value of sodium associated to its taste when NaCl and KCl are available. However, their estimate is equal to approximately half of the true value (indicated by a dotted line) when NaCl and LiCl are available.

Evidence suggests that gastric distension is an early inhibitory signal of water ingestion in thirsty rats (Hoffmann et al., 2006). Hoffmann et al. show that dehydrated rats will almost continuously drink water or saline, yet stop drinking after 5 to 8 min before reaching satiety. The oropharynx is also likely to be involved in the anticipatory control of drinking behavior, by signaling the amount of water consumed (Zimmerman et al., 2017). Considering these results on thirst, we can hypothesize that animals would stop drinking before reaching sodium homeostasis because of early inhibitory signals. Arguably, an experiment where rats would consume sodium until fully replete would probably need to be conducted in several sessions to study the choices of the rats at different levels of internal sodium to avoid potential confounding due to gastric distension. This way, within a session, if rats stop drinking the solutions, we may assume it is because their internal need is satisfied and not because of sickness.

Therefore, we chose to simulate such experimental conditions by representing the extended two-bottle intake test as a series of short sessions. This different representation also allows us to zoom into the first and last minutes of the long experiments and compare depleted and repleted behaviors. Figure 5 shows example data for the first and the last sessions of each version of the experiment (with NaCl and KCl or with NaCl and LiCl). In panels a, b and c, the data are collected from a sodium-depleted rat with access to NaCl and KCl, while in panels d, e and f, the data are collected from a sodium-depleted rat with access to NaCl and LiCl. As we can see in Figure 5A, during the first session the probability of drinking NaCl increases significantly with sodium deprivation (left). During the last session, the rats are replete, so the choice not to drink anything is the most frequent one (right). In Figure 5B, the choices of the rats are represented, and in Figure 5C, we can see the reward prediction error signals evoked by the outcomes of these choices. We can observe that during the first session, KCl evokes a negative reward prediction error signal because it has an energy cost in our model. This means no DA is released in the NAc, which is consistent with Fortin and Roitman's result [17] that KCl does not evoke a DA response in the NAc. During the last session, sodium chloride triggers a weak negative reward prediction error as the rat is replete in sodium. This corresponds to the absence of DA release in the NAc observed by Fortin and Roitman in sodium-repleted rats receiving intraoral NaCl infusions [17]. Interestingly, when we simulate an experiment where NaCl and LiCl are available, the probabilities of drinking NaCl and LiCl remain similar throughout the first and last sessions (Figure 5D). We can see the actual simulated choices in Figure 5E. We then can also track the reward prediction error signals evoked by the outcomes of these choices (see Figure 5F). We note that during the first session, LiCl and NaCl both evoke positive reward prediction error signals, consistent with Fortin and Roitman's results [17]: intraoral infusions of NaCl and LiCl both trigger DA signals in the NAcs of sodium-depleted rats. Drinking LiCl makes the predicted nutritive value associated with the taste of sodium (and lithium) decrease, which is why after several successive licks of LiCl, the reward prediction error is negative in response to LiCl and NaCl. Drinking NaCl increases the predicted value of the sodium taste, making the taste of sodium more rewarding. Licks of NaCl thus increase the reward prediction error signal. During the

last session, the rats reach the setpoint, so they drink sodium much less often. However, since rats lose some sodium after every trial, their internal sodium level can fall below the setpoint. In that case, NaCl and LiCl evoke a positive reward prediction error signal.
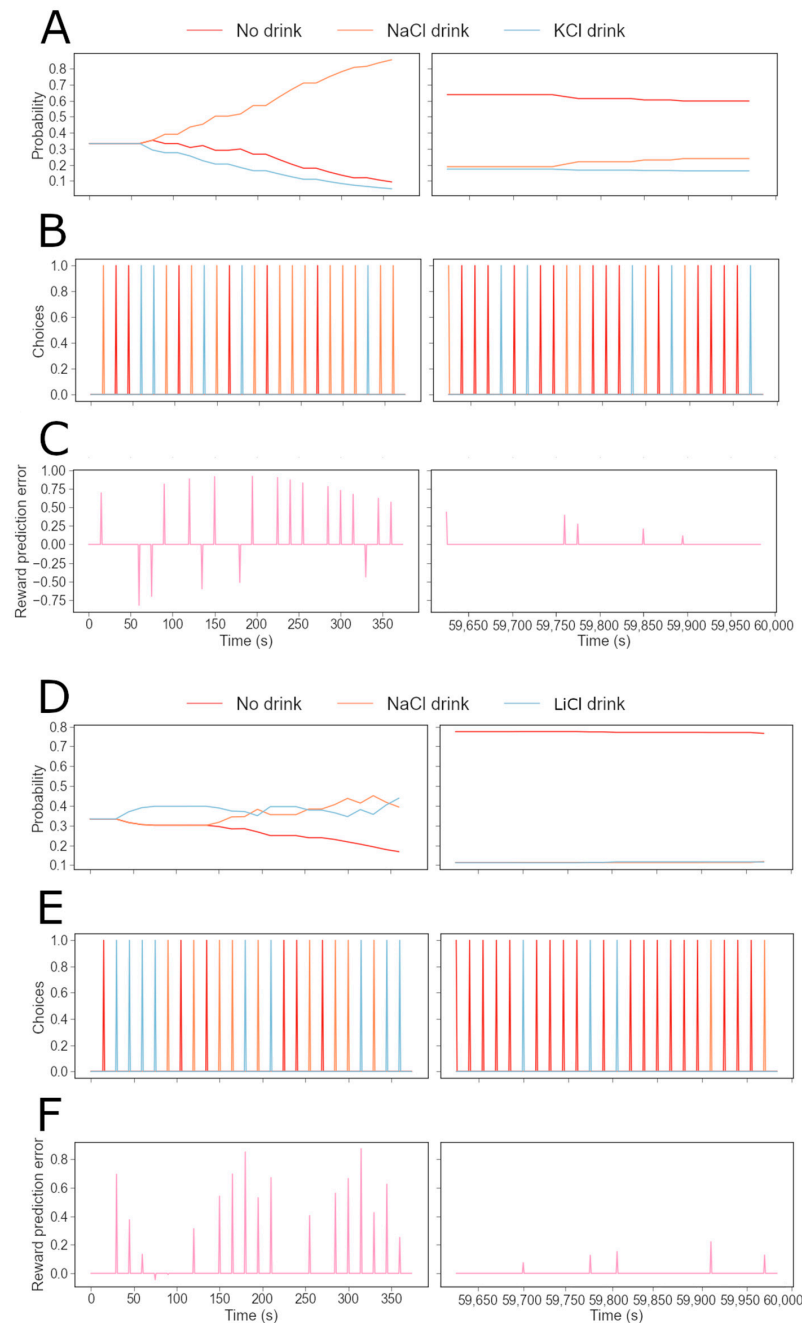

