Title	Homeostatic reinforcement learning explains foraging strategies
Author(s)	Arikawa, Etsushi; Yoshida, Naoto; Kanazawa, Hoshinori et al.
Citation	The 11th International Symposium on Adaptive Motion of Animals and Machines (AMAM2023). 2023, p. 160-161
Version Type	VoR
URL	https://doi.org/10.18910/92317
rights	
Note	

Homeostatic reinforcement learning explains foraging strategies

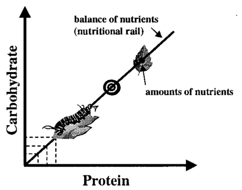
Etsushi Arikawa[1], Naoto Yoshida[1], Hoshinori Kanazawa[1 2], Yasuo Kuniyoshi[1 2]
[1]Department of Mechano-Informatics, The University of Tokyo, Japan
[2]Next Generation Artifitial Intelligence Research Center, The University of Tokyo, Japan
{*arikawa, n-yoshida, kanazawa, kuniyosh*}@*isi.imi.i.u-tokyo.ac.jp*

1 Introduction

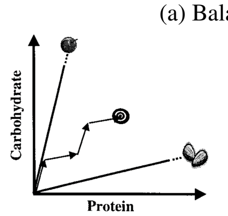
Animals need to ingest nutrients for survival, but they should maintain a balanced nutritional state, as an excess of nutrients can have negative effects on their vital functions. This maintaining process is known as homeostasis, which relates to animal behavior. Using the Geometric Framework (GF, [1]), three foraging strategies have been identified that can evaluate both homeostasis and animal behavior. However, the relationship between these strategies and the animals' internal mechanisms have not been well understood. Homeostatic reinforcement learning allows to discuss at the neural level and to explain behavior by designing internal mechanisms. Therefore, this study aims to explain the foraging strategies by using homeostatic reinforcement learning and changing the internal mechanisms of the agents.

2 Geometric Framework

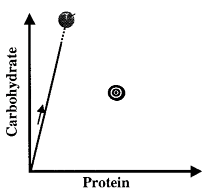
The GF can represent the relationship between amounts of nutrients and the intake target (Fig.1). The intake target can be achieved with a balanced food or complementary foods, but not with an imbalanced food.



(a) Balanced food



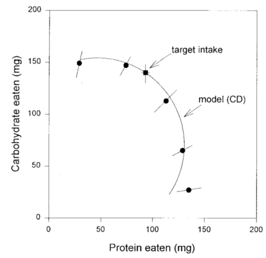
(b) Complementary foods



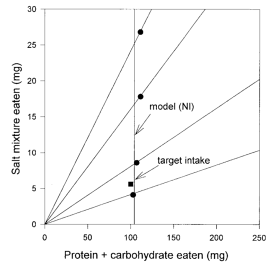
(c) Imbalanced food

Figure 1: Example of GF. Amount of nutrient can be represented by axes. Angle of lines (nutritional rail) describe the ratio of nutrient in foods. (Source: [1])

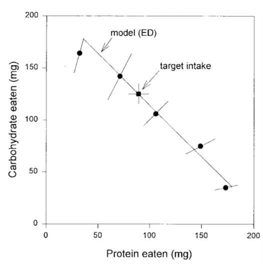
Animals need to ingest imbalanced foods if there are no balanced or complementary foods available in their environment, compromising their nutritional requirements for different nutrients. Three foraging rules have been reported to account for this compromise[1]: Closest Distance (CD) rule, No Interaction (NI) rule, Equal Distance (ED) rule (Fig.2). First, the CD rule (Fig. 2a) is the strategy that minimizes the total deviation in nutrient space from the intake target. Second, the NI rule (Fig. 2b) is a strategy that satisfies only one requirement, regardless of the excess or shortfall of the other. Finally, the ED rule (Fig. 2c) is a strategy that regulates intake such that the excess of one nutrient matches the deficit of the other.



(a) CD rule



(b) NI rule



(c) ED rule

Figure 2: Rules of compromise. (Source: [1])

3 Method

We used homeostatic reinforcement learning to explain foraging strategies related to homeostasis at the neural level. In our experiments, we adopted two resource environments (Fig. 3) in which agents had red and blue internal states, and they could ingest nutrients by foraging for food. Ingesting nutrients increased the internal states corresponding to the type of nutrients. The internal states of the agents moved depending on predefined updates that represented the movement of internal states from time step t to time step $t+1$. If their internal states deviated from set points, the episode terminated. The policy for homeostasis was optimized by proximal policy optimization [3] to maximize the expectation of the future cumulative sum of rewards. The reward function [2] is described as

$$r(\mathbf{X}_t, \mathbf{K}_t) = D(\mathbf{X}_t) - D(\mathbf{X}_{t+1}) = D(\mathbf{X}_t) - D(\mathbf{X}_t + \mathbf{K}_t), \quad (1)$$

where $\mathbf{X}_t = \{x_t^{\text{red}}, x^{\text{blue}_t}\}^\top$ is the vector of internal states, $\mathbf{X}_* = \{x_*^{\text{red}}, x_*^{\text{blue}}\}^\top$ is the vector of setpoints (interoceptive targets), $\mathbf{K}_t = \{k_t^{\text{red}}, k_t^{\text{blue}}\}^\top$ is the vector of inlets, $D(\mathbf{H}_t) = \sum_{i\in\{\text{red,blue}\}}^N |x_*^i - x_t^i|^2$ is the drive function. Updates of internal states are defined as follows.

