Simulation of foraging behavior using a decision-making agent with Bayesian and inverse Bayesian inference: Lévy and Brownian walk-like search patterns derived from differences in prey distributions

Shuji Shinohara[1*], Hiroshi Okamoto[1], Nobuhito Manome[1,2], Pegio-Yukio Gunji[3], Yoshihiro Nakajima[4], Toru Moriyama[5], Ung-il Chung[1]

[1] Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

[2] Department of Research and Development, SoftBank Robotics Group Corp., Tokyo, Japan

[3] Department of Intermedia Art and Science, School of Fundamental Science and Technology, Waseda University, Tokyo, Japan

[4] Graduate School of Economics, Osaka City University, Osaka, Japan

[5] Faculty of Textile Science, Shinshu University, Ueda, Japan

* Corresponding author

E-mail: shinohara@bioeng.t.u-tokyo.ac.jp

Author contributions

Conceptualization, S.S. and H.O.; Formal Analysis, S.S.; Methodology, S.S.; Software, S.S.; Writing—Original Draft Preparation, S.S.; Writing—Review and Editing, H.O., N.M., P.Y.G., Y.N., T.M. and U.C.; Supervision, P.Y.G. and U.C.; Project Administration, U.C.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Abstract

Lévy walks, random walks where the frequency of occurrence of a linear step length follows a power-law distribution, are found in the migratory behavior of organisms at various levels, from bacteria and T cells to humans. However, it has been pointed out that in human migratory behavior, the step length series may have temporal correlation (i.e., it is not random walk) and that there is some relationship between this time dependency and the fact that the frequency distribution of step length follows the power-law distribution. Furthermore, some large marine organisms have been found to switch between Lévy and Brownian walks, wherein the frequency of occurrence of the step length is characterized by an exponential distribution (EP), depending on the difficulty of prey acquisition. However, as of now it has not been clarified how the aforementioned three phenomena arise: the positive correlation created in the step length series, the relation between the positive correlation of the step length series and the form of an individual's step length distribution, and the switching between Lévy and Brownian behavior depending on the abundance of prey. The purpose of this study is to simulate foraging behavior by three Bayesian decision-making agents: an agent simultaneously performing both knowledge learning and knowledge-based inference, an agent performing only learning, an agent performing only inference, and to analyze how the aforementioned three phenomena arise. The simulation results show that only the agent with both learning and inference has a mechanism that simultaneously causes all the phenomena. This suggests that simultaneous learning on prey distribution and inference based on the knowledge gained in exploratory behavior under incomplete information may be the key to the emergence of Lévy walk-like patterns found in humans and marine organisms.

Introduction

Lévy walks are found in the migratory behavior of organisms at various levels, from bacteria and T cells to humans [1–6]. The Lévy walk is a type of random walk where the frequency of occurrence of a linear step length l follows a power-law distribution $P(l) \sim l^{-\eta}, 1 < \eta \leq 3$. Compared with the Brownian walk that is also a type of random walk (the frequency of occurrence of the step length l is characterized by an exponential distribution (EP) $P(l) \sim e^{-\lambda l}$), the Lévy walk is characterized by the appearance of linear movements over extremely long distances, and the reason behind the formation of such patterns in the movement of organisms has been probed in the past [7].

The formation of these movement patterns can be explained by the Lévy-flight foraging hypothesis [8,9] that states that when prey is sparse and randomly distributed and predators have no information (memory) about the prey, the Lévy walk, as a random search, becomes the optimal foraging behavior, whereas Brownian motion is theoretically optimal when prey is abundant [10,11]. In fact, it has been indicated that in a few marine organisms, when prey is abundant, a Brownian walk is observed, whereas when prey is scarce, a Lévy walk is observed [4,12]. However, Humphries et al. state that their analysis could not detect how the movement patterns arose, that is, whether the patterns identified were an adaptive behavior or whether observed patterns were an emergent property of the spatial distributions of prey; they further concluded that experimental studies were needed in vertebrates to test for the presence of Lévy behavior patterns in the absence of complex prey distributions [4]. Mechanisms for switching between Lévy and Brownian behavior have been the focus of considerable attention [13].

Humans have advanced learning and estimation capabilities and can acquire useful information through systematic search even in dynamic and uncertain non-stationary environments [14]. Ross et al. [15] demonstrated that in human exploration behavior, the mode of exploration changes depending on the encounter with prey. Particularly, they indicated that in response to encounters, hunters produce a more

tortuous search of patches of higher prey density and spend a large part of their search time in patches; however, they adopt more efficient unidirectional, inter-patch movements after failing to encounter prey over a sufficiently long period. This type of search behavior is called area-restricted search (ARS) [15,16]. Wang et al. [17] demonstrated that the frequency distribution of not only the step length, but also the duration of stay in a location, is a power-law distribution. In addition, Wang et al. [17], Rhee et al. [18], and Zhao et al. [19] demonstrated that the temporal variation of step length is autocorrelated; that is, the temporal variation of step length is time-dependent. Such a behavioral pattern, even if the frequency of occurrence of step length follows a power-law distribution, is not a pure random walk and therefore may not be called a Lévy flight because in the pure Lévy flight model, which describes the random walk completely with power-law distributed move-lengths (displacements), the expected correlation between two successive move-lengths is zero [17]. In addition, it has been emphasized that the dependence of the step length on time is related to the frequency distribution of the step length following a power-law distribution [17]. It is necessary to clarify the mechanisms that create these relationships.

Perhaps, in organisms with advanced cognitive functions, such as humans, various movement patterns may appear as a result of learning and decision-making through interaction with the environment and other individuals, rather than as a result of random search [14,20]. Gunji et al. [21,22] proposed a novel mathematical model called Bayesian and inverse Bayesian (BIB) inference that learns and modifies hypotheses based on observed data and subsequently performs inference using the hypotheses and observed data; in other words, learning and inference are performed simultaneously and seamlessly. Furthermore, Gunji et al. [23] proposed a swarm model that incorporates BIB inference into the self-propelled particle model and showed that Lévy walks emerge universally in swarming behaviors through simulations employing the model. The process of simultaneous learning and inference, similar to that in BIB inference, has been suggested in the human-related experiments. Zaki and Nosofsky conducted a behavioral experiment that demonstrated the influence of unlabeled test data on learning [24]. Their experiment demonstrated that

humans do not fix their hypothesis after training; instead, unlabeled test data influence the learned hypotheses of humans [25].

In this study, the authors propose a novel model of BIB inference based on Bayesian inference [26,27] as an algorithm for inference and the expectation–maximization (EM) algorithm [28,29], an unsupervised clustering method, as an algorithm for learning (inverse Bayesian inference). The model is used as a decision-making model for foraging agents. On the one hand, Bayesian inference is an algorithm that narrows down the hypothesis that best explains the observed data, to one of the several hypotheses (generating distributions) prepared in advance. On the other hand, the EM algorithm is an algorithm that, by modifying the initial distribution based on the observed data, approaches the distribution that best explains the observed data. Bayesian inference and EM algorithm have the same goal of estimating the most appropriate generative model based on observed data. However, the direction of the estimation is opposite. If the estimated generative model is considered as knowledge, then the EM algorithm can be assumed to be a type of learning because it forms and modifies knowledge based on observed data. Further, Bayesian inference can be considered to be knowledge-based inference because it decides the best among the several hypotheses (generative models) prepared in advance.

The purpose of this study is to simulate foraging behavior by agents each with one of three decision-making models (a BIB agent, an agent performing only learning and an agent performing only inference), and to analyze how the following three phenomena arise: 1) the positive correlation created in the displacement series, 2) the correlation between the positive correlation of the displacement series and the form of an individual's displacement distribution, and 3) the switching between Lévy and Brownian behavior depending on the abundance of prey. The results of the foraging simulations show that only the BIB agent has a mechanism that simultaneously causes the aforementioned phenomena. In addition, the BIB agent can acquire prey efficiently, eliminating the trade-off between speed and accuracy that the other agents encounter.

Results

Fig 1 Distribution of prey density in a two-dimensional space. Prey is acquired probabilistically in proportion to the prey density at the foraging point. (a) When prey is scarce (the number of patches is one). (b) When the prey is abundant (the number of patches is one). (c) When the number of patches is five.

The authors performed foraging simulations for the cases when prey was scarce (Fig 1(a)) and abundant (Fig 1(b)) in two-dimensional space. The simulation was also performed for the case wherein there were several patches (Fig 1(c)). The agent probabilistically decides where to forage according to its own estimated model that is represented by a normal distribution in this study. Prey is acquired stochastically in proportion to the density of prey at the foraging point. The objective of the agent is to acquire as much prey as possible within a certain period. In other words, the agent is required to search for the point with the highest prey density as soon as possible.

Case of single patchy prey distribution

Fig 2 Example of foraging behavior of BIB agent when prey is scarce. (a) Movement pattern. Two foraging points at adjacent time points are connected by a straight line. (b) Time evolution of the hypothesis with the highest confidence. (c) Time evolution of the variance in the x-direction of the model for the hypothesis with the highest confidence. (d) Time evolution of the variance in the y-direction of the model for the hypothesis with the highest confidence. (e) Time evolution of distance between the center of the model

for the hypothesis with the highest confidence and the center of the actual prey distribution. (f)Time evolution of the expected reward expressed by the prey density at the foraging point.

Fig 3. Example of foraging behavior of BIB agent when prey is abundant. (a) Movement pattern. In the figure, two foraging points at adjacent time points are connected by a straight line. (b) Time evolution of the hypothesis with the highest confidence. (c) Time evolution of the variance in the x-direction of the model for the hypothesis with the highest confidence. (d) Time evolution of the variance in the y-direction of the model for the hypothesis with the highest confidence. (e) Time evolution of distance between the center of the model for the hypothesis with the highest confidence and the center of the actual prey distribution. (f) Time evolution of the expected reward expressed by the prey density at the foraging point.

Fig 2 shows an example of the foraging behavior of a BIB agent when prey is scarce (Fig 1(a)). Fig 2(a) plots the foraging points at each time point in a two-dimensional space. In the figure, two points d^{t-1} and d^{t} at adjacent times are connected by straight lines. Fig 2(b) shows the time evolution of the hypothesis $h^{t}_{\max}$ with the highest confidence. Fig 2(c) and 2(d) show the time evolutions of the variance in the x- and y-directions of the model for the hypothesis $h^{t}_{\max}$, respectively. Fig 2(e) shows the time evolution of the distance between the center $\mu^{t}_{\max}$ of the model estimated by the BIB agent, that is, the mean of the generative distribution for the hypothesis $h^{t}_{\max}$ and the actual center μ_f of the prey distribution. Fig 2(f) shows the time evolution of the expected reward, expressed in terms of prey density $\Phi\left(d^{t} \mid \mu_f, \Sigma_f\right)$ at the foraging point d^{t}. The simulation was conducted up to the 10000th step; however, these figures are shown up to the 1000th step for convenience. The details of each agent model and simulation setup can be found in the Methods section.