Figure 5. Example data for two simulated rats taking the two-bottle intake test developed by Fortin and Roitman (2018) during an extended period of time, divided into several sessions. Both versions of the experiment last 1000 min (4000 trials), divided into 200 sessions of 5 min (20 trials). (**A**) Evolution of the probability for an individual simulated rat with optimized free parameters to drink NaCl, KCl or nothing over time during the first (left) or last session (right). (**B**) Choices between drinking NaCl, KCl and nothing over time during the first (left) or last session (right). (**C**) Reward prediction error signal over time during the first (left) or last session (right). (**D**) Evolution of the probability to drink NaCl, LiCl or nothing over time during the first (left) or last session (right). (**E**) choices between drinking NaCl, LiCl and nothing over time during the first (left) or last session (right). (**F**) Reward prediction error signal over time during the first (left) or last session (right).

Finally, we wondered what our model would predict for the drinking behavior of rats in the presence of only KCl or LiCl. In particular, would sodium-depleted rats keep drinking LiCl, which appears to have the same taste as sodium, to try to satisfy their need? Conversely, would these rats eventually learn that LiCl has no impact on their sodium balance? We used the model to simulate the condition where only KCl or LiCl is available to sodium-depleted rats. Before this experiment, these two rats were trained while sodium-depleted as in the experiments described above: the rat with access to KCl could choose from NaCl and KCl and learned not to drink either as it became replete in sodium. The rat with access to LiCl could choose between NaCl and LiCl and also learned not to drink either once it became replete in sodium.

In order to see clearer what happens in the agent behavior and the corresponding agent internal variable (i.e., the RPE), we show examples of individual behavior and DA dynamics in Figure 6F,G. Figure 6F describes the experiment in which only KCl is available. The individual choices are shown in the middle panels. The reward prediction error dynamics are represented in Figure 6, bottom panels. As we can see in the top panel, the individual agent's probability to consume KCl smoothly decreases (null behavior probability decreases) because of the energy cost it imposes without any internal state benefit. Due to this cost, the KCl choice leads to negative RPE signals, which would correspond experimentally to an absence of DA release in the NAc. By the last session, the rat is much more likely to drink nothing (with a probability of about 0.8) than to drink KCl (with a probability of about 0.2). The probability of engaging with KCl is not null even after extended exposure because it represents the probability of making an exploratory decision. During the last session, sodium chloride triggers a weak negative reward prediction error in the model as the agent is replete in sodium. This may correspond to the absence of DA release in the NAc observed by Fortin and Roitman in sodium-repleted rats (2018). Panels Figure 6D–F describe the experiment in which only LiCl is available. In Figure 6G, on the left, we show the evolution of the probability to drink LiCl or nothing over time during the first session. On the right: the probability during the last session is shown. We can see the evolution of the relevant probabilities. We first notice that the probability of drinking LiCl is higher than that of drinking nothing only during the first session and a part of the second one. At the end of the last session, the difference in probability between drinking LiCl and nothing is similar to the one between drinking KCl and nothing, during the last session. The middle panels show the individual choices made. The bottom panels show the reward prediction error dynamics. It appears that, during the first session, the rat learns that LiCl does not increase its internal sodium level and imposes an energy cost. The reward prediction error signals in response to LiCl are first positive and become negative as the rat learns the value of lithium.

