1 **Computational mechanisms of osmoregulation: a reinforcement** 2 **learning model for sodium appetite**

3 Short title

4 **Reinforcement learning model for sodium appetite**

5

6 **Yuuki Uchida^{1,2}, Takatoshi Hikida^{1*}, Yuichi Yamashita^{2*}**

7 ¹Laboratory for Advanced Brain Functions, Institute for Protein Research, Osaka
8 University, Osaka, Japan

9 ²Department of Information Medicine, National Center of Neurology and Psychiatry,
10 Tokyo, Japan

11

12 ***Corresponding authors**

13 Email: yamay@ncnp.go.jp (YY)

14 Email: hikida@protein.osaka-u.ac.jp (TH)

15

16

17 **Abstract**

18 Homeostatic control with oral nutrient intake is a vital complex system involving the
 19 orderly interactions between the external and internal senses, behavioral control, and
 20 reward learning. Sodium appetite is a representative system and has been intensively
 21 investigated in animal models of homeostatic systems and oral nutrient intake.
 22 However, the system-level mechanisms for regulating sodium intake behavior and
 23 homeostatic control remain unclear.

24 In the current study, we attempted to provide a mechanistic understanding of sodium
 25 appetite behavior by using a computational model, the homeostatic reinforcement
 26 learning model, in which homeostatic behaviors are interpreted as reinforcement
 27 learning processes. Through simulation experiments, we confirmed that our
 28 homeostatic reinforcement learning model successfully reproduced homeostatic
 29 behaviors by regulating sodium appetite. These behaviors include the approach and
 30 avoidance behaviors to sodium according to the internal states of individuals. In
 31 addition, based on the assumption that the sense of taste is a predictor of changes in
 32 the internal state, the homeostatic reinforcement learning model successfully
 33 reproduced the previous paradoxical observations of the intragastric infusion test,
 34 which cannot be explained by the classical drive reduction theory. Moreover, we
 35 extended the homeostatic reinforcement learning model to multi-modal data, and
 36 successfully reproduced the behavioral tests in which water and sodium appetite
 37 were mediated by each other. Finally, through an experimental simulation of chemical
 38 manipulation in a specific neural population in the brain stem, we proposed a testable
 39 hypothesis for the function of neural circuits involving sodium appetite behavior.

40 The study results support the idea that osmoregulation via sodium appetitive behavior
 41 can be understood as a reinforcement learning process and provide a mechanistic

42 explanation for the underlying neural mechanisms of sodium appetite and
43 homeostatic behavior.

44

45 **Author Summary**

46 The taste of high-concentration saltwater is rewarding during sodium depletion, while
47 it is aversive in a sodium sufficient state. This “sodium appetite” is a clear
48 manifestation of homeostasis maintenance and proper action selection in animals,
49 reflecting the internal environment. To reveal the computational mechanism of this
50 property, we applied a machine learning model, in which homeostatic stability is a
51 reward and the goal is to maximize the sum of the reward, and simulated animal
52 behavioral experiments. The results suggest that the mechanism of sodium-appetite
53 behavior is based on the machine learning model. Specifically, by replicating the
54 results of neural circuit manipulation, which controls sodium appetite, an algorithm in
55 which the function of a neural population affects sodium appetite behaviors was
56 proposed. Our results provide a fundamental computational model for a mechanism
57 by a function of a neural cell type to regulate animal behavior. More generally, this
58 study can be fundamental to understanding the computational process of decision
59 making reflecting the internal environment.

60

61

62

63 Introduction

64 Homeostatic control systems with oral nutrient intake are vital to sustain life. The
65 system is quite complex, in which orderly interactions between external and internal
66 senses, behavioral control, and reward learning are essential [1, 2]. Their failure to
67 properly develop or maintain precisely results in several disease symptoms. For
68 example, taste disorders may cause collapse of homeostasis and nutrient disorders
69 [3-5].

70

71 Sodium appetite is a representative system that has been intensively investigated in
72 animal models of the homeostatic systems that coordinate oral nutrient intake [6-11].
73 For example, at the behavioral level, it is known that preference for salty taste
74 changes depending on the internal sodium state. That is, when an animal is
75 sodium-depleted after adrenalectomy [12], administration of furosemide, or
76 low-sodium food [9, 10], it exhibits a positive sodium appetite, which means that
77 sodium intake works as a reward. On the other hand, when an animal is not deficient
78 in sodium, it will exhibit a negative sodium appetite, and sodium intake works as a
79 punishment [13, 14]. This property is related to various factors, and not only to
80 avoidance of harmful foods or the consumption of essential nutrients. Rather, it is a
81 complex phenomenon in which the reward value of a taste fluctuates, reflecting the
82 internal states of animals.