As can be seen from the figure, initially the BIB agent changes its hypothesis rapidly and travels

long distances; however, at around Step 135, when it starts to acquire more food, it narrows down to a single hypothesis. Subsequently, the model of the hypothesis is modified and the central coordinate $\mu^t_{\max}$ is set closer to the center μ_f of the prey distribution, and the variance $\Sigma^t_{\max}$ of the model for the hypothesis $h^t_{\max}$ is set closer to zero. As a result, the foraging points d^t become concentrated at the center of the prey distribution, and the expected reward increases further. Therefore, it can be seen that the search behavior of the BIB agent is divided into two stages: a rough search for an appropriate hypothesis in the initial stage and a detailed search through the model revision after the hypothesis has been fixed.

Fig 3 shows an example of the foraging behavior of a BIB agent when prey is abundant (Fig 1(b)). The vertical and horizontal axes in Fig 3(a)–3(f) represent the same as in Fig 2(a)–2(f), respectively. When prey is abundant, the hypothesis is fixed immediately after the start of the simulation, followed by a series of detailed searches. The expected reward is relatively high from the beginning because of the abundance of prey. Unlike the case where prey is scarce, the distance between the center μ_f of prey distribution and the center $\mu^t_{\max}$ of the model estimated by the BIB agent, and the variance $\Sigma^t_{\max}$ of the model does not approach zero over time. Therefore, when prey is abundant, unlike when prey is scarce, foraging behavior is not classified into two stages.

Fig 4. Differences in the movement patterns of BIB agents when prey is scarce and when it is abundant. (a) Time evolution of the step length when prey is scarce. (b) Time evolution of the step length when prey is abundant. (c) Complementary cumulative distribution function (CCDF) of the step length when prey is scarce. This figure also shows the results of fitting a TP model (green) and an EP model (red) to the simulation data. The exponent of the TP is η = 1.83. The fitting range is [0.098, 10.70]. (d) CCDF of the step length when prey is abundant. The exponent of the EP is λ = 1.85. The fitting range is [0.43, 6.24]. (e) Autocorrelation of the step-length time series when prey is scarce. (f) Autocorrelation of the step length time series when the prey is abundant.

Fig 4(a) and 4(b) show an example of the time evolution of the step length $l_t = \|d_t - d_{t-1}\|_2$ when prey is scarce and abundant, respectively. Fig 4(c) and 4(d) show the CCDF of the step length l_t when prey is scarce and abundant, respectively. These figures also show the results of fitting a truncated power-law distribution (TP) model (green) and an EP model (red) to the simulation data. Of the step length data, only the data within the range $\hat{l}_{\min} \le l_t \le \hat{l}_{\max}$ are used to improve the fitting as much as possible. See the Supplementary Information (SI) for details on how the fitting range $\left[\hat{l}_{\min}, \hat{l}_{\max}\right]$ and fitting method are determined. The CCDF graph is such that the CCDF value when $l = \hat{l}_{\min}$, is set to 1.

Fig 4(e) and 4(f) show the time dependence of the step length when prey is scarce and abundant, respectively. The horizontal axis represents the time lag τ, and the vertical axis represents the autocorrelation between the two time series l_t and $l_{t+\tau}$ of step length. As can be seen from these figures, when prey is scarce, there is a strong time dependence in the temporal variation of the step length, and the frequency distribution of the step length becomes a TP. Further, when prey is abundant, there is no strong time dependence and the frequency distribution of the step length becomes an EP.

Fig 5. Number of trials for which the frequency distribution of the step lengths was classified as TP and EP for each of the prey scarcity and abundance cases. The trials were run 100 times each with different random seeds. (a) BIB agent. (b) Sequential discounting expectation-maximization (SDEM) agent. (c) BID agent.

Fig 6. Comparison of the search performance of BIB agent with BID agent and SDEM agent; for SDEM agents, simulations were performed using three different variance-covariance matrices. The expected reward is the average value of 1000 trials run with different random seeds. (a) The time evolution of the expected reward for each agent. (b) The relationship between the average expected reward in the initial 1000 steps (1–1000 step) and the average expected reward in the final 1000 steps (9001–10000 step) for each agent. The regression line for the data of the three SDEM agents is also included in the figure.

Fig 5 shows the number of times the frequency distribution of the step length has been classified as TP and EP, as the results of 100 simulation trials with different random seeds for each of the cases of scarcity and abundance of prey. Fig 5(a), (b), and (c) represent the results of BIB agents that perform both learning and inference, SDEM agents that perform only learning, and BID agents that perform only inference, respectively. For all agents, the initial value of the variance-covariance matrix of the model is set to $\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$ and the discount factor as $\beta = 0.03$.

For BIB agents, when prey was scarce, the frequency distribution of the step length was classified as TP 93 times and EP 7 times, of the total 100 trials. Furthermore, when prey was abundant, the frequency distribution of the step length was classified as TP 5 times and EP 95 times. In both cases, the results of the binomial test showed significant differences between the number of occurrences of TP and EP ($p < 2.2\times10^{-16}$ and $p < 2.2\times10^{-16}$, respectively). The result of Fisher's exact test showed a significant relationship between

the amount of prey and the type of frequency distribution of step length ($p < 2.2\times10^{-16}$). The mean ± standard deviation of the power-law exponent for the 93 trials classified as TP when prey was scarce was 1.88 ± 0.41. Of the 93 trials, three (3.04, 3.29, and 3.27) had exponents outside the range $1 < \eta \leq 3$.

For SDEM agents, when prey was scarce, the frequency distribution of the step length was classified as TP 16 times and EP 84 times, of the total 100 trials. Further, when prey was abundant, the frequency distribution of the step length was classified as TP 8 times and EP 92 times. In both cases, the results of the binomial test showed significant differences between the number of occurrences of TP and EP ($p = 2.6\times10^{-12}$ and $p < 2.2\times10^{-16}$, respectively). For BID agents, when prey was scarce, the frequency distribution of the step length was classified as TP 19 times and EP 81 times, of the total 100 trials. Furthermore, when prey was abundant, the frequency distribution of the step length was classified as TP 0 times and EP 100 times. In both cases, the results of the binomial test showed significant differences between the number of occurrences of TP and EP ($p = 2.7\times10^{-10}$ and $p < 2.2\times10^{-16}$, respectively). Therefore, for SDEM and BID agents, regardless of the amount of prey, the frequency distribution of step lengths is often classified as EP and rarely as TP.

Next, a comparison of the search performance of BIB, BID, and SDEM agents is presented. Fig 6(a) shows the time evolution of the expected reward obtained by each agent when prey is scarce. However, these values represent the average of 1000 trials with different random seeds. Fig 6(b) shows the relationship between the average expected reward in the initial 1000 steps (1–1000 step) and the average expected reward in the final 1000 steps (9001–10000 step) for each agent. For the BIB and BID agents, the initial value of the variance-covariance matrix of the model was set to $\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$. For the SDEM agents, the initial values of the variance-covariance matrix of the model were set to $\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$, $\begin{pmatrix} 5.0 & 0 \\ 0 & 5.0 \end{pmatrix}$, and $\begin{pmatrix} 10.0 & 0 \\ 0 & 10.0 \end{pmatrix}$. The discount factor was set as $\beta = 0.03$, for all agents. The details of each agent model and simulation setup can be found in the Methods section. For SDEM agents, the higher the variance, the worse is the start-up and the higher is the final expected reward. In other words, there is a trade-off between the speed and accuracy of

search. This is because the larger the variance, the wider the area to be searched and the less likely it is to be overlooked, nonetheless, the longer it takes to converge the mean and variance. However, the BID agent has an appropriate start; nonetheless the final expected reward is lower because the hypothesis is not revised. It can be seen that the BIB agent is both faster and more accurate in its search than the SDEM and BID agents.

Case of five patchy prey distributions

Fig 7. Example of foraging behavior of BIB agent in the case of five prey patches. (a) Movement pattern. In the figure, two foraging points at adjacent time points are connected by a straight line. (b) Time evolution of the hypothesis with the highest confidence. (c) Time evolution of the x-coordinate of the foraging point. (d) Time evolution of the y-coordinate of the foraging point.

Fig 7 shows an example of the foraging behavior of BIB agents when there are five patchy prey distributions (Fig 1(c)). Fig 7(a) plots the foraging points d^t at each time point in a two-dimensional space. Fig 7(b) shows the time evolution of the hypothesis $h^t_{\max}$ with the highest confidence. Fig 7(c) and 7(d) show the time evolution of the x- and y-coordinates of the foraging point $d^t = \left(d^t_x, d^t_y\right)$, respectively. The central coordinates of the prey distribution in the five patches were (1.67, 4.66), (-1.93, -2.60), (3.37, 2.00), (4.98, 0.068), and (4.01, 2.39), respectively. These figures show that, similar to that in the case of one patchy prey distribution, the foraging point moves closer to the center of a patch; however, as food consumption reduces the food supply, the BIB agent moves to another patch.

Fig 8. Example of the CCDF of the duration of the hypothesis with the highest confidence when there are five prey patches. This figure also shows the results of fitting a TP model (green) and an EP model (red) to the simulation data. The exponent of TP is η = 1.98. The fitting range is [4, 2198].

Fig 8(a) shows an example of the CCDF for the duration of $h^t_{\max}$, that is, the length of time from the time a hypothesis becomes the hypothesis with the highest confidence until the time when it is superseded by another hypothesis. As shown in Fig 4(c) and 4(d), this figure is also accompanied by the results of fitting the TP and EP to the data. It can be seen from the figure that the frequency distribution of the durations follows a TP. After 100 trials of simulation with different random seeds, the frequency distribution of the durations was classified as TP in all 100 trials. For the 100 trials classified as TP, the mean ± standard deviation of the power-law exponent of TP was 2.04 ±0.23.

Discussion

First, a novel model of BIB inference is proposed wherein knowledge learning and knowledge-based inference are performed simultaneously. The learning was modeled using SDEM, a type of online EM algorithm, and the inference was modeled using BID, a Bayesian inference with discount factor. Second, foraging simulations were conducted using a decision-making agent incorporating the BIB inference model.

The following are the three results of the simulations. First, the BIB agents that perform simultaneous learning and inference can acquire prey relatively more efficiently than the agents that perform only inference (Bayesian inference) or only learning (inverse Bayesian inference). Second, when prey is abundant, the frequency distribution of the step length (that is, the distance between the foraging point at one time and the foraging point the next time) of BIB agents follows an EP, whereas when prey is scarce, the distribution becomes a TP. In other words, from a single model of BIB inference, various patterns of migratory behavior with TP or EP emerge, depending on the difficulty of obtaining food. This result is consistent with the finding that a few marine organisms switch between Brownian and Lévy walks depending on the amount of prey [4,12]. Third, when prey is scarce, there is a strong autocorrelation or time-dependence in the temporal

variation of the step length. Along with the second finding, this implies that the frequency distribution of the step length tends to follow a power-law distribution when there is a temporal dependence of the change in step length. This result is consistent with the finding [17] that there is a relationship between the dependence of the step length on time and the fact that the frequency distribution of the step length follows a power-law distribution.

When prey is scarce, the search behavior of BIB agents is divided into two modes: a rough search to select the best hypothesis and a detailed search designed by revising the hypothesis model after the hypothesis is fixed. This switch is triggered by the acquisition of food and may be considered a model of ARS.

When prey is scarce, BIB agents show extremely long steps in the initial hypothesis search mode; however, after fixing the hypothesis, the variance of the estimated model converges to zero in the hypothesis correction mode; accordingly, the step length approaches zero. Therefore, the range of the observed step lengths is wide. It is interesting to note here that although the foraging points are sampled from a normal distribution, the frequency distribution of the distances between foraging points, that is, the step lengths, results in a TP.