Figure 6 gives example data, represented over the entire duration of the experiments to show evolutions more clearly. It also gives average data that can be tested with future experiments. Figure 6A shows the evolution of the probability for the rat to drink KCl or nothing over time (left), and the evolution of the probability for the rat to drink LiCl or nothing over time (right). Figure 6B represents the reward prediction error signal of the same rats over time. This different representation gives us a global view of the evolution of the behavior of these individual rats throughout the whole experiment. Figure 6C shows the average predicted value of a lick of sodium or lithium based on its taste for rats at the end of the 25 min experiment. The true nutritive value of one lick of sodium is indicated with a dotted line. This estimate is equal to half of the true value when KCl is available because it does not change in the absence of exposure to NaCl or LiCl. When LiCl is available, this value is close to zero ($p = 2.02 \times 10^{-30}$). Figure 6E confirms that at the end of the experiment, rats are on average just as unlikely to drink LiCl (N = 15) as they are to drink KCl (N = 14), which does not taste like NaCl.
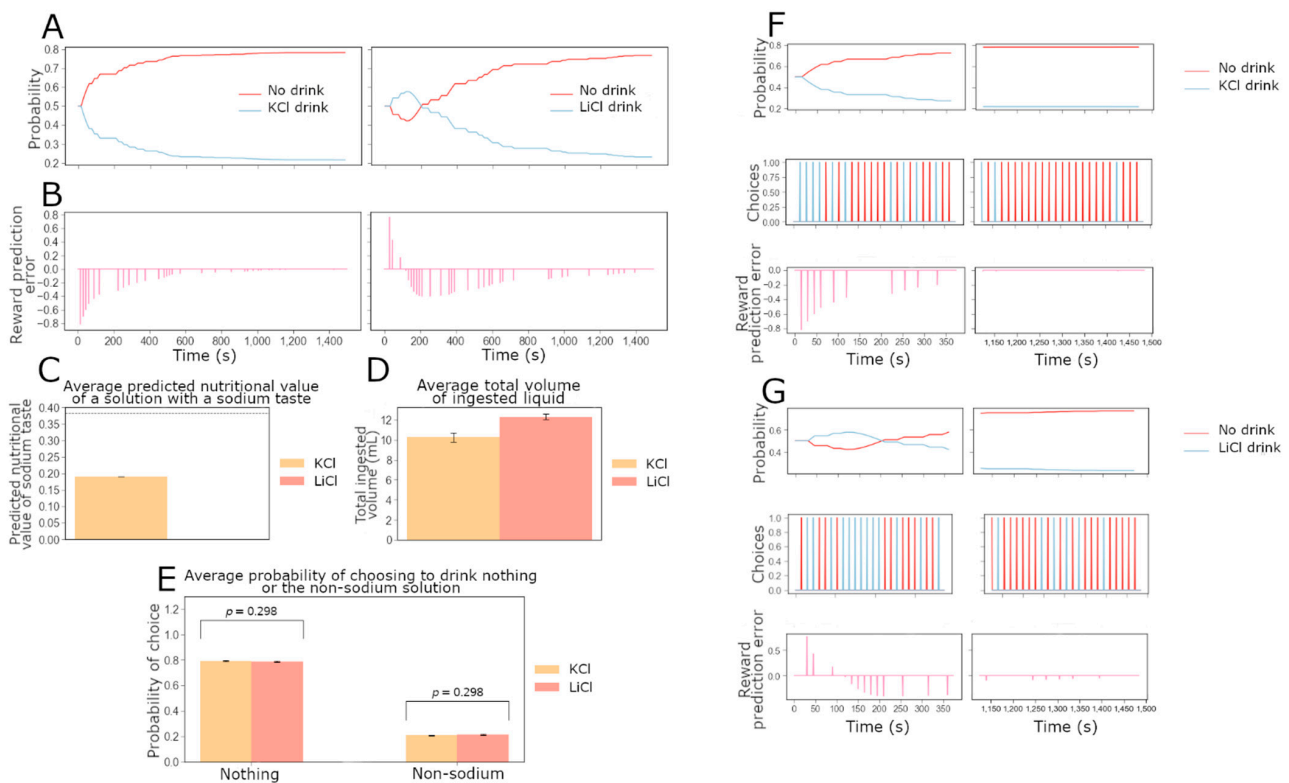
Figure 6. Model predictions on sodium-depleted rat behavior and dopamine dynamics when only non-sodium salt is available. Here we produce example simulated data for an experiment in which two sodium-depleted agents have access to only one salt solution: KCl only and LiCl only. Both versions of the experiment last 25 min (100 trials), divided into 5 sessions of 5 min (20 trials). (**A**–**E**) give example data for an optimized agent and average data for the experiment in which sodium-depleted agents have access to only one salt solution. For average data, 14 agents have access to KCl, while 15 agents have access to LiCl. The free parameter values for the agents in this group (N = 28) are all different and differ slightly from the optimized ones used for the example data. Both versions of the experiment last 25 min (100 trials), divided into 5 sessions of 5 min (20 trials). We show data for the entire 25 min experiment. (**A**) Evolution of the probability for an agent to drink KCl or nothing over time (left); probability to drink LiCl (right). (**B**) Reward prediction error signals for KCl (left) and LiCl (right). (**C**) Average predicted value of a lick of KCl (yellow) or LiCl (orange) based on the taste at the end of the 25 min experiment. The true nutritive value of one lick of sodium is indicated with a dotted line ($p = 2.02 \times 10^{-30}$). (**D**) Average volumes of ingested liquid during the 25 min simulated experiments: KCl (N = 14) is shown in yellow and LiCl (N = 15) is shown in red ($p = 0.001$). (**E**) Average probabilities to drink the available non-sodium solution (KCl in yellow or LiCl in orange). ((**F**) **top panels**): Example evolution of the probability of an agent to drink KCl or nothing over time during the first session (left) and the last session (right). ((**F**) **middle panels**): Choices between drinking KCl and nothing over time during the first (left) or last session (right). ((**F**) **bottom panels**): Reward prediction signal over time during the first (left) or last session (right). ((**G**) **top**): Evolution of the probability for a simulated rat with optimized free parameters to drink LiCl or nothing over time during the first (left) or the last session (right). ((**G**) **middle**): Choices between drinking LiCl and nothing over time during the first (left) or last session (right). ((**G**) **bottom**): Reward prediction error signal over time during the first (left) or last session (right).