83

84 At the physiological level, multiple hormones are involved in controlling sodium
85 appetite. For instance, adrenalectomized rats, which have difficulty secreting
86 aldosterone, show increased sodium appetite [6]. In relation to this, sodium deficiency
87 increases the level of angiotensin II and stimulates the secretion of aldosterone [15].

88 Furthermore, the pacemaker-like firing of aldosterone-sensing neurons in the nucleus
89 of the solitary tract (NTS^{HSD2} neurons) has been observed in sodium-deficient
90 animals [16].

91

92 At the neural substrate level, it is known that several neuronal nuclei (e.g., limbic
93 system, pons, and basal ganglia) control sodium appetite. For example, activation of
94 the ventral tegmental area (VTA) dopaminergic neurons, which show robust
95 correlations with reward systems, decreases salt intake [17]. Dopaminergic neurons
96 in the midbrain may encode appetitive properties of sodium [18]. Some excitatory
97 neurons in the pre-locus coeruleus decrease sodium appetite [10]. Conversely, the
98 activation of neurons in the subfornical organ increases sodium appetite [9].

99

100 Despite these findings, the system-level mechanisms for regulating sodium intake
101 behavior and homeostatic controls remain unclear. The drive reduction theory [19] is
102 a classical depiction that has provided a basis for the nature of sodium appetite
103 behavior; that is, the discrepancy from the optimal state drives behavior to reduce the
104 discrepancy. However, some important biological observations cannot be explained
105 by the drive reduction theory. For example, direct infusion of sodium into the stomach,
106 not through the mouth, does not reduce sodium appetite for several minutes [2], even
107 though the internal sodium state may approach the optimal level.

108

109 To address this issue, here we propose a computational model to provide a
110 mechanistic understanding of sodium appetite, the homeostatic reinforcement
111 learning (HRL) model, according to which homeostatic behaviors are interpreted as
112 reinforcement learning processes [2, 20, 21]. In the HRL model, changes in the

internal state approaching the ideal states were regarded as rewards, while changes in the internal state moving away from the ideal state were considered as punishment. Based on this idea, the values of the optimal behavior to maintain internal states were acquired through an incremental learning process. In addition, by treating the taste modality and the actual change of internal state separately, the HRL model provides explanations about the mechanisms that integrate taste, behavior, and maintenance of inner environments.

The HRL model was originally proposed to explain homeostatic controls for body temperature, internal water state [2], and pathological mechanisms of cocaine addiction [21]. This study is the first attempt to use the HRL model to explain sodium appetite behavior. In addition, although the HRL model can handle multi-dimensional internal states, previous studies have utilized it to treat one-dimensional changes in the internal state [2, 20, 21]. However, sodium appetite is a complex process in which the interactions and homeostatic balance between the preference for water and sodium are involved. Therefore, in the current study, we extended the HRL model to a multi-dimensional model by simulating two internal states of water and sodium. Based on this extension, we attempted to provide a mechanistic understanding of previous findings using experiments with dynamic interactions between the internal states of water and sodium. Finally, by simulating an experimental manipulation of the neural activity of a particular brain nucleus known to be involved in sodium appetitive behavior, we provide a hypothesis for the function of a specific neural population in the control of sodium appetite.

Results

Model overview

In the current study, sodium appetitive behavior was modeled using the HRL model [2, 20, 21]. This model is based on the assumption that homeostasis is an RL process, in which the minimization of deviations of internal states from an optimal level (i.e., homeostasis) is treated as a computation for maximizing the sum of rewards. In the HRL model, a multi-dimensional metric space in which each dimension represents an internal state (such as body temperature, blood glucose density, water state, and sodium level) was defined as a “homeostatic space”. In this homeostatic space, the drive function $D(H_t)$ was defined as the distance between the internal state of the i -th component (e.g., water or sodium) at time t , $H_{i,t}$, and the ideal internal state H_i^* :

$$D(H_t) = \sqrt[m]{\sum_{i=1}^N |H_i^* - H_{i,t}|^n} \quad \#(1)$$

Where m and n are free parameters to define the distance, and N is the total number of dimensions for internal states (e.g., water, sodium, etc.). When the internal state approaches the ideal state, the value of the drive function should be reduced. Based on this drive function, the reward r is determined as a change in the values of the drive function from time t to time $t+1$. Specifically, to implement nutrient intake, the internal state at time $t+1$ should contain the amount of nutrient intake at time t , defined as K_t :

$$\begin{aligned} r(H_t, K_t) &= D(H_t) - D(H_{t+1}) \\ &= D(H_t) - D(H_t + K_t) \quad \#(2) \end{aligned}$$

159

160 As described later, in the HRL model, the intake of taste stimuli ($K^\wedge$) can be modeled
161 as a predictor of the actual nutrient intake (K). Under this assumption, the reward was
162 calculated as follows:

163

$$r(H_t, \hat{K}_t) = D(H_t) - D(H_t + \hat{K}_t) \# (3)$$

164

165 The RL process was modeled as a Rescorla-Wagner model. In this model, the values
166 of action a (e.g., sodium intake, do nothing...), $Q(a)$, are renewed based on the
167 reward prediction error:

$$Q_{t+1}(a) = Q_t(a) + \alpha(r_t - Q_t(a)) \# (4)$$

168 Where, α is the learning rate.

169

170

171 **Experiment 1: Sodium homeostasis according to the HRL.**

172 First, we confirmed that the homeostatic control of the internal states of sodium can
173 be replicated using the framework of the HRL model. In this experiment, model mice
174 were able to choose to either perform *saltwater intake* or *do nothing* (Fig 1B). The
175 action values of intake and do nothing were both set to 0. Therefore, the initial action
176 selection was random ($P(\text{intake}) = 0.5$) (Fig 1C). After several random choices of
177 sodium intake, the action value of intake was reinforced, and the internal state
178 approached the ideal state. In approximately 20 trials, $P(\text{intake})$ was nearly 1, and the
179 internal state rapidly reached the ideal state. When the internal state exceeded the
180 ideal state after approximately 80 trials, sodium intake became a punishment, and the
181 action value of *do nothing* increased. After several trials of *do nothing*, at around trial

number 140, due to the decay assumed to be a natural loss of internal sodium (see Method for more details), the internal state became lower than the ideal state. Therefore, the action value of sodium intake increased again. Through the repetitions of this cycle, the model successfully achieved homeostatic control of the internal states of sodium (Fig 1C).

Fig 1. Homeostatic behavior according to the homeostatic reinforcement learning (HRL) model. (A) Schematic drawing of the computational process of the HRL model. (B) Definition of a state and two actions in experiments 1 and 2. (C) Example of homeostatic behavior. Changes in internal sodium state, the value of each action (Q-value), probability of sodium intake ($P(\text{Intake})$), and magnitude of reward over time were plotted. The dotted line in the panel of the internal state indicates the ideal point ($H = 50$) of sodium taste. In the panel of the probability of sodium intake ($P(\text{Intake})$), the blue marks on axis $P = 1.0$ represent 'Intake behavior', and those on axis $P = 0.0$ indicate 'Do nothing'. At the beginning of the simulation, the internal sodium state and Q-values for each action were set to 0. After several random selections of action, the sodium Q-value of intake was increased, and the internal sodium state quickly reached the ideal point, maintaining homeostatically-regulated behaviors.

Experiment-2: Sense of taste as a predictor of changes in internal states.

In the homeostatic control of sodium states, as well as in the monitoring of the internal states, the sense of taste may play an important role. In the HRL model, this assumption can be tested by implementing this sensory modality as a predictor of

changes in the internal states induced by nutrient intake [2]. In this study, we simulated an intragastric infusion test [10]. In the intragastric infusion test, there were three groups of animals: (1) a control group of sodium-depleted mice, (2) a NaCl-IG group of sodium-depleted animals treated with intragastric infusion of saltwater before the test, and (3) a NaCl-oral group with sodium-depleted mice stimulated with sodium (salty taste) during the test (Fig 2B). The definitions of the states and actions were the same as in the previous experiment (Fig 1B). At the beginning of the intragastric infusion test, the internal states for the control group and NaCl-oral group were set at $H_t = 0$, while for the NaCl-IG group, the internal state was set at $H_t = H^*/2$, corresponding to sodium partially supplied through a gastric infusion. Each animal model was tested in 100 trials, with a total duration of 600 s in the actual experiments. During the 100 trials, the model animals of the NaCl-oral group were assumed to have constant salty taste stimulation (Fig 2A).

As a result of the simulation, there was no significant difference in the total numbers of NaCl intakes between the control and NaCl-IG groups (Fig 2C). On the other hand, the total intake of the NaCl-oral group was clearly lower than that of the control and NaCl-IG groups. These trends were similar to those observed in animal experiments [10] (Fig 2C).

To provide a mechanistic overview of these results, the transitions of each variable during the experiment are plotted in Fig. 2D-2F. In the NaCl-IG group, even though sodium was partly supplied through gastric infusion, an increase in the action value of sodium intake led to an increase in sodium intake behavior (Fig 2E), resulting in the number of intakes not changing significantly compared with the control group (Fig 2C).

On the other hand, in the NaCl-oral group, the action value of do nothing was reinforced (Fig 2F) based on the expectation of an increase in the internal sodium state (K^A) due to the continuous application of the salty stimulus (sodium) during the test. There were no clear differences between the action values of salt intake and do nothing (Fig 2F), and the number of intakes of the group that received NaCl as drinking water was lower than that of the control group (Fig 2C).