In general, the EM algorithm or Bayesian inference attempts to accurately estimate the parameters (mean and variance) of the generating distribution that produces the observed data. It is important to note that the observed data are usually sampled from the generating distribution; therefore, regardless of how the EM algorithm or Bayesian inference estimates the generating distribution, it has no effect on the observed data.

However, in the simulation task of this study, foraging points are determined based on the model estimated by the agent; this implies that the observed data depends on the distribution that the agent estimates. This can lead to the variance of the model estimated by the BIB agent approaching zero, even though the variance of the actual prey distribution is not zero. The structure of this task is similar to that of the multi-armed bandit problem [30].

In the case of prey abundance, the time dependence is weakened and the frequency distribution of the step length becomes an EP. In the case of prey abundance, it is not necessary to cause the center of the estimated model to coincide with the center of the prey distribution and to converge the variance of the

estimated model to zero, because prey can be obtained with a high probability without foraging at the center of the prey distribution. For this reason, short step lengths do not tend to appear. In addition, because there is a certain probability of acquiring prey regardless of which hypothesis is chosen at the initial stage, there tends to be less need for hypothesis search and therefore fewer large movements. Therefore, the range of observed step lengths is narrower in situations where prey is abundant. Thus, it can be noted that depending on the difficulty of prey acquisition, various migratory behavioral patterns emerge. Interestingly, these patterns emerge from a single decision-making model: BIB inference.

In the case of multiple patches, BIB agents move to another patch when the amount of food in one patch decreases and the reward earned becomes smaller. It is also shown that the frequency distribution of the duration of the hypothesis becomes a TP.

The exploration and preferential return (EPR) model incorporates the random walk process of exploration and the human propensity to revisit places (preferential return). In EPR, one of these two competing mechanisms is chosen probabilistically each time [31]. Exploration and exploitation are two essential components of any optimization algorithm [32]. Determining an appropriate balance between exploration and exploitation is the most challenging task in the development of any meta-heuristic algorithm, owing to the randomness of the optimization process [33]. In the present simulations, two patterns emerged: a rough search using knowledge (hypotheses) and a detailed search where the target was narrowed down by modifying the knowledge. The former may be associated with exploitation, whereas the latter with exploration. It is interesting to note that the emergence of these behavioral patterns in BIB agents is not explicitly controlled by any parameters; they can occur automatically depending on the situation.

The limitation of this study is that the present simulations address only the case of a single cluster with respect to the mixture distribution of the SDEM. In the future, the authors intend to perform simulations using a generative distribution comprising several normal distributions. In the present simulation, the authors regard the generation distribution and density distribution of prey in each patch as a normal distribution. In the future, the authors intend to simulate a case wherein the density distribution of the prey inside the patch is

uniform or non-parametric.

The authors have shown that when a BIB agent is placed in a severe environment, i.e., when it has to take its search behavior more seriously, the temporal variation of step length becomes autocorrelated and the frequency distribution of step length becomes a power-law distribution. In the future, it is needed to clarify whether the temporal variation of step length is time-dependent in animals other than humans, and if so, whether it is correlated with the form of the frequency distribution. Furthermore, the fact that these features appeared only in a BIB agent means that simultaneous learning of knowledge and inference based on knowledge in exploratory behavior under incomplete information may be the key to the emergence of Levy-Walk-like behavioral patterns. This point needs to be clarified experimentally in the future.

Methods

SDEM algorithm

The EM algorithm [28] is a machine-learning technique used for unsupervised clustering. The classification of a given set of N data $\left\{d^1,\cdots,d^i,\cdots d^N\right\}$ into K clusters is considered here. First, a mixture normal distribution comprising K normal distributions is considered for generating data d.

$$P(d)=\sum_{k=1}^{K}\pi_k N(d\mid\mu_k,\Sigma_k) \qquad (1)$$

Here, d is an m-dimensional data vector and $N(d\mid\mu_k,\Sigma_k)$ is an m-dimensional normal distribution with mean μ_k and variance-covariance matrix Σ_k, as shown below.

$$N(d\mid\mu_k,\Sigma_k)=\frac{1}{\sqrt{(2\pi)^m\left|\Sigma_k\right|}}\exp\left(-\frac{1}{2}(d-\mu_k)^{\mathrm{T}}\Sigma_k^{-1}(d-\mu_k)\right) \qquad (2)$$

Here, $\sum_{k=1}^{K}\pi_k=1$ is assumed. π_k is called the “mixing weights” and represents the weight of each

normal distribution. The aim of the EM algorithm is to estimate the mixing weights π_k, mean μ_k, and variance Σ_k of K normal distributions from N data $\left\{d^1,\cdots,d^i,\cdots d^N\right\}$. First, in Step E, the responsibility is calculated. The responsibility for the normal distribution *k* of data d^i is calculated as follows:

$$q_k^i = \frac{\pi_k N\left(d^i \mid \mu_k, \Sigma_k\right)}{\sum_{j=1}^{K} \pi_j N\left(d^i \mid \mu_j, \Sigma_j\right)} \qquad (3)$$

Next, the weights of data d^i with respect to the normal distribution *k* are defined as follows:

$$w_k^j = \frac{q_k^i}{\sum_{j=1}^{N} q_k^j} \qquad (4)$$

Next, in Step M, the mixing weights, mean, and variance of each normal distribution are updated.

$$\pi_k \leftarrow \frac{\sum_{i=1}^{N} q_k^i}{N} \qquad (5)$$

$$\mu_k \leftarrow \sum_{i=1}^{N} w_k^j d^i \qquad (6)$$

$$\Sigma_k \leftarrow \sum_{i=1}^{N} w_k^j \left(d^i - \mu_k\right)\left(d^i - \mu_k\right)^{\mathrm{T}} \qquad (7)$$

In the EM algorithm, the values converge by alternately repeating Step E and M. However, in the EM algorithm, it is necessary to provide all observational data simultaneously. In practice, there are cases where data processing must be performed sequentially, each time after the data are observed. Various online algorithms have been proposed to address this situation [34–41]. For example, Yamanishi et al. [29] proposed an SDEM. In SDEM, Step E and M were executed once for each data observed sequentially. In addition, the SDEM algorithm introduces the effect of forgetting to deal dealing with unsteady situations where the inferred target changes. Specifically, weighting is performed to weaken the influence of the older observation data by introducing the discount rate β (0<β<1). When data d^t are observed at time t, the mixing weights, mean, and variance of each normal distribution are updated, as shown below.

$$\tilde{\pi}_k^{t+1} \leftarrow (1-\beta)\tilde{\pi}_k^t + \beta q_k^t = \beta \sum_{i=1}^{t} (1-\beta)^{t-i} q_k^i \quad (8)$$

$$\tilde{\mu}_k^{t+1} \leftarrow (1-\beta)\tilde{\mu}_k^t + \beta q_k^t d^t = \beta \sum_{i=1}^{t} (1-\beta)^{t-i} q_k^t d^i \quad (9)$$

$$\tilde{\Sigma}_k^{t+1} \leftarrow (1-\beta)\tilde{\Sigma}_k^t + \beta q_k^t \left(d^t - \mu_k^t\right)\left(d^t - \mu_k^t\right) = \beta \sum_{i=1}^{t} (1-\beta)^{t-i} q_k^t \left(d^i - \mu_k^t\right)\left(d^i - \mu_k^t\right) \quad (10)$$

$$\pi_k^{t+1} \leftarrow \frac{\tilde{\pi}_k^{t+1}}{\sum_{j=1}^{K} \tilde{\pi}_j^{t+1}} \quad (11)$$

$$\mu_k^{t+1} \leftarrow \frac{\tilde{\mu}_k^{t+1}}{\tilde{\pi}_k^{t+1}} \quad (12)$$

$$\Sigma_k^{t+1} \leftarrow \frac{\tilde{\Sigma}_k^{t+1}}{\tilde{\pi}_k^{t+1}} \quad (13)$$

In this study, for simplicity, the case of a single normal distribution, that is, K=1, is treated as the generating distribution of the data. In the case of K=1, q_1^t and $\tilde{\pi}_1^t$ are always 1. For simplicity, μ_1^t and Σ_1^t are denoted as μ^t and Σ^t, respectively. At this time, Formula (12) and (13) show that the new estimated value is obtained as a convex combination of the current estimated value and current observed data, as follows:

$$\mu^{t+1} \leftarrow (1-\beta)\mu^t + \beta d^t \quad (14)$$

$$\Sigma^{t+1} \leftarrow (1-\beta)\Sigma^t + \beta\left(d^t - \mu^t\right)\left(d^t - \mu^t\right)^{\mathrm{T}} \quad (15)$$

These represent the exponential moving averages of μ^t and Σ^t.

Bayesian inference with discount rate (BID)

In Bayesian inference, first, several hypotheses (generative models) are prepared for an inferred target. Next, the evidence (the data generated by the target) is observed and the confidence in each hypothesis

is updated by quantitatively evaluating the fit between each hypothesis and the observed data. Finally, based on the confidence level, the single best hypothesis is narrowed down.

In general, it is better to obtain as much observational data (information) as possible for the accurate estimation of the target. However, this is true only in a stationary environment. When the target varies dynamically, it is necessary to discard the distant past observation data for a more accurate estimation. To address this situation, BID is introduced here [26,27].

First, a discrete version of Bayesian inference is described. In Bayesian inference, first a number of hypotheses h_k are defined and a model (the generating distribution of the data) for each hypothesis is prepared in the form of a conditional probability $P(d \mid h_k)$. Specifically, d is assumed to be an m-dimensional vector, as in the case of SDEM, and the model $P(d \mid h_k)$ is assumed to be an m-dimensional normal distribution, as follows:

$$P(d \mid h_k) = N(d \mid \mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^m |\Sigma_k|}} \exp\left(-\frac{1}{2} (d - \mu_k)^{\mathrm{T}} \Sigma_k^{-1} (d - \mu_k) \right) \qquad (16)$$

This conditional probability is called the likelihood if the data are fixed and considered as a function of the hypothesis h_k, that is, the parameter $\theta_k = (\mu_k, \Sigma_k)$ of the normal distribution. An initial confidence $P^1(h_k)$ value for each hypothesis is provided as a prior probability. If the confidence in hypothesis h_k at time t is $P^t(h_k)$ and data d^t are observed, the posterior probability $P^t(h_k \mid d^t)$ is calculated using Bayes' theorem, as follows:

$$P^t(h_k \mid d^t) = \frac{P^t(h_k) P(d^t \mid h_k)}{P^t(d^t)} \qquad (17)$$

where $P^t(d^t)$ is the marginal probability of the data at time t and is defined, as follows:

$$P^t(d^t) = \sum_k P^t(h_k) P(d^t \mid h_k) \qquad (18)$$

The following Bayesian update accordingly replaces the posterior probability with the confidence the next

time.

$$P^{t+1}\left(h_k\right) \leftarrow P^t\left(h_k \mid d^t\right) \qquad (19)$$

Combining Equation (17) and (19),

$$P^{t+1}\left(h_k\right) \leftarrow \frac{P^t\left(h_k\right) P\left(d^t \mid h_k\right)}{P^t\left(d^t\right)} \qquad (20)$$

The estimation proceeds by updating the confidence in each hypothesis by Equation (20), each time the data are observed. Here, the confidence in each hypothesis satisfies $\sum_k P^{t+1}\left(h_k\right)=1$. Focusing on the recursive nature of $P^t\left(h_k\right)$, Equation (20) can be rewritten as follows:

$$P^{t+1}\left(h_k\right) \leftarrow P^1\left(h_k\right) \prod_{i=1}^{t} \frac{P\left(d^i \mid h_k\right)}{P^i\left(d^i\right)} \qquad (21)$$

Here, $P^i\left(d^i\right)$ is common to all hypotheses and can be regarded as a constant. Omitting the normalization process, Equation (21) can be written as follows:

$$P^{t+1}\left(h_k\right) \leftarrow P^1\left(h_k\right) \prod_{i=1}^{t} P\left(d^i \mid h_k\right) \qquad (22)$$

That is, the current confidence of each hypothesis is proportional to the prior probability multiplied by the likelihood of the data observed each time.