These average data allow us to propose an explanation for the individual behavior in Figure 6A,B. The agent in panel a has already learned it is not worth it to drink KCl (the action value is initialized with the value the rat previously learned to satisfy its need in sodium). The probability of consuming potassium thus quickly drops to about 0.2. The agent in Figure 6B has learned not to drink LiCl as it has kept its associated action value

learned while sodium replete. This agent discovers through an exploratory decision that lithium is rewarding because of the previously learned $\hat{K}_t$ value, but lithium does not increase the internal state, so $\hat{K}_t$ decreases to 0. The reward prediction error signal therefore becomes negative in our model when the agent consumes LiCl, which has a cost but no homeostatic reward. Thus, the probability of consuming lithium quickly decreases to about 0.2. In Figure 6D, we chose to represent the total liquid consumed during each version of the experiment. It appears that rats drink more liquid when they have access to LiCl than when KCl is the available solution ($p = 0.001$).

4. Discussion

In this contribution, we show how sodium-directed behaviors and DA responses evoked by sodium and non-sodium stimuli following the induction of sodium appetite, can be accurately described by the homeostatic reinforcement (HRRL) theoretical framework. To do so, we optimized a simple HRRL agent to learn about the consequences of consuming different sodium (NaCl)- and non-sodium (KCl, LiCl)-containing solutions across various states of sodium balance. We show that the behavior of the HRRL model agents closely reproduced experimental data in which sodium-depleted rats were given access to NaCl, KCl or LiCl. Much like the rats, following the induction of sodium appetite, our HRRL agents strongly preferred solutions that contained sodium or lithium over potassium. We then examined the reward prediction error signals on single trials for the HRRL agents in response to consumption of salt-containing solutions. In our HRRL model, the RPE signal should correspond with phasic DA release in the NAc during sodium appetite. We show that the HRRL RPE is strongly modulated by internal state and the properties of the salt stimulus, which aligns well with experimentally observed DA signals in the NAc of sodium-depleted rats that are exposed to NaCl, LiCl and KCl. As in rats with sodium appetite, the HRRL RPE signals were positive for sodium tasting salts (NaCl and LiCl) under deplete conditions and attenuated as the agent approached a sodium-repleted state. Furthermore, the HRRL RPE for KCl was negligible, which also matched the experimentally observed phasic DA release.

We then used the HRRL framework to model sodium-seeking behavior in prolonged salt discrimination experiments, where sodium-depleted rats are allowed to consume sodium to satiety. Interestingly, our simulated agents began to reduce their consumption of NaCl before becoming sated. This suggested that the agent behavior reflects the anticipation of a future replete state and makes this as a prediction for the experiments. We further observed that after extensive experience with salt solutions while sodium depleted, the HRRL agent that had access to sodium and potassium chloride could predict the nutritive value of a lick of NaCl based on its taste. However, the estimated value of NaCl consumption was halved when the HRRL agents had access to both sodium and lithium chloride, which have been shown to taste similarly. We finally investigated the drinking behavior when only lithium chloride was available. Recall that lithium chloride tastes similar to sodium chloride but cannot restore the lost sodium. Our simulations predict that sodium-deprived rats learn that lithium chloride does not fulfill their sodium deficiency and thus should stop consuming it. Our simulation results can be tested with experiments that could give further insight into homeostatic state regulation, goal-directed behavior, reinforcement learning and how these phenomena depend on mesolimbic DA signaling.

In our HRRL framework, we modeled the state of sodium balance through a drive function that represents the deviation between the agents' current state and an idealized homeostatic setpoint. In the model, the source of this sodium drive is ambiguous. However, in biological systems, sodium need is sensed via distributed neural systems. There are aldosterone-sensitive neurons in the nucleus of the solitary tract (NTS) that express 11β-hydroxysteroid dehydrogenase type 2 (HSD2+) that are activated by sodium depletion [35]. Activation of NTS HSD2+ neurons can drive sodium consumption in sodium-repleted mice, whereas inhibition reduces sodium consumption in depleted animals [36]. Importantly, NTS HSD2+ neurons project to other hindbrain areas such as the parabrachial complex (PB) and

pre-locus coeruleus (pre-LC), which contribute to sodium appetite in distinct ways [37,38]. In addition to the NTS, sodium need is also sensed by neurons in the subfornical organ (SFO), which respond to changes in blood osmolality, among other things [39], and has been shown to regulate sodium appetite [40]. Notably, NTS HSD2+ neurons and SFO neurons that respond to osmolality challenges may influence sodium appetite through synergistic influences on neurons in the Bed Nucleus of the Stria Terminalis [41]. Thus, in biological systems, state sensing, even for a single nutrient, is subject to complex multi-pathway regulation. This is to say nothing of how multiple competing needs interaction to influence behavior. In our HRRL models, the drive function was one-dimensional, as the current state of sodium balance was the only input under consideration.

In addition to sensing challenges to sodium balance to tune drive states, another key aspect of sodium appetite is the ability to detect sources of sodium in the environment. In the HRRL model, a "tastant" was evaluated by comparing its estimated reward value with the current state of sodium balance. If the agent was below an ideal setpoint, the tastant was evaluated as positively reinforcing. Otherwise the tastant was evaluated negatively as it further exacerbated deviations from homeostasis and cost energy to approach and consume. In vivo, sensing sodium primarily begins via Na+ selective epithelial sodium channels (eNACs), which are critical for sodium detection and discrimination [42–44]. Sodium taste information is relayed to the brain via the chorda tympani (CT; [45,46]) and CT transection disrupts the expression of sodium appetite [47,48]. Interestingly, CT responses to sodium solutions are augmented by sodium appetite [25], suggesting that changes in sodium balance can exert widespread effects on taste processing, even outside the CNS.

In the CNS, sodium appetite alters sodium taste responses in numerous brain areas, including hindbrain structures such as the NTS [49,50] and PB [51] as well as circuits linked to reinforcement learning such as the striatum [52] and the mesolimbic DA system [16,17]. Such widespread changes in sodium taste processing likely contribute to the robust changes in the consummatory responses and affective orosensory evaluation of sodium-containing solutions typically observed in sodium appetite [23,53].