Fig 2. Intra-gastric and oral stimulation tests were explained using the sense of taste as a predictor of changes in the internal state. (A) Schematic illustration of experiments 2 and 3: Taste perception (K^A) was a predictor of an increase in the internal state. (B) Three groups of behavioral tests. (C) Comparison of results of the behavioral experiments [10] and the computational model. The intra-gastric infusion did not change sodium intake, while oral stimulation decreased the intake. (D-F) Transitions of the simulation. (D) The control group shows an increased Q-value of intake. (E) The congenial increase of the Q-value is seen in the NaCl-IG group. (F) The values of 'Do nothing' and 'Intake' are reinforced in the NaCl-oral group.

Experiment 3: Multi-dimensional HRL.

In this experiment, in order to describe the dynamic interaction between the internal states of water and sodium, we extended the HRL model to multiple dimensions [2]. The HRL model with multi-dimensional internal states (water state and sodium state) was assessed with a two-bottle preference test, which is a behavioral procedure used

to compare the preference toward contents in two bottles (Fig 3B). In the two-bottle preference test, the model mice were able to either perform “saltwater intake,” “water intake,” or “do nothing” (Fig 3A). In this experiment, there were four groups of animals: a control group with fulfilled initial internal states, a sodium-depleted group, a water-depleted group, and a water/salt-depleted group.

The results of the simulations revealed that the control group with fulfilled initial internal states showed continuously decreased action values of both water intake and salt intake, and the individuals in this group mostly chose to perform the do-nothing behavior. As a result, both internal states remained flat in the ideal state (Fig 3I). Accumulating these intakes, consumption from water bottles and saltwater bottles was low-keyed (Fig 3C-3E). In contrast, in the water/salt-depleted group, the action values of both water intake and salt intake rapidly increased. Reflecting these increases, both the water and sodium states also increased (Fig 3F). The accumulation of these consumptions was evidently larger in the depleted than in the control group (Fig 3C). In other circumstances, the sodium-depleted group, which was provided with large amounts of water, showed an increased action value of salt intake, resulting in an increased sodium state (Fig 3G). The total intake of water was slight, whereas salt intake was dominant (Fig 3D). The water-deficient group exhibited notable behaviors. First, both the action values of water intake and salt intake increased. The action value of water intake was larger than that of salt intake. Saltwater intake slightly increased (Fig 3H). These trends were similar to those observed in the actual animals assessed using the two-bottle preference test [9] (Fig 3C).

Fig 3. Multi-dimensional homeostatic reinforcement learning (HRL) model as a suitable explanation of the result of the two-bottle preference test. (A) Definition of a state and three actions in experiments 3 and 4. (B) Schematic drawing of experiment 3 (Two-bottle preference test). (C) Comparison of behavioral data [9] and simulation data in control (all-satisfied) and water/sodium-depleted group. Control groups show minor intakes, and both depleted groups demonstrate copious volumes of intake in both sets of data. (D) In the Na-depleted group, water intakes are slight, and saltwater intake is increased. (E) The water-deficient groups show abundant water intakes and non-negligible saltwater intakes (F-I) Transitions of the HRL models (F) The water/sodium-deficient HRL model shows strong increases of values of water and saltwater intake. (G) The action value of sodium intake singly soars in the Na-depleted model. (H) The water-depleted model shows increased values of water and saltwater intake. (I) The control model refuses both intakes.

Experiment-4: Simulated chemogenetic neural manipulation in the salt appetite network. Among the brain regions related to sodium appetite, it is assumed that neurons in the lateral parabrachial nucleus (LPBN^{Htr2c} neurons) have a suppressive function toward sodium appetite because artificial inhibition of that neuron with chemogenetic neural manipulation (designer receptors exclusively activated by designer drugs/DREADD) increases sodium appetite [22]. LPBN^{Htr2c} neurons project to the central amygdala (CeA). Based on these previous findings, in this experiment, we hypothesized that LPBN^{Htr2c} neurons have the function to inflict a negative bias on

the action value of sodium appetitive behavior. The hypothesis was implemented as an account of action values in the HRL model, as follows:

$$Q'_s = Q_s - LPBN \#(5)$$

where Q' is the value of an action selection function. Additionally, the inhibition of LPBN^{Htr2c} neurons by DREADD was represented by the cancelation of the negative bias (i.e., LPBN in equation 5 was set to 0). With this implementation, we tested our hypothesis about LPBN neurons by comparing the simulation and the observation in an actual animal experiment (Fig 4A) [22].

This experiment was performed in a two-bottle preference test using water and saltwater [22]. There were four groups of mice based on the depletion of water/salt and application of DREADD, namely water-depleted/control (no DREADD), sodium-depleted/control, water-depleted/DREADD, and sodium-depleted/DREADD.