Next, the function of forgetting to deal with unsteady situations is incorporated by introducing a discount factor into the Bayesian inference, as follows [26,27]. However, in the following, C is used instead of P, to distinguish between Bayesian inference and Bayesian inference, by introducing a discount factor. If the normalization process is omitted, the updating formula for the confidence of each hypothesis is defined as follows.

$$C^{t+1}\left(h_k\right) \leftarrow \left[\frac{C^t\left(h_k\right)}{C^t\left(d^t\right)}\right]^{1-\beta} C\left(d^t \mid h_k\right) \qquad (23)$$

where β is the discount factor and Equation (23) agrees with Equation (20) when β = 0. After Equation (23),

normalization is conducted as follows, to satisfy $\sum_k C^{t+1}(h_k)=1$.

$$C^{t+1}(h_k) \leftarrow \frac{C^{t+1}(h_k)}{\sum_j C^{t+1}(h_j)} \qquad (24)$$

Focusing on the recurrence of $C^t(h_k)$, Equation (23) can be transformed as follows:

$$C^{t+1}(h_k) \leftarrow \left[C^1(h_k)\right]^{(1-\beta)^t} \prod_{i=1}^{t} \frac{\left[C(d^i \mid h_k)\right]^{(1-\beta)^{t-i}}}{\left[C^i(d^i)\right]^{(1-\beta)^{t+1-i}}} \qquad (25)$$

In Equation (25), the denominator $C^i(d^i)$ on the right-hand side is common to each hypothesis and can be regarded as a constant; therefore, omitting the normalization process, the following equation is obtained:

$$C^{t+1}(h_k) \leftarrow \left[C^1(h_k)\right]^{(1-\beta)^t} \prod_{i=1}^{t} \left[C(d^i \mid h_k)\right]^{(1-\beta)^{t-i}} \qquad (26)$$

In other words, it is understood that the current confidence is multiplied by the likelihood in the distant past with less weight. When β = 0, using $C^{t+1}(h_k) \leftarrow C^1(h_k)C(d^1 \mid h_k)C(d^2 \mid h_k)\cdots C(d^t \mid h_k)$ in the Bayesian inference, the current likelihood and the likelihood in the distant past are evaluated with equal weight. When β = 1, $C^{t+1}(h_k) \leftarrow C(d^t \mid h_k)$ is used to calculate the next confidence using only the likelihood of the current observation. In the BID, the confidence for each hypothesis is updated according to Equation (23), each time data is observed, and the best hypothesis $h^t_{\max}$ is determined based on the confidence level.

$$h^t_{\max} = \arg\max_{h_i} C^t(h_i) \qquad (27)$$

BIB inference

SDEM and BID are already explained. If the generative model is regarded as knowledge, SDEM can be regarded as a type of learning because knowledge is formed and modified based on the observed data. DBI, however, can be called knowledge-based inference because it decides which of the several hypotheses

(generative models) prepared in advance is the best.

In this section, a model of BIB inference is proposed, that employs SDEM and BID to perform learning and inference simultaneously. Here, BID and SDEM are the models for Bayesian inference and inverse Bayesian inference, respectively. This is a model where the hypotheses are revised based on inverse Bayesian inference, while Bayesian inference is used to access which hypothesis is correct. Because the model $C(d \mid h_k)$ of the hypothesis h_k changes with time through inverse Bayesian inference, it is denoted by $C^t(d \mid h_k)$, with time t as a subscript. First, data d^t at time t is observed, and subsequently, the confidence of each hypothesis is updated using the BID formula.

$$C^{t+1}(h_k) \leftarrow \left[\frac{C^t(h_k)}{C^t(d^t)}\right]^{1-\beta} C^t(d^t \mid h_k) \qquad (28)$$

Next, the hypothesis $h^t_{\max}$ with the highest confidence is selected using Equation (27). SDEM is thereafter used to update the model $C^{t+1}(d \mid h^t_{\max}) = N(d \mid \mu^{t+1}_{\max}, \Sigma^{t+1}_{\max})$, for the following hypothesis $h^t_{\max}$, as follows.

$$\mu^{t+1}_{\max} \leftarrow (1-\beta)\mu^t_{\max} + \beta d^t \qquad (29)$$

$$\Sigma^{t+1}_{\max} \leftarrow (1-\beta)\Sigma^t_{\max} + \beta(d^t - \mu^t_{\max})(d^t - \mu^t_{\max})^{\mathrm{T}} \qquad (30)$$

This is repeated every time data is observed. If $K = 1$ is set, $C^t(h_1) = 1$ is always obtained because there is only one hypothesis. This implies that no Bayesian inference is performed and only inverse Bayesian inference is performed employing SDEM. However, if SDEM is not employed and the model of the hypothesis $h^t_{\max}$ is not modified, then only BID is implemented.

Foraging simulation

Case of single patchy prey distribution

Consider a situation in which the preys are distributed in a restricted two-dimensional space (Fig 1(a) and (b)). The density distribution of the prey at a certain point $d = \left(d_x, d_y\right)$ is defined as follows:

$$\Phi\left(d \mid \mu_f, \Sigma_f\right) = \exp\left(-\frac{1}{2}\left(d-\mu_f\right)^{\mathrm{T}} \Sigma_f^{-1}\left(d-\mu_f\right)\right) \quad (31)$$

The distribution takes a maximum value when the condition $d = \mu_f$ is satisfied, and the value is $\Phi\left(d \mid \mu_f, \Sigma_f\right) = \Phi\left(\mu_f \mid \mu_f, \Sigma_f\right) = 1$. As moving away from the center μ_f of the distribution, the density of prey decreases. Σ_f is a variance parameter, and the larger $\left|\Sigma_f\right|$ is, the more widely the prey is distributed and the larger is the total amount of prey.

Next, consider three foraging agents that perform BID, SDEM, and BIB inference. Each agent decides probabilistically which point $d^t = \left(d_x^t, d_y^t\right)$ to forage according to the model of the hypothesis $h_{\max}^t$, with the highest confidence at time t, as follows:

$$d^t \sim N\left(d \mid \mu_{\max}^t, \Sigma_{\max}^t\right) \quad (32)$$

Next, whether a prey is acquired is probabilistically determined based on the prey density at point d^t.

$$success^t = \begin{cases} 1 & if\ rnd \le \Phi\left(d^t \mid \mu_f, \Sigma_f\right) \\ 0 & else \end{cases} \quad (33)$$

where rnd represents a uniform random number in the [0,1] interval; $success^t = 1$ represents successful prey acquisition, and $success^t = 0$ represents unsuccessful prey acquisition. Each agent performs the estimation on the basis of the information on the location of the prey acquisition, only if the prey acquisition is successful. If the agent fails to acquire prey, it does not modify the confidence level of each hypothesis or the model of the hypothesis; nonetheless, it only changes the next foraging point according to Equation (32). The objective of the agent is to acquire as many preys as possible within a certain period of time, and it always succeeds in acquiring prey when it chooses the center of the prey distribution as its foraging point, that is, $d^t = \mu_f$ because $\Phi\left(d^t \mid \mu_f, \Sigma_f\right) = \Phi\left(\mu_f \mid \mu_f, \Sigma_f\right) = 1$. In other words, the objective of the agent is

to converge the parameters of its estimated model to $\mu_{\max}^{t} \to \mu_f$ and $\Sigma_{\max}^{t} \to 0$ as soon as possible.

The group of processes described in this section is summarized as an algorithm below.

1. Determine the parameters $\left(\mu_f, \Sigma_f\right)$ of the prey density distribution.
2. Set the values for parameters β, *K*.
3. Establish the initial values for $\theta_k^1 = \left(\mu_k^1, \Sigma_k^1\right), C^1\left(h_k\right) \quad (k = 1, 2, \cdots K)$.
4. Repeat the following whenever data d^t are observed.
   - 4.1 The agent chooses the hypothesis $h_{\max}^t$ it believes in the most, according to Equation (27).
   - 4.2 Based on the model of the hypothesis $h_{\max}^t$, the agent decides the foraging point d^t according to Equation (32).
   - 4.3 Equation (33) determines the success or failure of foraging.
   - 4.4 The agent executes the following processes simultaneously, only if it is successful in foraging:
     - 4.4.1 Bayesian inference (BID): The confidence $C^t\left(h_k\right)$ of each hypothesis is updated according to Equation (28).
     - 4.4.2 Inverse Bayesian inference (SDEM): The model $C^t\left(d \mid h_{\max}^t\right) = N\left(d \mid \mu_{\max}^t, \Sigma_{\max}^t\right)$ of the hypothetical $h_{\max}^t$ is modified according to Equation (29) and (30).

Case of multiple patchy prey distributions

In this section, a situation is considered wherein there are multiple prey patches. Specifically, a prey distribution is considered, such that there are P prey patches, as shown in Fig 1(c).

$$\Phi^t\left(d\right) = \sum_{p=1}^{P} w_p^f \Phi\left(d \mid \mu_f^p, \Sigma_f^p\right) \qquad (34)$$

where μ_f^p and Σ_f^p are the center and variance of patch p, respectively. w_p^f is a parameter that controls

the total amount of prey of patch p, at time t. As an initial value, it is set as $w_p^1 = 1$ for all p.

In the simulation task described in the previous section, prey distribution did not change even when the agent acquired prey. In this section, it is assumed that the agent consumes food every time it acquires, and the prey amount is set to decrease by Δ. If the agent acquires prey at the location d^t at time t, then w_p^t in each patch decreases according to the amount of prey at the foraging location d^t as follows:

$$w_p^{t+1} \leftarrow w_p^t - \frac{w_p^t \Phi\left(d^t \mid \mu_f^p, \Sigma_f^p\right)}{\sum_{j=1}^{P} w_j^t \Phi\left(d^t \mid \mu_f^j, \Sigma_f^j\right)} \Delta \qquad (35)$$

In the simulation task of the previous section, the agent updated the confidence of each hypothesis according to Equation (28) only when it succeeded in acquiring prey. In the task of this section, the agent also modifies its confidence when it fails to acquire prey, as follows:

$$C^{t+1}\left(h_k\right) \leftarrow \left[\frac{C^t\left(h_k\right)}{C^t\left(d^t\right)}\right]^{1-\beta} \left[1 - C^t\left(d^t \mid h_k\right)\right] \quad (36)$$

In other words, contrary to the successful case, the higher the likelihood of the hypothesis for d^t, the lower is the confidence level.