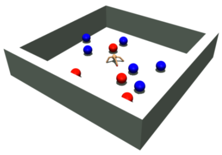


Figure 3: Two Resource Enviromnent. Red and blue balls represent foods for ingesting nutrients. (Source: [4])

3.1 Updates for CD rule

The updates for expression of CD rule are described as

$$x_{t+1}^i = x_t^i - \delta_{\text{default}}^i + \delta_{\text{food}}^i I_t^i, \tag{2}$$

where $i \in \{\text{red}, \text{blue}\}$ is the nutritional type, $\delta\,_{\text{default}}^{\,i} = 0.00015$ is the default consumption of the nutrient, $\delta\,_{\text{food}}^{\,i}$ is the inlet of the nutrient when agents captures a food. I_t^i is one if agents ingest a food; otherwise zero [4]. We consider that this equation can express CD rule, but it is insufficient for other rules. Hence, we proposed two equations as follows.

3.2 Updates for NI rule

Our proposed updates for expression of NI rule, which has a mechanism to ignore the error of one nutrient, are described as

$$\begin{cases} x_{t+1}^{\text{blue}} &= x_t^{\text{blue}} - \delta_{\text{default}}^{\text{blue}} + \delta_{\text{food}}^{\text{blue}} I_t^{\text{blue}} \\ x_{t+1}^{\text{red}} &= x_*^{\text{red}}. \end{cases} \tag{3}$$

3.3 Updates for ED rule

We propose the updates for expression of ED rule, which has converting mechanism(Fig.4), and are described as

$$x'^i_t = x_t^i - \text{sgn}(e_t^i) c_t, \tag{4}$$

where x'^i_t is the internal state after converting, e_t^i is the error from the setpoints($e_t^i = x_t^i - x_*^i$). c_t represents converted excess and are described as

$$c_t = \{\text{sgn}(e_t^{\text{red}}) \oplus \text{sgn}(e_t^{\text{blue}})\}\min(|e_t^{\text{red}}|, |e_t^{\text{blue}}|), \tag{5}$$

where $\oplus$ means exclusive OR, and takes one if signs of each errors are different; otherwise zero.

4 Experiments and Results

First of all, agents learn foraging behavior in the environment where the intake target can be reached by ingesting two complementary foods. After learning the policy for foraging behavior, the foraging strategies of agents are evaluated in another environment where the intake target cannot be reached by ingesting imbalanced food.

Our experimental results (Fig.5) are consistent with the foraging strategies observed in animals (Fig.2), suggesting that homeostatic reinforcement learning can explain these strategies.

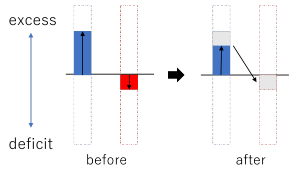
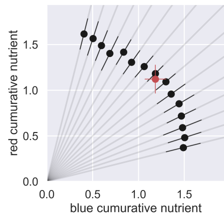
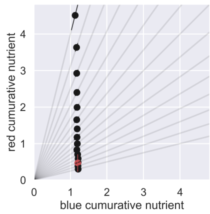


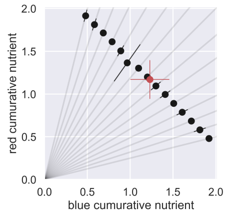
Figure 4: Example of nutritional conversion. Before conversion, the internal blue state is excess and the internal red state is deficit. The excess of blue nutrients are converted to red nutrients, and the internal red states close to the setpoint.



(a) CD rule



(b) NI rule



(c) ED rule

Figure 5: Cumulative ingested nutrients with different ratio. Black circles represent means of cumulative nutrients, red circles represent intake targets, gray lines represent nutritional rails, solid lines represent standard deviation of cumulative nutrients.

5 Conclusion

We proposed two updating rules of the agent's internal states to express the NI rule and the ED rule. Our results suggest that three foraging strategies can be explained as emergent behaviors of homeostatic reinforcement learning [4]. We assumed a fixed foraging strategy in an animal; rather, in real animals, an individual species may change foraging strategies depending on environmental conditions. For example, it has been reported that gregarious insects exhibit the ED rule, while solitary insects exhibit the CD rule [5]. These effects of environmental conditions on foraging strategies are left for our future work.

References

[1] D. Raubenheimer and S. J. Simpson, "Integrating nutrition: a geometrical approach," in *Proceedings of the 10th International Symposium on Insect-Plant Relationships*, pp. 67–82. Springer, 1999.

[2] M. Keramati and B. Gutkin, "A reinforcement learning theory for homeostatic regulation," *Advances in neural information processing systems*, vol. 24, 2011.

[3] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[4] N. Yoshida, T. Daikoku, Y. Nagai, and Y. Kuniyoshi, "Embodiment perspective of reward definition for behavioural homeostasis," in *Deep RL Workshop NeurIPS 2021*, 2021.

[5] S. J. Simpson and D. Raubenheimer, "Assuaging nutritional complexity: a geometrical approach," *Proceedings of the Nutrition Society*, vol. 58, no. 4, pp. 779–789, 1999.