Both DREADD groups showed more saltwater intake than the control group. However, no clear differences were observed between the no DREADD groups. In the sodium depletion groups, water intake was slight in both the no DREADD and DREADD groups. In both groups, saltwater intake rapidly increased. The intake of the DREADD group was higher than that of the no DREADD group. Water intake increased sharply in the dehydration groups. There was no clear difference between the DREADD and no DREADD groups. The control (no-DREADD) model showed maintained homeostasis (Fig 4B). Saltwater intake of the dehydration groups was slight, but the DREADD group showed a much higher intake than that of the no

DREADD group. These trends successfully replicated those observed in the actual animal experiments (Fig 4C and 4D).

Fig 4. Simulation of designer receptors exclusively activated by designer drug (DREADD), which inhibits tonic neuronal suppression of sodium appetite. (A) Schematic illustration of experiment 4: DREADD (B) Transitions of control (no-DREADD) models. Sodium intake values were subtracted from these plots by a negative bias toward sodium appetite. (C) DREADD experiments increased sodium intake [22]. (D) Homeostatic reinforcement models demonstrate the equivalent behaviors.

Discussion

In this study, we attempted to provide a mechanistic understanding of sodium appetite behavior using the HRL model. In the simple simulation of experiment 1, we confirmed that the HRL model successfully reproduced homeostasis-like behaviors by regulating sodium appetite, (i.e., approach and avoidance behavior to sodium). In addition, based on the assumption that the sense of taste is a predictor of changes in internal states, the HRL model successfully reproduced the previous observations of the intragastric infusion test that cannot be explained by the classical drive reduction theory [19]. These results support the idea that sodium appetitive behavior can be understood as a reinforcement learning process.

This idea is consistent with previous findings that the reward learning system is involved in sodium appetite behaviors. For instance, the activity of dopaminergic neurons in the VTA, which is thought to have a robust relationship with the reinforcement learning process in the brain [23, 24], was increased when sodium-depleted mice licked saltwater. In contrast, pharmacological inactivation of neural projections from the VTA to the nucleus accumbens decreased sodium intake [18]. In addition, optogenetic excitation of VTA dopaminergic neurons suppresses sodium intake in sodium-depleted mice [17]. The HRL model may provide insight into integrating these previous findings.

In the HRL model, the sense of taste was hypothesized to predict changes in the internal state. The latter means that the salty stimulus represents an immediate inducer of reinforcement, but this was not the case for the actual changes in the internal state. Nevertheless, only actual intake may result in the satiation of internal states. This assumption is consistent with the fact that gastric infusion of water does not work as a reinforcer, in contrast to oral intake of water, which can work as a reinforcer [2, 25]. In addition, artificial sweeteners, including saccharine and sucralose, can function as reinforcers [26-28], although their effects are relatively weaker than those of sucrose, which induces substantial changes in the internal state. From the perspective of computational theory, this assumption of the HRL model corresponds to the predictive processing (also referred to as predictive coding, or active inference) theory, in the sense that homeostasis is understood as the prediction of interoceptive sensory states and minimization of prediction error [29]. As such, homeostasis and sodium appetite behavior may provide an ideal research field to unify these computational theories.

378

379 In experiment 3, we extended the HRL model to multi-modal data, and successfully
 380 reproduced the behavioral tests in which water and sodium appetite regulate each
 381 other. In the current study, as the simplest attempt, the internal states of sodium and
 382 water were assumed to contribute equally. However, in an actual biological system,
 383 the homeostatic maintenance of water and sodium may not be exactly equal. For
 384 example, the effects of intragastric infusion of water and saltwater on the appetite of
 385 water and sodium may have different timescales [9, 11]. A more detailed
 386 implementation of such differences in water and sodium appetite may provide novel
 387 insights for understanding the system-level mechanisms of sodium appetite.

388

389 Finally, we successfully replicated the characteristic features of LPBN^{Htr2c}
 390 suppression experiments using DREADD. In the proposed model, we assumed that
 391 LPBN^{Htr2c} neurons provide a tonic negative bias toward sodium appetite via the
 392 central amygdala. That is, the LPBN-amygdala projection provides a negative bias of
 393 action selections and has no direct contribution to the learning of the Q-value. Based
 394 on this assumption, the learning of the Q-value canceled this negative bias, resulting
 395 in the inhibition of LPBN neurons, enhancing the relative value of salt intake, and
 396 provoking a rapid increase in salt-intake behavior. This assumption is consistent with
 397 previous findings that an immediate increase in sodium craving by sodium depletion
 398 may not be mediated by learning processes [30]. In addition, CeA is an essential
 399 region for both hedonic and aversive intakes. For example, CeA
 400 prepronociceptin-expressing neurons are activated by hedonic intake and promote
 401 palatable food consumption [31]. In contrast, activation of PKC- δ^+ neurons in the
 402 lateral subdivision of CeA inhibits feeding [32]. Further investigation of the neural

connections from LPBN to CeA, together with the HRL model, may provide fundamental information for the development of more precise algorithms.