The group of processes described in this section and the previous section are summarized as an algorithm below.

1 Determine the number of patches P of the prey distribution and the parameters $\left(\mu_f^p, \Sigma_f^p\right)$ of each patch.

2 Set the values for parameters β, *K*.

3 Establish the initial values for $\theta_k^1 = \left(\mu_k^1, \Sigma_k^1\right), C^1\left(h_k\right) \quad (k = 1, 2, \cdots K)$.

4 Repeat the following whenever data d^t are observed.

4.1 The agent chooses the hypothesis $h_{\max}^t$ it believes in the most, according to Equation (27).

4.2 Based on the model of the hypothesis $h_{\max}^t$, the agent decides the foraging point d^t according to Equation (32).

4.3 Equation (33) determines the success or failure of foraging.

4.4 The following processes are conducted according to the results of foraging.

4.4.1 Success to forage:

4.4.1.1 The confidence $C^t\left(h_k\right)$ of each hypothesis is updated according to Equation (28).

4.4.1.2 The model $C^t\left(d \mid h_{\max}^t\right) = N\left(d \mid \mu_{\max}^t, \Sigma_{\max}^t\right)$ of the hypothetical $h_{\max}^t$ is modified according to Equation (29) and (30).

4.4.1.3 The prey distribution $\Phi^t\left(d\right)$ changes according to Equation (35).

4.4.2 Failure to forage:

4.4.2.1 The confidence $C^t\left(h_k\right)$ of each hypothesis is updated according to Equation (36).

Parameter setting

The number of hypotheses for the BIB and BID agents was set as $K = 20$. The number of hypotheses for the SDEM agent was set as $K = 1$. For the initial values of the model parameters for each hypothesis, the mean μ_k^1 was determined randomly in the range [-5,5] for both x and y coordinates. For the BIB and BID agents, the variance was set as $\Sigma_k^1 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$ for all the hypotheses. The variances of the SDEM agents were set as $\Sigma_k^1 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$, $\begin{pmatrix} 5.0 & 0 \\ 0 & 5.0 \end{pmatrix}$, and $\begin{pmatrix} 10.0 & 0 \\ 0 & 10.0 \end{pmatrix}$. The initial value of confidence (that is, prior probability) for each hypothesis was set as $C^1\left(h_k\right) = 1/K = 0.05$, such that they would be equal. For all agents, the discount rate was set as $\beta = 0.03$.

For the single patch, two cases were simulated: scarce prey and abundant prey. First, the central coordinates μ_f of prey distribution were randomly determined in the range [-5,5] for both the x and y

coordinates. The variance of the prey distribution was set as $\Sigma_f = \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix}$ when food was scarce (Fig 1(a)), while it was set as $\Sigma_f = \begin{pmatrix} 5.0 & 0 \\ 0 & 5.0 \end{pmatrix}$ when food was abundant (Fig 1(b)). In the case of multiple patches, the number of patches was set to 5. The center coordinates μ_f^p of each patch were randomly determined within the range [-5,5] for both the x and y coordinates (Fig 1(c)). The variance of each patch was set to $\Sigma_f^p = \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix}$. The amount of food loss when the agent acquired food was set to $\Delta = 0.001$. The simulation was run for up to 10000 steps.

The simulation program was developed based on the C++ language. The compiler was MinGW 8.1.0 64-bit for C++ [42]. The Qt library (Qt version Qt 5.15.2 MinGW 64-bit) was also used for development [43].

Acknowledgments

This research is partially supported by the Center of Innovation Program from the Japan Science and Technology Agency, JST. This work was supported by the JSPS KAKENHI Grant Numbers JP21K12009.

Data availability statement

The source code and data used to produce the results and analyses presented in this manuscript are available from Bitbucket Git repository:

https://bitbucket.org/ShujiShinohara/levyflightforagingsimulation1/src/master/

References

[1] Harris TH, Banigan EJ, Christian DA, Konradt C, Wojno EDT, Norose K, et al. Generalized Lévy walks and the role of chemokines in the migration of effector CD8+ T cells. Nature. 2012;486: 545–548. DOI: 10.1038/nature11098.

[2] Ariel G, Rabani A, Benisty S, Partridge JD, Harshey RM, Be'Er A. Swarming bacteria migrate by Lévy Walk. Nat Commun. 2015;6: 8396. DOI: 10.1038/ncomms9396.

[3] Shokaku T, Moriyama T, Murakami H, Shinohara S, et al.: Morioka . Development of an automatic turntable-type multiple T-maze device and observation of pill bug behavior. Rev Sci Instrum. 2020;91: 104104. DOI: 10.1063/5.0009531.

[4] Humphries NE, Queiroz N, Dyer JRM, Pade NG, Musyl MK, Schaefer KM, et al. The environmental context explains the Lévy and Brownian movement patterns of marine predators. Nature. 2010;465: 1066–1069. DOI: 10.1038/nature09116.

[5] Humphries NE, Weimerskirch H, Queiroz N, Southall EJ, Sims DW. Foraging success of biological Lévy flights recorded in situ. Proc Natl Acad Sci U S A. 2012;109: 7169–7174. DOI: 10.1073/pnas.1121201109.

[6] Raichlen DA, Wood BM, Gordon AD, Mabulla AZP, Marlowe FW, Pontzer H. Evidence of Lévy walk foraging patterns in human hunter–gatherers. Proc Natl Acad Sci U S A. 2014;111: 728–733. DOI: 10.1073/pnas.1318616111.

[7] Sims DW, Southall EJ, Humphries NE, Hays GC, Bradshaw CJA, Pitchford JW, et al. Scaling laws of marine predator search behavior Nature. 2008;451: 1098–1102. DOI: 10.1038/nature06518.

[8] Viswanathan GM, Raposo EP, da Luz MGE. Lévy flights and superdiffusion in the context of biological encounters and random searches Phys Life Rev. 2008;5: 133–150. DOI: 10.1016/j.plrev.2008.03.002.

[9] Bartumeus F, da Luz MGE, Viswanathan GM, Catalan J. Animal search strategies: A quantitative random-walk analysis. Ecology. 2005;86: 3078–3087. DOI: 10.1890/04-1806.

[10] Viswanathan GM, Buldyrev SV, Havlin S, Da Luz MGE, Raposo EP, Stanley HE. Optimization of the success of random searches Nature. 1999;401: 911–914. DOI: 10.1038/44831.

[11] Bartumeus F, Catalan J, Fulco UL, Lyra ML, Viswanathan GM. Optimizing the encounter rate in biological interactions: Lévy versus Brownian strategies. Phys. Rev. Lett. 2002 Mar 04;88(9):097901. DOI: 10.1103/PhysRevLett.88.097901.

[12] Sims DW, Humphries NE, Bradford RW, Bruce BD. Lévy flight and Brownian search patterns of free-ranging predators reflect different prey field characteristics. J Anim Ecol. 2012 Mar;81(2): 432-442. doi: 10.1111/j.1365-2656.2011.01914.x. Epub 2011 Oct 17.

[13] Ross C, Pacheco-Cobos L, et al.. A general model of forager search: Adaptive encounter-conditional heuristics outperform Lévy flights in the search for patchily distributed prey. J Theor Biol. 2018 Oct 14;455: 357-369. doi: 10.1016/j.jtbi.2018.07.031.

[14] Namboodiri VM, Levy JM, Mihalas S, Sims DW, et MG. Rationalizing spatial exploration patterns of wild animals and humans through a temporal discounting framework Proc Natl Acad Sci U S A. 2016;113: 8747–8752. DOI: 10.1073/pnas.1601664113.

[15] Ross CT, Winterhalder B. Evidence for encounter-conditional, area-restricted search in a preliminary study of Colombian blowgun hunters. PLOS ONE. 2018;13: e0207633. DOI: 10.1371/journal.pone.0207633.

[16] Hills TT, Kalff C, Wiener JM. Adaptive Lévy Processes and Area-Restricted Search in Human Foraging PLOS ONE. 2013;8(4): e60488. https://doi.org/10.1371/journal.pone.0060488.

[17] Wang XW, Han XP, Wang BH. Correlations and scaling laws for human mobility. PLOS ONE. 2014;9: e84954. DOI: 10.1371/journal.pone.0084954.

[18] Rhee I, Shin M, Hong S, Lee K, Kim SJ, Chong S. Levy-walk nature of human mobility. IEEE/ACM Trans Networking 2011;19: 630–643. DOI: 10.1109/TNET.2011.2120618.

[19] Zhao K, Musolesi M, Hui P, Rao W, et al.: Explaining the power-law distribution of human mobility through transportationmodality decomposition. Sci Rep. 2015;5: 9136. https://doi.org/10.1038/srep09136.

[20] Boyer D, Ramos-Fernández G, Miramontes O, Mateos JL, Cocho G, Larralde H, et al. Scale-free foraging by primates emerges from their interactions with complex environments. Proc Biol Sci. 2006;273: 1743–1750. DOI: 10.1098/rspb.2005.3462.

[21] Gunji YP, Shinohara S, Haruna T, Basios V. Inverse Bayesian inference as a key of consciousness featuring a macroscopic quantum logical structure. Biosystems. 2017;152: 44–65. DOI: 10.1016/j.biosystems.2016.12.003.

[22] Gunji YP, Murakami H, Tomaru T, Basios V. Inverse Bayesian inference in swarming behaviour of soldier crabs. Philos Trans A Math Phys Eng Sci. 2018;376. DOI: 10.1098/rsta.2017.0370.

[23] Gunji YP, Kawai T, Murakami H, Tomaru T, Minoura M, Shinohara S. Lévy Walk in Swarm Models Based on Bayesian and Inverse Bayesian Inference. Comput Struct Biotechnol J. 2021;19: 247–260. DOI: 10.1016/j.csbj.2020.11.045.

[24] Zaki SR, Nosofsky RM. A high-distortion enhancement effect in the prototype learning paradigm: Dramatic effects of category learning during the test. Mem Cognit. 2007;35: 2088–2096. DOI: 10.3758/bf03192940.

[25] Zhu X, Goldberg AB. Introduction to Semi-Supervised Learning Morgan Claypool: San Rafael, CA, USA. 2009.

[26] Shinohara S, Manome N, Suzuki K, Chung UI, Takahashi T, Gunji PY, et al. Extended Bayesian inference incorporating symmetry bias. Biosystems. 2020;190: 104104. DOI: 10.1016/j.biosystems.2020.104104.

[27] Shinohara S, Manome N, Suzuki K, Chung UI, Takahashi T, Okamoto H, et al. A new Bayesian causal inference method in nonstationary environments. PLOS ONE. 2020;15: e0233559. DOI: 10.1371/journal.pone.0233559.

[28] Bishop CM. Pattern Recognition and Machine Learning. Springer, New York, NY, USA. 2011.

[29] Yamanishi K, Takeuchi J, Williams G, Milne P. Online unsupervised outlier detection using finite mixtures with discounting learning algorithms. Data Min Knowl Discov. 2004;8: 275–300. https://doi.org/10.1023/B:DAMI.0000023676.72185.7c.

[30] Richard SS, Andrew GB. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, USA. 1998.