Linking the gustatory properties of sodium-containing stimuli (taste, phasic) with sodium need (state, tonic) is essential for sodium appetite. Changes in the physiological state (tonic) serve to bias the organism's behavioral repertoire towards sodium-seeking behaviors such that sodium intake is invigorated when tastants that contain sodium are identified. For example, consummatory responses such as bursts of licking, which are thought to reflect the palatability of a taste stimulus [54], increase following sodium depletion [53]. Prior work suggests that taste sensing and state interact in an additive fashion [53], but if either is disrupted, the expression of sodium appetite is impaired [36,55].

Perhaps the key feature of the HRRL framework is that by incorporating a drive function that tracks deviations from homeostasis into a traditional RL-based agent, this enabled us to model the taste–state interactions that drive both current and future sodium-seeking behaviors associated with sodium appetite. At the moment of consumption (e.g., when an RPE is triggered), it is likely that state-dependent changes in taste information are relayed to DA neurons via the PB and pre-LC, both of which innervate the VTA [38]. PB neurons express cFos following sodium consumption in sodium-depleted rats and this corresponds with reduced cFos staining in NTS HSD2+ neurons [5]. Moreover, pre-LC FoxP2+ neurons that receive efferents from the NTS are strongly implicated in sodium appetite [56] and project to the VTA [17]. By tuning RPEs according to the current state of the animal, the HRRL model may mimic PB and pre-LC influences on VTA DA signaling that track physiological need.

Future work could explore whether the HRRL model also captures longer term plasticity in reinforcement circuits associated with sodium appetite. For example, sodium appetite is augmented by a single prior experience with sodium depletion [57], while repeated depletions increase the need for free sodium intake long after sodium levels are replenished [58,59]. In terms of RPE signaling, mesolimbic DA neurons appear to encode sodium cues as reward predictive only after extensive experience with cue–sodium pairings while

sodium depleted [16]. However, after such associations are learned, phasic DA release to the sodium cue is only observed when the animals are in a state of need [16]. It is likely the HRRL model would already account for this in its underlying math. Another challenging issue is how the proposed drive function is encoded in neural activity and how multiple drives may be prioritized with respect to each other (see also the discussion on multiple constraint satisfaction above).

An interesting point to consider in light of our results is whether the taste can be considered as another conditioned stimulus (CS) whose predicted impact of the consumed unconditioned stimulus (US) on the internal state is learned. A study on DA in mice consuming sucrose draws conclusions supporting this hypothesis [32]. According to this study, a taste evokes a DA response only if the animal has associated through prior learning this taste with a nutritional value. In other words, the hedonic property of taste cannot alone be responsible for reinforcement. The study suggests that taste acts as a conditioned stimulus predictive of a food reward. This implies that the association between a taste and a nutritional value can be extinguished by giving the animal extensive experience with the taste in the absence of a value (an artificial sweetener instead of sucrose, for example). When it comes to sodium, however, taste does not seem to act simply as a conditioned stimulus and appetite appears to be regulated by a more complex interplay of both innate and learned mechanisms. This is reflected in our model by the non-linear drive function and the non-zero initial value of $\hat{K}_t$. However, in our simulations, it is possible to extinguish seeking behavior towards the sodium taste after the exposure of sodium-depleted rats to LiCl. This suggests that the subjective value attached to sodium taste can be modified with learning, which should be investigated with new experiments.

Intriguingly, a recent proposal suggests that gustatory information could act as a hedonic predictor of the long-run worth of a good [60]. This suggestion appears at first glance to be compatible with our work; for example, just as with the gustatory "hedonic" account in our simulation, the gustatory "reward/value" rapidly leads to the motivation of consumption behavior and is ultimately dominated by the nutritive lack of impact for LiCl with the end value converging to zero.

Our model, as all computational models, faces several limitations: the actions are necessarily chosen and performed by the agent at discrete and regular time steps. Moreover, the solutions are consumed at a fixed volume. A continuous time model, in which the internal state of the agent has to be constantly regulated with actions that maintain its homeostasis, would be closer to reality. Indeed, the temporality of physiological regulation is important as the deviation from the setpoint must be reduced as soon as possible. We have previously shown that taking the shortest path to reach the setpoint optimizes fitness and that HRRL agents learn to preemptively avoid life-threatening excursions far from the homeostatic optima [61].

The proposed framework remains quite an abstract algorithmic model: how it relates to the neural circuitry mechanistically remains to be studied. For example, as discussed above, the circuits that mediate taste–state interactions in the brain are multi-faceted and could be communicated with the DA system through many convergent pathways. Another challenge that remains is how one might observe the drive function? A hint of predictive internal state encoding has been shown experimentally (see discussion above and [4]); however, it is not clear if such neural traces represent the internal state itself or the drive function necessary for learning and generating the behaviors. In fact, perhaps the drive function properties could be a key to individual differences in homeostatically motivated learned behaviors. Thus, understanding how one may measure the properties of the drive function from behavioral observations remains a key challenge.

In conclusion, we showed that a homeostatic reinforcement learning theory can account for behaviors motivated by sodium appetite. A recent study also put forth a similar argument [62]—focusing on a different dataset, they showed that HRRL agents can match two-bottle test behavior for sodium and water choices. Here, we critically extend these ideas to show that the HRRL framework can also quantitatively account for the taste- and

state-dependent DA responses, arguing that such DA signaling may in fact be causal for the learning of such appetitive behaviors.