Here, we discuss the significance of constructing a computational model for sodium appetite. To understand complex systems such as the brain, investigations from three levels are essential, namely computational theory, representation and algorithm, and hardware implementation [33]. In this study, we provided a mechanistic explanation of sodium appetite behavior by bridging previous findings from the aspects of these three levels. Although the model behaviors were evaluated only for their quantitative similarities with the actual animal experiments, the model can also provide quantitative predictions of unobservable latent variables, such as reward prediction error, values of actions (motivation toward nutrient intake), and prediction of internal states. Investigations of neural correlates of such latent variables may provide a deeper understanding of the underlying neural mechanisms of sodium appetite and homeostatic behavior.

Methods

Sodium homeostasis

An overview of the HRL model is described in the Results section (see Equations 1–4). To investigate the applicability of the HRL model to sodium appetite behavior, we performed a sodium intake test (Experiment 1). The computation algorithm is illustrated in Fig 1A. In this experiment, only the internal state of sodium was considered. An external state (S_0) and two actions, do nothing (a_0) and intake (a_1),

were assessed (Fig 1B). Action selection depends on the relative magnitudes of the values of each action (Q-value), following the soft-max function:

$$P_t(a_k) = \frac{\exp(\beta \cdot Q_t(a_k))}{\sum_j \exp(\beta \cdot Q_t(a_j))} \quad \#(6)$$

where $P_t(a_k)$ is the probability of an action a_k to be selected at time t , and β is the inverse temperature, a parameter controlling the randomness of an action. In Experiment 1, the Q-values of both actions were set to 0. Therefore, the first action was randomly chosen. When intake behavior was performed, the internal sodium state increased with K , a constant parameter defining the size of sodium intake. When nothing was chosen, K was set to 0. At $t = 0$, to represent sodium depletion, the first internal state ($H = 0$) was far lower than the ideal state ($H^* = 50$). At this stage, the value of the drive function was large because the drive function corresponds to a type of distance from the internal state of time t (H_t) to the ideal state ($H^* = 50$) (Equation 1). If an agent performed the intake behavior at this moment, the internal state increased and the drive function became smaller, resulting in a positive reward (Equation 2).

In addition, the natural decrease of sodium state was implemented as follows using the temporal decay constant τ (Equation 7).

$$H_{t+1} = \left(1 - \frac{1}{\tau}\right) H_t \quad \#(7)$$

As a result, the calculation of the reward value was determined as follows (Equation 8).

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

$$= D(H_t) - D\left(\left(1 - \frac{1}{\tau}\right)H_t + K_t\right) \#(8)$$

448

449 To renew Q-values with the reward value, we used a conventional Rescorla-Wagner
450 model [35], where i indicates each action, α is learning rate, and $r - Q$ represents
451 reward prediction error [35]. After this update of the Q-values, the agent chooses the
452 next action. The detailed values of the parameters of this simulation are listed in
453 Table S1.

454

455

Oral sense as a predictor of changes in internal states

457 In experiments 2–4, we introduced oral sense $K^\wedge$, which represents the prediction of
458 changes in the internal state. The definition of the reward (Equation 3) was also
459 updated as follows:

$$r(H_t, \widehat{K}_t) = D(H_t) - D\left(\left(1 - \frac{1}{\tau}\right)H_t + \widehat{K}_t\right) \#(9)$$

460

461 Thus, the reward was defined as changes in internal states and taste sensory input
462 as predictors for changes in internal states. The other functions were the same as
463 those in experiment 1.

464

465 The behavioral experiment of intragastric infusion is shown in Fig. 2A. The animals
466 were grouped into three groups: control, intragastric, and oral stimulation. Only the
467 control group was salt-depleted. In the intragastric (IG) infusion group was sodium
468 depleted. In addition, an intragastric cannula was inserted into the gut of the mice.
469 Before the intake test, saltwater was directly infused into the gut. In the oral

stimulation group, the animals were stimulated with a strong salty stimulus in the intake test. To simulate the IG infusion group, we hypothesized that the internal state of sodium was set to the level of half-satisfaction at the beginning of the intake test. To model the oral stimulation group, a salty taste was supplied, regardless of the selected actions. The definitions of states and actions were the same as in Experiment 1 (Fig 1B). The algorithm including taste sensory input as a predictor of changes in the internal state is illustrated in Fig 2A. The detailed values of the parameters used in this experiment are shown in Table (in Table S2).