[31] Pappalardo L, Rinzivillo S, Simini F. Human mobility modeling: exploration and preferential return meet the gravity model. Procedia Comput Sci. 2016;83: 934–939. DOI: 10.1016/j.procs.2016.04.188.

[32] Emary E, Zawbaa HM. Impact of chaos functions on modern swarm optimizers. PLOS ONE. 2016;11: e0158738. DOI: 10.1371/journal.pone.0158738.

[33] Wu X, Zhang S, Xiao W, et al.. Exploration/exploitation tradeoff in the whale optimization algorithm. IEEE Access. 2019;7–125928. DOI: 10.1109/ACCESS.2019.2938857, 125919.

[34] Neal RM, Hinton GE. The EM algorithm justifies incremental, sparse, and other variants. In Jordan MI, editor. Learning in graphical models Dordrecht: Kluwer Academic Publishers; 1998. pp. 355–368. https://doi.org/10.1007/978-94-011-5014-9_12.

[35] Sato M, Ishii S. Online EM algorithm for the normalized Gaussian network. Neural Comput. 2000;12: 407–432. DOI: 10.1162/089976600300015853.

[36] Liang P, Klein D. Online EM for unsupervised models. Proceedings of Human Language Technologies 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2009. pp. 611–619.

[37] Cappe O, Moulines E. Online expectation–maximization algorithm for latent data models. J R Stat Soc B. 2009;71: 593–613. https://doi.org/10.1111/j.1467-9868.2009.00698.x.

[38] Dias JG, Cortinhal MJ. SKM algorithm: k-means algorithm for clustering sequential data. In Lect Notes Comput Sci Geffner H, Prada R, Machado AI, David N, editors. New York: Springer. 2008: 173–182. https://doi.org/10.1007/978-3-540-88309-8_18.

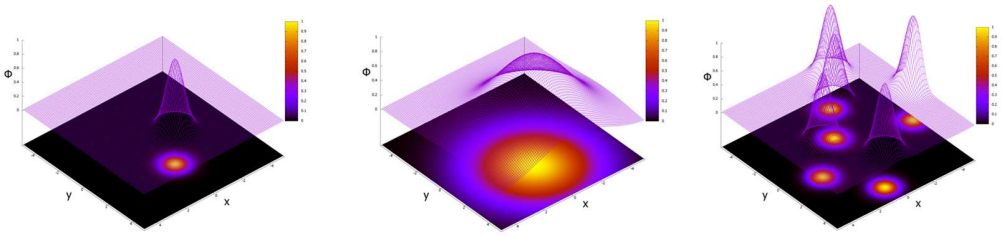
[39] Zhong S. Efficient online spherical k-means clustering. Proceedings of the IEEE International Joint Conference on Neural Networks 2005;5: 3180–3185. https://doi.org/10.1109/ijcnn.2005.1556436.

[40] Andrews NO, Fox EA. *Recent developments in document clustering*. Technical Report TR-07. Available at: https://vtechworks.lib.vt.edu/handle/10919/19473-35; 2007.

[41] Tehrani AF, Ahrens D. Modified sequential k-means clustering by utilizing response: A case study for fashion products. Expert Syst. 2017;34: e12226. https://doi.org/10.1111/exsy.12226.

[42] Available: http://www.mingw.org/. (Accessed 23 January 2021).

[43] Available: https://www.qt.io/. (Accessed 23 January 2021).



(a) **(b)** **(c)**

Fig 1 Distribution of prey density in a two-dimensional space. Prey is acquired probabilistically in proportion to the prey density at the foraging point. (a) When prey is scarce (the number of patches is one). (b) When the prey is abundant (the number of patches is one). (c) When the number of patches is five.

691

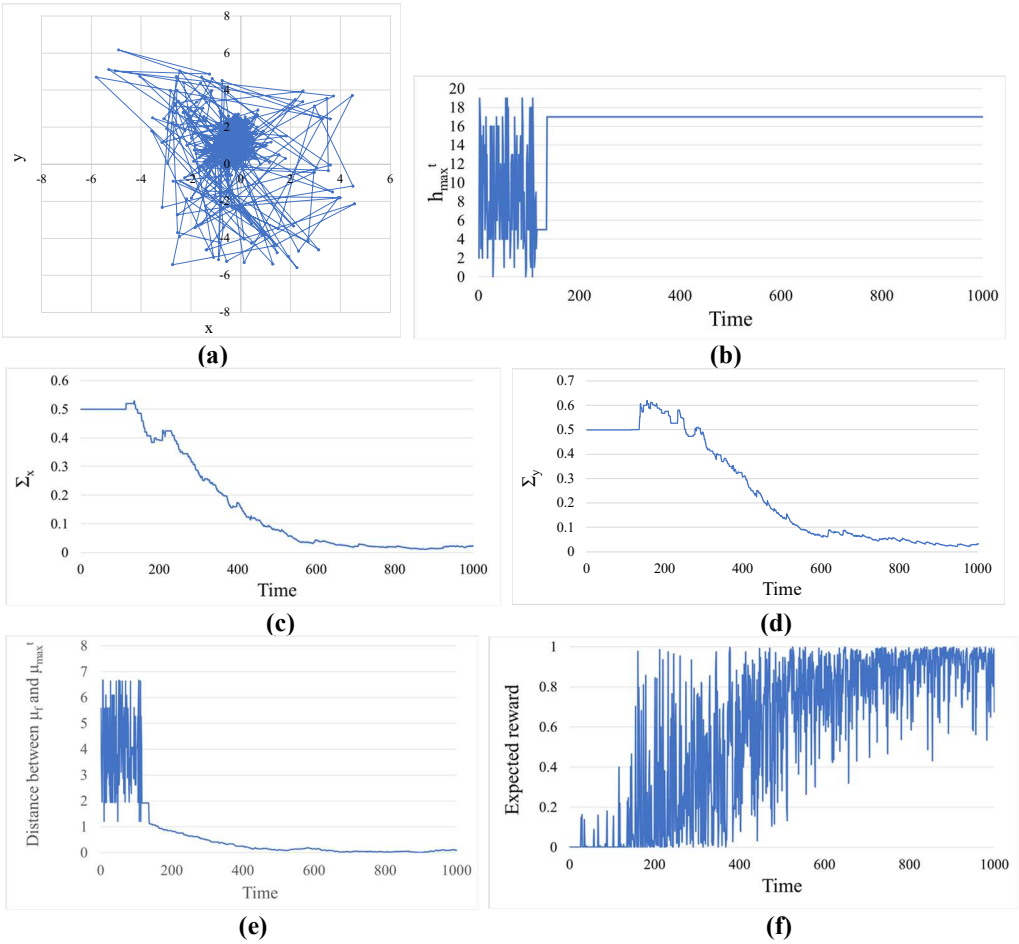


Fig 2 Example of foraging behavior of BIB agent when prey is scarce. (a) Movement pattern. Two foraging points at adjacent time points are connected by a straight line. (b) Time evolution of the hypothesis with the highest confidence. (c) Time evolution of the variance in the x-direction of the model for the hypothesis with the highest confidence. (d) Time evolution of the variance in the y-direction of the model for the hypothesis with the highest confidence. (e) Time evolution of distance between the center of the model for the hypothesis with the highest confidence and the center of the actual prey distribution. (f)Time evolution of the expected reward expressed by the prey density at the foraging point.

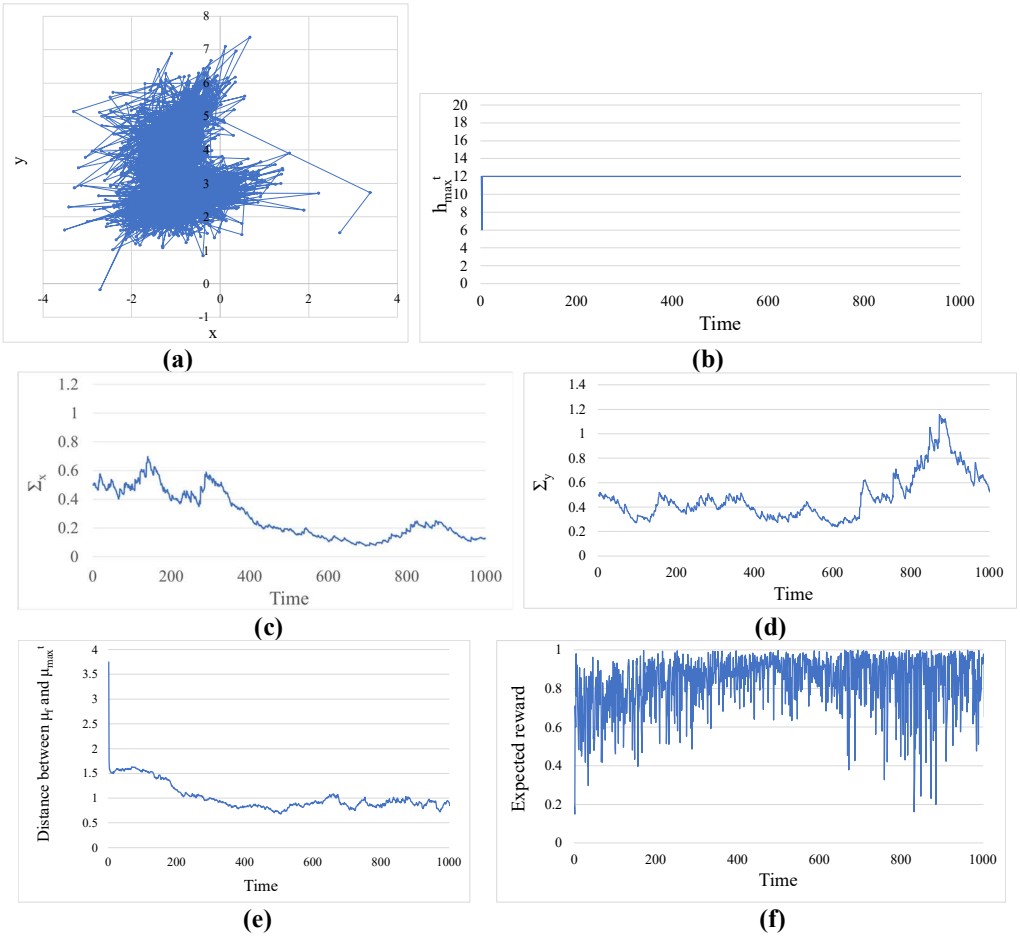


Fig 3. Example of foraging behavior of BIB agent when prey is abundant. (a) Movement pattern. In the figure, two foraging points at adjacent time points are connected by a straight line. (b) Time evolution of the hypothesis with the highest confidence. (c) Time evolution of the variance in the x-direction of the model for the hypothesis with the highest confidence. (d) Time evolution of the variance in the y-direction of the model for the hypothesis with the highest confidence. (e) Time evolution of distance between the center of the model for the hypothesis with the highest confidence and the center of the actual prey distribution. (f) Time evolution of the expected reward expressed by the prey density at the foraging point.