Author Contributions: Experimental data curation, J.J.C. and M.F.R.; investigation, formal analysis, A.D. and C.B.; supervision, B.G. and M.F.R.; writing—original draft, A.D., M.F.R., J.J.C. and B.G.; writing—review and editing, B.G., M.F.R., J.J.C. and A.D. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The model code and the generated numerical data are available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Free parameters of the model and their values.

Parameter	Value	Explanation
ϵ	0.1446	Learning rate
β	1.5896	Rate of exploration
m	3	Root of the drive function
n	4	Power of the drive function
K	0.3819	Outcome of an action and its effect on the internal sodium level
cost	0.8180	Energy cost of going to drink from a sipper tube
loss	0.0559	Quantity of sodium metabolized by rats at every trial
setpoint H^*	211.7066	Optimal sodium internal level

References

1. Rescorla, R.A.; Wagner, A.R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory*; Black, A.H., Prokasy, W.F., Eds.; Appleton Century Crofts: New York, NY, USA, 1972; pp. 64–99.
2. Tian, J.; Huang, R.; Cohen, J.Y.; Osakada, F.; Kobak, D.; Machens, C.K.; Callaway, E.M.; Uchida, N.; Watabe-Uchida, M. Distributed and Mixed Information in Monosynaptic Inputs to Dopamine Neurons. *Neuron* **2016**, *91*, 1374–1389. [CrossRef] [PubMed]
3. Schultz, W. Multiple reward signals in the brain. *Nat. Rev. Neurosci.* **2000**, *1*, 199–207. [CrossRef] [PubMed]
4. Livneh, Y.; Ramesh, R.N.; Burgess, C.R.; Levandowski, K.M.; Madara, J.C.; Fenselau, H.; Goldey, G.J.; Diaz, V.E.; Jikomes, N.; Resch, J.M.; et al. Homeostatic circuits selectively gate food cue responses in insular cortex. *Nature* **2017**, *546*, 611–616. [CrossRef] [PubMed]
5. Geerling, J.C.; Loewy, A.D. Central regulation of sodium appetite. *Exp. Physiol.* **2008**, *93*, 177–209. [CrossRef]
6. Augustine, V.; Lee, S.; Oka, Y. Neural Control and Modulation of Thirst, Sodium Appetite, and Hunger. *Cell* **2020**, *180*, 25–32. [CrossRef]
7. Schultz, W. Predictive Reward Signal of Dopamine Neurons. *J. Neurophysiol.* **1998**, *80*, 1–27. [CrossRef]
8. Keiflin, R.; Janak, P.H. Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. *Neuron* **2015**, *88*, 247–263. [CrossRef]
9. Dabney, W.; Kurth-Nelson, Z.; Uchida, N.; Starkweather, C.K.; Hassabis, D.; Munos, R.; Botvinick, M. A distributional code for value in dopamine-based reinforcement learning. *Nature* **2020**, *577*, 671–675. [CrossRef]
10. Kutlu, M.G.; Zachry, J.E.; Melugin, P.R.; Cajigas, S.A.; Chevee, M.F.; Kelley, S.J.; Kutlu, B.; Tian, L.; Siciliano, C.A.; Calipari, E.S. Dopamine release in the nucleus accumbens core signals perceived saliency. *Curr. Biol.* **2021**, *31*, 4748–4761.e8. [CrossRef]
11. Liu, S.; Borgland, S. Regulation of the mesolimbic dopamine circuit by feeding peptides. *Neuroscience* **2015**, *289*, 19–42. [CrossRef]
12. Cone, J.J.; Roitman, J.D.; Roitman, M.F. Ghrelin regulates phasic dopamine and nucleus accumbens signaling evoked by food-predictive stimuli. *J. Neurochem.* **2015**, *133*, 844–856. [CrossRef] [PubMed]
13. Mietlicki-Baase, E.G.; Reiner, D.J.; Cone, J.J.; Olivos, D.R.; McGrath, L.E.; Zimmer, D.J.; Roitman, M.F.; Hayes, M.R. Amylin Modulates the Mesolimbic Dopamine System to Control Energy Balance. *Neuropsychopharmacology* **2014**, *40*, 372–385. [CrossRef] [PubMed]
14. Mebel, D.M.; Wong, J.C.Y.; Dong, Y.J.; Borgland, S.L. Insulin in the ventral tegmental area reduces hedonic feeding and suppresses dopamine concentration via increased reuptake. *Eur. J. Neurosci.* **2012**, *36*, 2336–2346. [CrossRef] [PubMed]
15. Cone, J.; McCutcheon, J.; Roitman, M.F. Ghrelin Acts as an Interface between Physiological State and Phasic Dopamine Signaling. *J. Neurosci.* **2014**, *34*, 4905–4913. [CrossRef]