Two-bottle preference test

For the simulation of the two-bottle preference test in experiment-3, we set two internal states corresponding to water and sodium. Thus, in Equation 1, the number of dimensions of internal state N was set to 2. Each internal state updates as follows:

$$H_{w,t+1} = \left(1 - \frac{1}{\tau_w}\right) H_{w,t} \quad \#(10)$$

$$H_{s,t+1} = \left(1 - \frac{1}{\tau_s}\right) H_{s,t} \quad \#(11)$$

where H_w is water state, and H_s is sodium state.

The detailed parameters were partially different from those of the taste expectation.

The detailed values of the parameters in experiment 3 are shown in Table S3.

DREADD experiment

To simulate the DREADD experiment, we assumed that LPBN^{Htr2c} neurons provided tonic suppression of sodium appetite as implemented by a tonic negative bias in the selection of sodium intake action as follows:

$$\text{Control: } Q'_s = Q_s - \text{LPBN} \#(12)$$

where LPBN is a positive constant value corresponding to the negative bias of the LPBN^{Htr2c} neuron. Note that Q' is used only for action selections and Q-values are updated with previous Q-values and the reward, regardless of Q' and the tonic bias (Equation 4).

DREADD treatment was implemented as the cancelation of this negative bias by adding the positive value of *drd* as follows:

$$\text{DREADD: } Q'_s = Q_s - \text{LPBN} + \text{drd} \#(13)$$

In the current experiment, the values of LPBN and *drd* were set to be the same. At the beginning of the saltwater intake test, the initial Q-value for salt intake behavior was set to LPBN, corresponding to the state, where Q' for both actions was 0 (i.e., the probabilities for salt intake and do-nothing were 0.5). The initial water state was set to 0 for the dehydration group, and the initial sodium state was set to 0 for the sodium-depleted group. The detailed values of the parameters in Experiment 3 are shown in Table S4.

Supporting information captions

S1 Table. Free parameters for experiment 1.

513 Parameters for experiment of sodium homeostasis according to the HRL.

514 **S2 Table. Free parameters for experiment 2.**

515 Parameters for a simulation with sense of taste as a predictor of changes in internal
516 states.

517 **S3 Table. Free parameters for experiment 3.**

518 Parameters used for multi-dimensional HRL.

519 **S4 Table. Free parameters for experiment 4.**

520 Parameters for a simulation of chemogenetic neural manipulation

521

References

1. Cannon WB. Organization for physiological homeostasis. *Physiol Rev.* 1929;9: 399-431.
2. Keramati M, Gutkin B. Homeostatic reinforcement learning for integrating reward collection and physiological stability. *eLife.* 2014;3: 1-26.
3. Feng P, Huang L, Wang H. Taste bud homeostasis in health, disease, and aging. *Chem Senses.* 2014;39: 3-16.
4. Steinbach S, Hummel T, Böhner C, Berkold S, Hundt W, Kriner M, et al. Qualitative and quantitative assessment of taste and smell changes in patients undergoing chemotherapy for breast cancer or gynecologic malignancies. *J Clin Oncol.* 2009;27:1899-1905.
5. Sánchez-Lara K, Sosa-Sánchez R, Green-Renner D, Rodríguez C, Laviano A, Motola-Kuba D, et al. Influence of taste disorders on dietary behaviors in cancer patients under chemotherapy. *Nutr J.* 2010;9: 1-6.
6. Richter P. Increased salt appetite in adrenalectomized rats. *Am J Physiol.* 1935;115: 155-161.
7. Watanabe E, Fujikawa A, Matsunaga H, Yasoshima Y, Sako N, Yamamoto T, et al. Nav2/NaG channel is involved in control of salt-intake behavior in the CNS. *Journal Neurosci.* 2000;20: 7743-7751.
8. Oka Y, Butnaru M, von Buchholtz L, Ryba NJ, Zuker CS. High salt recruits aversive taste pathways. *Nature.* 2013;494: 472-475.
9. Matsuda T, Hiyama TY, Niimura F, Matsusaka T, Fukamizu A, Kobayashi K, et al. Distinct neural mechanisms for the control of thirst and salt appetite in the subfornical organ. *Nature Neurosci.* 2017;20: 230-241.