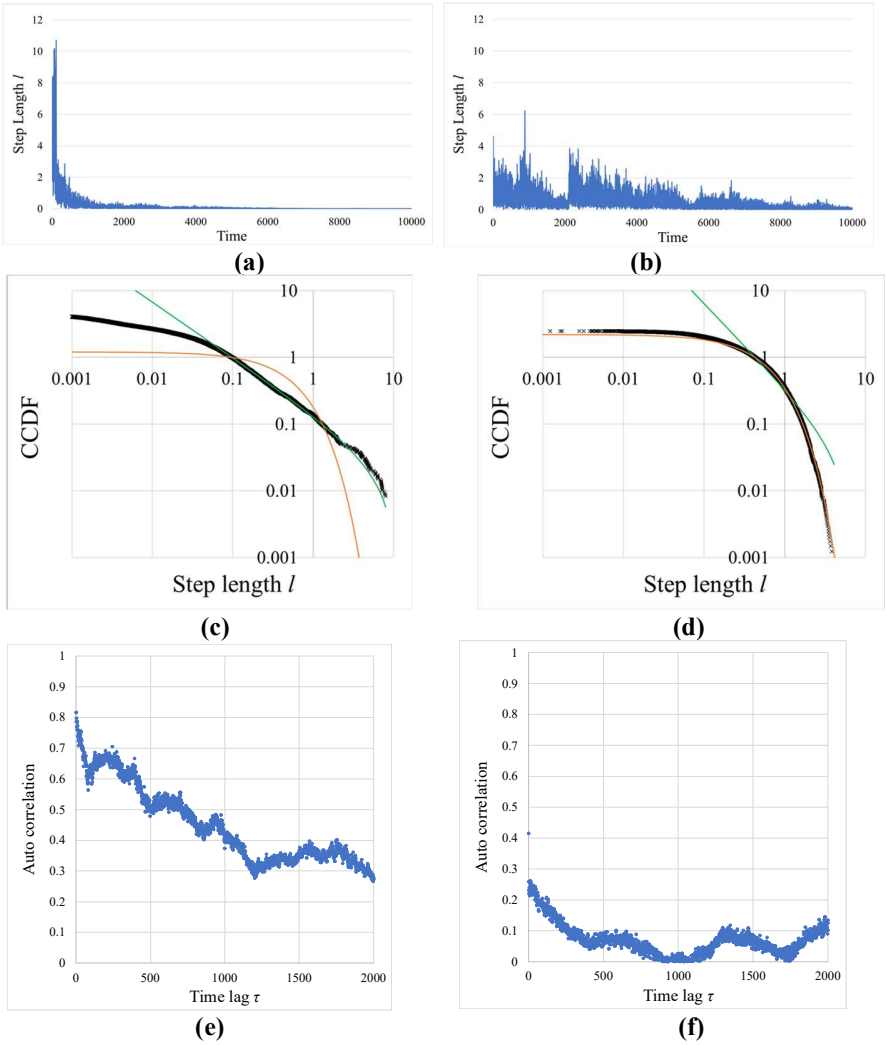


Fig 4. Differences in the movement patterns of BIB agents when prey is scarce and when it is abundant. (a) Time evolution of the step length when prey is scarce. (b) Time evolution of the step length when prey is abundant. (c) Complementary cumulative distribution function (CCDF) of the step length when prey is scarce. This figure also shows the results of fitting a TP model (green) and an EP model (red) to the simulation data. The exponent of the TP is $\eta = 1.83$. The fitting range is [0.098, 10.70]. (d) CCDF of the step length when prey is abundant. The exponent of the EP is $\lambda = 1.85$. The fitting range is [0.43, 6.24]. (e) Autocorrelation of the step-length time series when prey is scarce. (f) Autocorrelation of the step length time series when the prey is abundant.

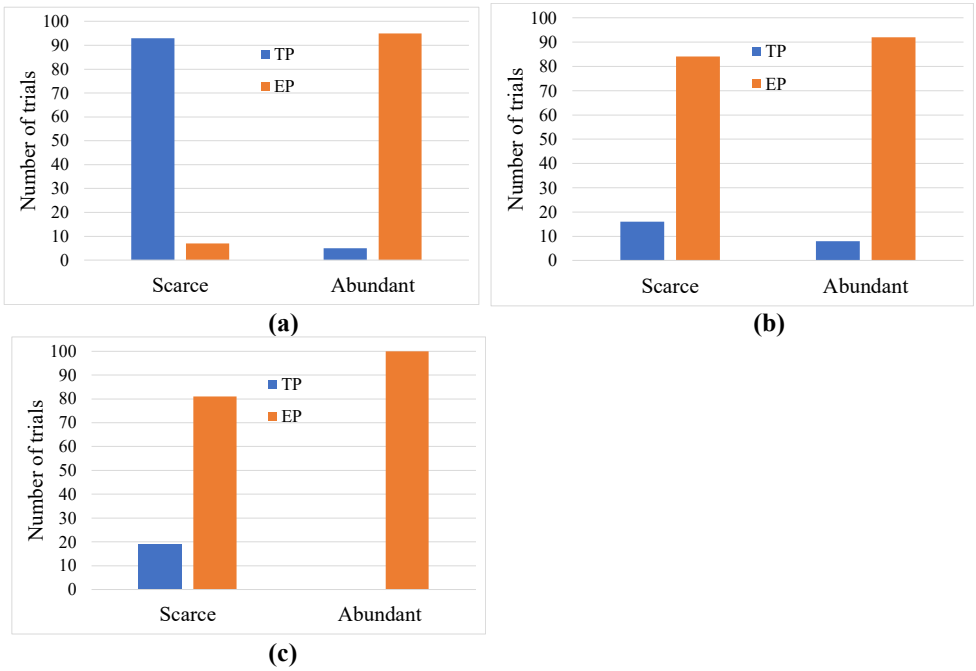
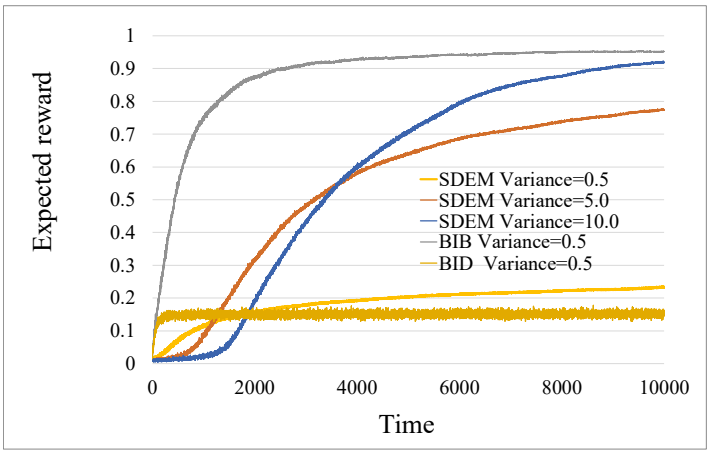
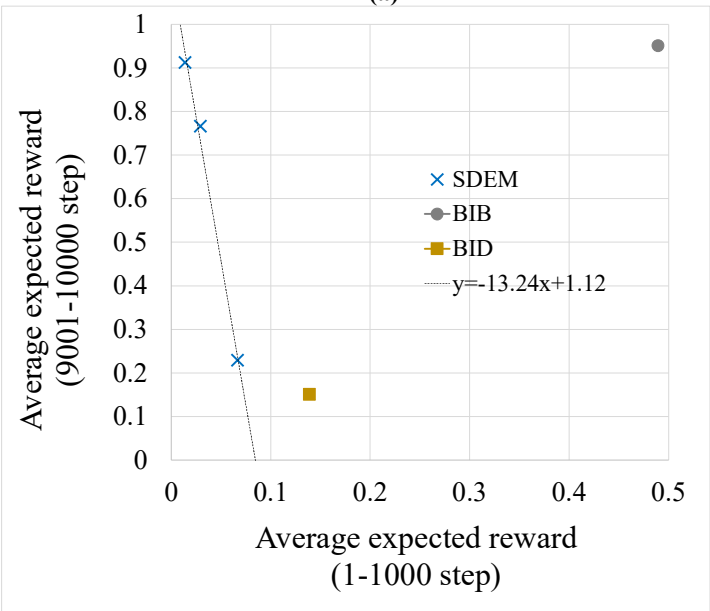


Fig 5. Number of trials for which the frequency distribution of the step lengths was classified as TP and EP for each of the prey scarcity and abundance cases. The trials were run 100 times each with different random seeds. (a) BIB agent. (b) Sequential discounting expectation-maximization (SDEM) agent. (c) BID agent.



(a)



(b)

Fig 6. Comparison of the search performance of BIB agent with BID agent and SDEM agent; for SDEM agents, simulations were performed using three different variance-covariance matrices. The expected reward is the average value of 1000 trials run with different random seeds. (a) The time evolution of the expected reward for each agent. (b) The relationship between the average expected reward in the initial 1000 steps (1–1000 step) and the average expected reward in the final 1000 steps (9001–10000 step) for each agent. The regression line for the data of the three SDEM agents is also included in the figure.

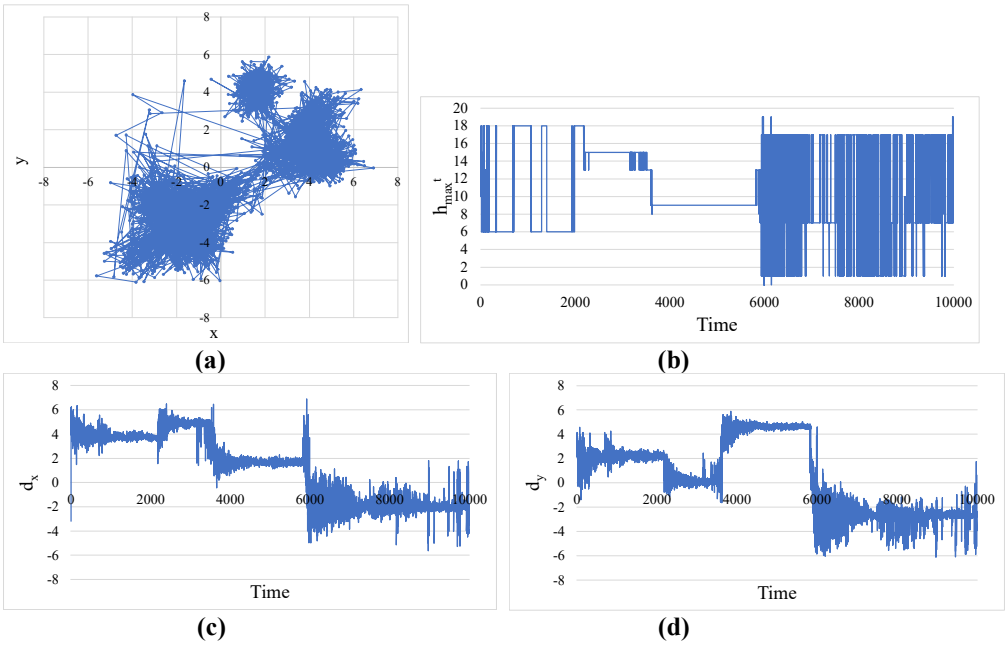


Fig 7. Example of foraging behavior of BIB agent in the case of five prey patches. (a) Movement pattern. In the figure, two foraging points at adjacent time points are connected by a straight line. (b) Time evolution of the hypothesis with the highest confidence. (c) Time evolution of the x-coordinate of the foraging point. (d) Time evolution of the y-coordinate of the foraging point.