16. Cone, J.J.; Fortin, S.M.; McHenry, J.A.; Stuber, G.D.; McCutcheon, J.E.; Roitman, M.F. Physiological state gates acquisition and expression of mesolimbic reward prediction signals. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1943–1948. [CrossRef]
17. Fortin, S.M.; Roitman, M.F. Challenges to Body Fluid Homeostasis Differentially Recruit Phasic Dopamine Signaling in a Taste-Selective Manner. *J. Neurosci.* **2018**, *38*, 6841–6853. [CrossRef]
18. Hsu, T.M.; Bazzino, P.; Hurh, S.J.; Konanur, V.R.; Roitman, J.D.; Roitman, M.F. Thirst recruits phasic dopamine signaling through subfornical organ neurons. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 30744–30754. [CrossRef]
19. Grove, J.C.R.; Gray, L.A.; Medina, N.L.S.; Sivakumar, N.; Ahn, J.S.; Corpuz, T.V.; Berke, J.D.; Kreitzer, A.C.; Knight, Z.A. Dopamine subsystems that track internal states. *Nature* **2022**, *608*, 374–380. [CrossRef]
20. Reichenbach, A.; Clarke, R.E.; Stark, R.; Lockie, S.H.; Mequinion, M.; Dempsey, H.; Rawlinson, S.; Reed, F.; Sepehrizadeh, T.; DeVeer, M.; et al. Metabolic sensing in AgRP neurons integrates homeostatic state with dopamine signalling in the striatum. *eLife* **2022**, *11*, e72668. [CrossRef]
21. Denton, D. *The Hunger for Salt: An Anthropological, Physiological, and Medical Analysis Berlin*; Springer: New York, NY, USA, 1982; ISBN 0387112863.
22. Richter, C.P. Increased Salt Appetite in Adrenalectomized Rats. *Am. J. Physiol. -Leg. Content* **1936**, *115*, 155–161. [CrossRef]
23. Berridge, K.C.; Flynn, F.W.; Schulkin, J.; Grill, H.J. Sodium depletion enhances salt palatability in rats. *Behav. Neurosci.* **1984**, *98*, 652–660. [CrossRef]
24. Nachman, M. Learned aversion to the taste of lithium chloride and generalization to other salts. *J. Comp. Physiol. Psychol.* **1963**, *56*, 343–349. [CrossRef] [PubMed]
25. Contreras, R.J.; Frank, M. Sodium deprivation alters neural responses to gustatory stimuli. *J. Gen. Physiol.* **1979**, *73*, 569–594. [CrossRef] [PubMed]
26. Mehdi Keramati Boris Gutkin 2014 Homeostatic reinforcement learning for integrating reward collection and physiological stability. *eLife* **1943**, *3*, e04811.
27. Hulme, O.; Melville, T.; Gutkin, B. Neurocomputational Theories of Homeostatic Control. *Phys. Life Rev.* **2019**, *31*, 214–232. [CrossRef] [PubMed]
28. Hull, C.L. *Principles of Behavior: An Introduction to Behavior Theory*; Appleton-Century-Crofts: New York, NY, USA, 1943.
29. Keramati, M.; Durand, A.; Girardeau, P.; Gutkin, B.; Ahmed, S.H. Cocaine addiction as a homeostatic reinforcement learning disorder. *Psychol. Rev.* **2017**, *124*, 130–153. [CrossRef] [PubMed]
30. Gonçalves, P.J.; Lueckmann, J.-M.; Deistler, M.; Nonnenmacher, M.; Öcal, K.; Bassetto, G.; Chintaluri, C.; Podlaski, W.F.; Haddad, S.A.; Vogels, T.P.; et al. Training deep neural density estimators to identify mechanistic models of neural dynamics. *eLife* **2020**, *9*, e56261. [CrossRef] [PubMed]
31. Schultz, W. Dopamine reward prediction error coding. *Dialog. Clin. Neurosci.* **2016**, *18*, 23–32. [CrossRef] [PubMed]
32. Beeler, J.A.; McCutcheon, J.E.; Cao, Z.F.; Murakami, M.; Alexander, E.; Roitman, M.F.; Zhuang, X. Taste uncoupled form nutrition fails to sustain the reinforcing properties of food. *Eur. J. Neurosci* **2012**, *36*, 2533–2546. [CrossRef]
33. Sutton, R.S.; Barto, A.G. *Reinfrocement Learning: An Introduction*; MIT Press: Cambridge, UK, 1998; Volume 1.
34. Schultz, W.; Dayan, P.; Montague, P.R. A neural substrate of prediction and reward. *Science* **1997**, *14*, 1593–1599. [CrossRef]
35. Geerling, J.C. Aldosterone Target Neurons in the Nucleus Tractus Solitarius Drive Sodium Appetite. *J. Neurosci.* **2006**, *26*, 411–417. [CrossRef] [PubMed]
36. Jarvie, B.C.; Palmiter, R.D. HSD2 neurons in the hindbrain drive sodium appetite. *Nat. Neurosci.* **2016**, *20*, 167–169. [CrossRef] [PubMed]
37. Geerling, J.C.; Loewy, A.D. Sodium deprivation and salt intake activate separate neuronal subpopulations in the nucleus of the solitary tract and the parabrachial complex. *J. Comp. Neurol.* **2007**, *504*, 379–403. [CrossRef] [PubMed]
38. Shin, J.-W.; Geerling, J.C.; Stein, M.K.; Miller, R.L.; Loewy, A.D. FoxP2 brainstem neurons project to sodium appetite regulatory sites. *J. Chem. Neuroanat.* **2011**, *42*, 1–23. [CrossRef]
39. Hiyama, T.; Watanabe, E.; Okado, H.; Noda, M. The Subfornical Organ is the Primary Locus of Sodium-Level Sensing by Nax Sodium Channels for the Control of Salt-Intake Behavior. *J. Neurosci.* **2004**, *24*, 9276–9281. [CrossRef]
40. Matsuda, T.; Hiyama, T.; Niimura, F.; Matsusaka, T.; Fukamizu, A.; Kobayashi, K.; Kobayashi, K.; Noda, M. Distinct neural mechanisms for the control of thirst and salt appetite in the subfornical organ. *Nat. Neurosci.* **2016**, *20*, 230–241. [CrossRef]
41. Resch, J.M.; Fenselau, H.; Madara, J.C.; Wu, C.; Campbell, J.N.; Lyubetskaya, A.; Dawes, B.A.; Tsai, L.T.; Li, M.M.; Livneh, Y.; et al. Aldosterone-Sensing Neurons in the NTS Exhibit State-Dependent Pacemaker Activity and Drive Sodium Appetite via Synergy with Angiotensin II Signaling. *Neuron* **2017**, *96*, 190–206.e7. [CrossRef]
42. Geran, L.C.; Spector, A.C. Sodium taste detectability in rats is dependent of anion size: The psychophysical characteristics of the transcellular sodium taste transduction pathway. *Behav. Neurosci.* **2000**, *114*, 1229–1238. [CrossRef]
43. Smith, K.R.; Treesukosol, Y.; Paedae, A.B.; Contreras, R.J.; Spector, A.C. Contribution of the TRPV1 channel to salt taste quality in mice as assessed by conditioned taste aversion generalization and chorda tympani nerve responses. *Am. J. Physiol. -Regul. Integr. Comp. Physiol.* **2012**, *303*, R1195–R1205. [CrossRef]
44. Treesukosol, Y.; Lyall, V.; Heck, G.L.; DeSimone, J.A.; Spector, A.C. A psychophysical and electrophysiological analysis of salt taste in Trpv1 null mice. *Am. J. Physiol. -Regul. Integr. Comp. Physiol.* **2007**, *292*, R1799–R1809. [CrossRef]
45. O'Keefe, G.B.; Schumm, J.; Smith, J.C. Loss of sensitivity to low concentrations of NaCl following bilateral chorda tympani nerve sections in rats. *Chem. Senses* **1994**, *19*, 169–184. [CrossRef] [PubMed]