- 546 10. Lee S, Augustine V, Zhao Y, Ebisu H, Ho B, Kong D, et al. Chemosensory
547 modulation of neural circuits for sodium appetite. *Nature*. 2019;568: 93-97.
- 548 11. Augustine V, Lee S, Oka Y. Neural control and modulation of thirst, sodium
549 appetite, and hunger. *Cell*. 2020;180: 25-32.
- 550 12. Catalanotto FA, Sweeney EA. Salivary sodium and potassium concentrations
551 in adrenalectomized rats. *Behav Biol*. 1978;24: 467-473.
- 552 13. Galaverna OG, Seeley RJ, Berridge KC, Grill HJ, Epstein AN, Schulkin J.
553 Lesions of the central nucleus of the amygdala I: Effects on taste reactivity, taste
554 aversion learning and sodium appetite. *Behav Brain Res*. 1993; 59: 11-17.
- 555 14. Chandrashekar J, Kuhn C, Oka Y, Yarmolinsky DA, Hummler E, Ryba NJP,
556 et al. The cells and peripheral representation of sodium taste in mice. *Nature*.
557 2010;464: 297-301.
- 558 15. Eaton DC, Pooler J, Vander AJ. Vander's renal physiology. 7th ed. New York:
559 McGraw-Hill Medical; 2009.
- 560 16. Resch JM, Fenselau H, Madara JC, Wu C, Campbell JN, Lyubetskaya A, et
561 al. Aldosterone-sensing neurons in the NTS exhibit state-dependent pacemaker
562 activity and drive sodium appetite via synergy with angiotensin II signaling. *Neuron*.
563 2017;96: 190-206.e7.
- 564 17. Sandhu EC, Fernando ABP, Irvine EE, Tossell K, Kokkinou M, Glegola J, et
565 al. Phasic stimulation of midbrain dopamine neuron activity reduces salt consumption.
566 *eNeuro*. 2018;5.
- 567 18. Verharen JPH, Roelofs TJM, Menting-Henry S, Luijendijk MCM,
568 Vanderschuren LJMJ, Adan RAH. Limbic control over the homeostatic need for
569 sodium. *Sci Rep*. 2019;9: 1-11.

- 570 19. Hull CL. Principles of behavior: an introduction to behavior theory. New York:
571 D. Appleton-Century Co.; 1943.
- 572 20. Hulme OJ, Morville T, Gutkin B. Neurocomputational theories of homeostatic
573 control. *Phys Life Rev.* 2019;31: 214-232.
- 574 21. Keramati M, Durand A, Girardeau P, Gutkin B, Ahmed SH. Cocaine addiction
575 as a homeostatic reinforcement learning disorder. *Psychol Rev.* 2017;124: 130-153.
- 576 22. Park S, Williams KW, Liu C, Sohn JW. A neural basis for tonic suppression of
577 sodium appetite. *Nat Neurosci.* 2020;23: 423-432.
- 578 23. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and
579 reward. *Science.* 1997;275: 1593-1599.
- 580 24. Nakanishi S, Hikida T, Yawata S. Distinct dopaminergic control of the direct
581 and indirect pathways in reward-based and avoidance learning behaviors.
582 *Neuroscience.* 2014;282: 49-59.
- 583 25. McFarland D. Separation of satiating and rewarding consequences of
584 drinking. *Physiol Behav.* 1969;4: 987-989.
- 585 26. Fernandes AB, Alves da Silva J, Almeida J, Cui G, Gerfen CR, Costa RM, et
586 al. Postingestive modulation of food seeking depends on vagus-mediated dopamine
587 neuron activity. *Neuron.* 2020;106: 778-88 e6.
- 588 27. Hughes LH. Saccharine reinforcement in a T maze. *J Comp Physiol Psychol.*
589 1957;50: 431-435.
- 590 28. Collier G, Siskel M, Jr. Performance as a joint function of amount of
591 reinforcement and inter-reinforcement interval. *J Exp Psychol.* 1959;57: 115-20.
- 592 29. Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci.*
593 2010;11: 127-38.

- 594 30. Tindell AJ, Smith KS, Pecina S, Berridge KC, Aldridge JW. Ventral pallidum
595 firing codes hedonic reward: when a bad taste turns good. J Neurophysiol. 2006;96:
596 2399-2409.
- 597 31. Hardaway JA, Halladay LR, Mazzone CM, Pati D, Bloodgood DW, Kim M, et
598 al. Central amygdala prepronociceptin-expressing neurons mediate palatable food
599 consumption and reward. Neuron. 2019;102: 1037-1052 e7.
- 600 32. Cai H, Haubensak W, Anthony TE, Anderson DJ. Central amygdala
601 PKC-delta(+) neurons mediate the influence of multiple anorexigenic signals. Nat
602 Neurosci. 2014;17: 1240-1248.
- 603 33. Marr D. Vision: a computational investigation into the human representation
604 and processing of visual information. USA: Henry Holt and Co., Inc.; 1982.
- 605 34. Sutton RS, Barto AG. Reinforcement learning: an introduction. 2nd edition.
606 Massachussets: MIT Press; 2018.
- 607

Control

IG infusion

Oral stimulation

Trial

A **B** **C** **D** **E** **F** **G** **H** **I**