46. Golden, G.J.; Ishiwatari, Y.; Theodorides, M.L.; Bachmanov, A.A. Effect of Chorda Tympani Nerve Transection on Salt Taste Perception in Mice. *Chem. Senses* **2011**, *36*, 811–819. [CrossRef] [PubMed]

47. Sollars, S.I.; Bernstein, I.L. Sodium appetite after transection of the chorda tympani nerve in Wistar and Fischer 344 rats. *Behav. Neurosci.* **1992**, *106*, 1023–1027. [CrossRef] [PubMed]

48. Breslin, P.A.; Spector, A.C.; Grill, H.J. Chorda tympani section decreases the cation specificity of depletion-induced sodium appetite in rats. *Am. J. Physiol. -Regul. Integr. Comp. Physiol.* **1993**, *264*, R319–R323. [CrossRef]

49. Jacobs, K.M.; Mark, G.P.; Scott, T.R. Taste responses in the nucleus tractus solitarius of sodium-deprived rats. *J. Physiol.* **1988**, *406*, 393–410. [CrossRef]

50. McCaughey, S.A.; Scott, T.R. Rapid induction of sodium appetite modifies taste-evoked activity in the rat nucleus of the solitary tract. *Am. J. Physiol. -Regul. Integr. Comp. Physiol.* **2000**, *279*, R1121–R1131. [CrossRef]

51. Shimura, T.; Komori, M.; Yamamoto, T. Acute sodium deficiency reduces gustatory responsiveness to NaCl in the parabrachial nucleus of rats. *Neurosci. Lett.* **1997**, *236*, 33–36. [CrossRef]

52. Loriaux, A.L.; Roitman, J.D.; Roitman, M.F. Nucleus accumbens shell, but not core, tracks motivational value of salt. *J. Neurophysiol.* **2011**, *106*, 1537–1544. [CrossRef]

53. Breslin, P.A.; Kaplan, J.M.; Spector, A.C.; Zambito, C.M.; Grill, H.J. Lick rate analysis of sodium taste-state combinations. *Am. J. Physiol. -Regul. Integr. Comp. Physiol.* **1993**, *264*, R312–R318. [CrossRef]

54. Johnson, A.W. Characterizing ingestive behavior through licking microstructure: Underlying neurobiology and its use in the study of obesity in animal models. *Int. J. Dev. Neurosci.* **2017**, *64*, 38–47. [CrossRef]

55. Bernstein, I.L.; Hennessy, C.J. Amiloride-sensitive sodium channels and expression of sodium appetite in rats. *Am. J. Physiol. Integr. Comp. Physiol.* **1987**, *253*, R371–R374. [CrossRef]

56. Lee, S.; Augustine, V.; Zhao, Y.; Ebisen, H.; Ho, B.; Okay, Y. Chemosensory modulation of neural circuits for sodium appetite. *Nature* **2019**, *568*, 93–97. [CrossRef]

57. Sakai, R.R.; Fine, W.B.; Epstein, A.N.; Frankmann, S.P. Salt appetite is enhanced by one prior episode of sodium depletion in the rat. *Behav. Neurosci.* **1987**, *101*, 724–731. [CrossRef]

58. Sakai, R.R.; Frankmann, S.P.; Fine, W.B.; Epstein, A.N. Prior episodes of sodium depletion increase the need-free sodium intake of the rat. *Behav. Neurosci.* **1989**, *103*, 186–192. [CrossRef]

59. Dietz, D.M.; Curtis, K.S.; Contreras, R.J. Taste, Salience, and Increased NaCl Ingestion after Repeated Sodium Depletions. *Chem. Senses* **2005**, *31*, 33–41. [CrossRef]

60. Dayan, P. "Liking" as an early and editable draft of long-run affective value. *PLoS Biol.* **2022**, *20*, e3001476. [CrossRef]

61. Keramati, M.; Gutkin, B.S. A Reinforcement Learning Theory for Homeostatic Regulation. In *Advances in Neural Information Systems*; Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K.Q., Eds.; Curran Associates: Red Hook, NY, USA, 2011, ISBN 9781618395993.

62. Uchida, Y.; Hikida, T.; Yamashita, Y. Computational Mechanisms of Osmoregulation: A Reinforcement Learning Model for Sodium Appetite. *Front. Neurosci.* **2022**, *16*, 857009. [CrossRef]