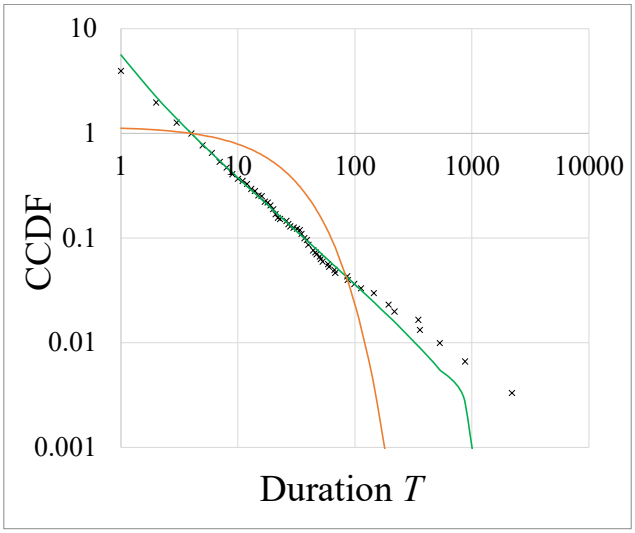


Fig 8. Example of the CCDF of the duration of the hypothesis with the highest confidence when there are five prey patches. This figure also shows the results of fitting a TP model (green) and an EP model (red) to the simulation data. The exponent of TP is $\eta = 1.98$. The fitting range is [4, 2198].

Supplementary Information for

Simulation of foraging behavior using a decision-making agent with Bayesian and inverse Bayesian inference: Lévy and Brownian walk-like search patterns derived from differences in prey distributions

Fitting to simulation data

Here, a method is described to fit the frequency distribution of the data observed by simulation to the truncated power-law distribution (TP) model and the exponential distribution (EP) model. Here, the case in which the data take continuous values is discussed. For discrete values, please refer to [1].

Fitting to TP

Here, a method is described to fit the frequency distribution of the step length l observed by simulation to the TP. The method is based on previous studies [2–6]. Specifically, the aim is to determine the minimum $\hat{l}_{\min}$ and maximum values $\hat{l}_{\max}$ of the observed data to be fitted to the TP model and the exponent $\hat{\eta}$ of the TP model that best fits the data in the range $\hat{l}_{\min} \le l \le \hat{l}_{\max}$. First, $\hat{l}_{\max}$ is the longest step length of the observation data. Next, the calculation method $\hat{l}_{\min}$ is described. In the case of a continuous distribution, TP in the range $l_{\min} \le l \le l_{\max}$ is expressed by the following formula:

$$p\left(l;\eta,l_{\min},l_{\max}\right)=\frac{\eta-1}{l_{\min}^{1-\eta}-l_{\max}^{1-\eta}}l^{-\eta} \qquad \text{(S37)}$$

The CDF of $p(l;\eta,l_{\min},l_{\max})$ is expressed in the following equation.

$$P(l;\eta,l_{\min},l_{\max})=1-\frac{l_{\min}^{1-\eta}-l^{1-\eta}}{l_{\min}^{1-\eta}-l_{\max}^{1-\eta}}=\frac{l^{1-\eta}+l_{\max}^{1-\eta}}{l_{\min}^{1-\eta}-l_{\max}^{1-\eta}} \quad \text{(S38)}$$

If the observed data in the range of $l_{\min}\le l\le l_{\max}$ are $\{l_1,l_2,\cdots,l_n\}$, then the log-likelihood of these data for TP is calculated using Equation (S37) as follows:

$$L(\eta;l_{\min},l_{\max})=\sum_{i=1}^{n}\ln p(l_i;\eta,l_{\min},l_{\max})=n\left(\ln(\eta-1)-\ln\left(l_{\min}^{1-\eta}-l_{\max}^{1-\eta}\right)\right)-\eta\sum_{i=1}^{n}\ln l_i \quad \text{(S39)}$$

The exponent $\hat{\eta}(l_{\min},l_{\max})$ of the TP model that best fits the data in the range $l_{\min}\le l\le l_{\max}$ is η that maximizes $L(\eta;l_{\min},l_{\max})$. Specifically, η is varied from 0.5 to 3.5, in increments of 0.01, to obtain $\hat{\eta}(l_{\min},l_{\max})$ that numerically maximizes Equation (S39).

The authors introduce the Kolmogorov-Smirnov statistic $D(l_{\min},l_{\max})$ to measure the closeness of the cumulative frequency distribution $S(l;l_{\min},l_{\max})$ obtained from the data in the range $l_{\min}\le l\le l_{\max}$ and the theoretical cumulative frequency distribution $P(l;\hat{\eta}(l_{\min},l_{\max}),l_{\min},l_{\max})$ represented by Equation (S38).

$$D(l_{\min},l_{\max})=\max_{l_{\min}\le l\le l_{\max}}\left|S(l;l_{\min},l_{\max})-P(l;\hat{\eta}(l_{\min},l_{\max}),l_{\min},l_{\max})\right| \quad \text{(S40)}$$

If $l_{\max}=\hat{l}_{\max}$ is fixed, then $D(l_{\min},\hat{l}_{\max})$ is a function of $l_{\min}$. $l_{\min}$ that minimizes $D(l_{\min},\hat{l}_{\max})$ is numerically chosen from the observed data. That is, $\hat{l}_{\min}=\arg\min_{l_{\min}} D(l_{\min},\hat{l}_{\max})$. From the above equations, $\hat{l}_{\min}$ and $\hat{l}_{\min}$ are obtained. Finally, the exponent $\hat{\eta}=\hat{\eta}(\hat{l}_{\min},\hat{l}_{\max})$ of the TP model that best fits the data in the range $\hat{l}_{\min}\le l\le\hat{l}_{\max}$, is determined using Equation (S39).

Fitting to EP

In this section, the aim is to determine the minimum value $\hat{l}_{\min}$ of the observed data to be fitted to the EP model and the exponent $\hat{\lambda}$ of the EP model that best fits the data in the range $\hat{l}_{\min} \le l$. The EP in the range $l_{\min} \le l$ is expressed by the following equation:

$$p\left(l;\lambda,l_{\min}\right)=\lambda e^{-\lambda\left(l-l_{\min}\right)} \qquad \text{(S41)}$$

The CDF of $p\left(l;l_{\min},\lambda\right)$ is expressed as follows:

$$P\left(l;\lambda,l_{\min}\right)=e^{-\lambda\left(l-l_{\min}\right)} \qquad \text{(S42)}$$

If the data in the range of $l_{\min} \le l$ are $\left\{l_1,l_2,\cdots,l_m\right\}$, then the log likelihood for these data is expressed as

$$L\left(\lambda;l_{\min}\right)=\sum_{i=1}^{m}\ln p\left(l_i;\lambda,l_{\min}\right)=m\ln\lambda-\lambda\sum_{i=1}^{m}\left(l_i-l_{\min}\right) \qquad \text{(S43)}$$

The exponent $\hat{\lambda}\left(l_{\min}\right)$ that maximizes $L\left(\lambda;l_{\min}\right)$ is obtained as a solution to $\frac{\partial L\left(\lambda;l_{\min}\right)}{\partial\lambda}=0$ using the following formula:

$$\hat{\lambda}\left(l_{\min}\right)=n\left(\sum_{i=1}^{n}\left(l_i-l_{\min}\right)\right)^{-1} \qquad \text{(S44)}$$

The Kolmogorov-Smirnov statistic $D\left(l_{\min}\right)$ to measure the closeness of the cumulative frequency distribution $S\left(l;l_{\min}\right)$ obtained from the data in the range $l_{\min} \le l$ and the theoretical cumulative frequency distribution $P\left(l;\hat{\lambda}\left(l_{\min}\right),l_{\min}\right)$ represented by Equation (S42).

$$D\left(l_{\min}\right)=\max_{l_{\min}\le l}\left|S\left(l;l_{\min}\right)-P\left(l;\hat{\lambda}\left(l_{\min}\right),l_{\min}\right)\right| \text{(S45)}$$

$\hat{l}_{\min}$ is calculated from $\hat{l}_{\min}=\arg\min_{l_{\min}} D\left(l_{\min}\right)$. The final value is $\hat{\lambda}=\hat{\lambda}\left(\hat{l}_{\min}\right)$.

Comparison of TP and EP

In this section, the authors describe a method for determining the most suitable distribution model (TP or EP) for the simulation data. Akaike information criteria weights (AICw) are used for comparison [6]. First, the Akaike information criterion (AIC) for data in the range $l_{\min} \le l \le l_{\max}$ is defined as follows:

$$\begin{aligned} AIC_{TP} &= -2\ln\left(L\left(\hat{\eta};l_{\min},l_{\max}\right)\right)+2 \\ AIC_{EP} &= -2\ln\left(L\left(\hat{\lambda};l_{\min}\right)\right)+2 \end{aligned} \tag{S46}$$

The AIC difference Δ is thereafter calculated as follows:

$$\begin{aligned} AIC_{\min} &= \min\left(AIC_{TP}, AIC_{EP}\right) \\ \Delta_{TP} &= AIC_{TP} - AIC_{\min} \\ \Delta_{EP} &= AIC_{EP} - AIC_{\min} \end{aligned} \tag{S47}$$

Finally, AICw are calculated as follows:

$$\begin{aligned} w_{TP} &= \frac{e^{-\Delta_{TP}/2}}{e^{-\Delta_{TP}/2}+e^{-\Delta_{EP}/2}} \\ w_{EP} &= \frac{e^{-\Delta_{EP}/2}}{e^{-\Delta_{TP}/2}+e^{-\Delta_{EP}/2}} \end{aligned} \tag{S48}$$

First, using the data in the range $\hat{l}_{\min} \le l \le \hat{l}_{\max}$ calculated during the fitting of TP, the most appropriate exponents, $\hat{\eta}$ and $\hat{\lambda}$, are determined for each model.

Next, these exponents are used to calculate and compare AICw. Subsequently, the authors changed $\hat{l}_{\min}$ to the one calculated during the fitting of the EP and compared. If $w_{TP} > w_{EP}$ for both data, TP is considered to fit the simulated data better. However, if $w_{TP} < w_{EP}$ for both data, EP is considered to fit the simulated data better. In case of discrepancies between the results in both datasets, the following indicators are defined and examined according to reference [2].

$$\begin{aligned} D_{adj,TP} &= \frac{\ln N}{\ln n_{TP}} D_{TP} \\ D_{adj,EP} &= \frac{\ln N}{\ln n_{EP}} D_{EP} \end{aligned} \tag{S49}$$

where D_{TP} and D_{EP} are the Kolmogorov-Smirnov statistics calculated during the model fitting of TP and EP, respectively. N is the total number of observed data points, and n_{TP} and n_{EP} are the number of observed data points used in each model fitting. In other words, the index considers a model that can fit more observational data. In the case of $D_{adj,TP} < D_{adj,EP}$, TP is considered to fit the simulation data better. Conversely, when $D_{adj,TP} > D_{adj,EP}$, EP is considered to fit the simulation data better.

REFERENCES

1. Shinohara S, Manome N, Nakajima Y, Gunji YP, Moriyama T, Okamoto H, et al. Power Laws Derived from a Bayesian Decision-Making Model in Non-Stationary Environments. Symmetry. 2021;13(4): 718. doi: 10.3390/sym13040718.
2. Humphries NE, Weimerskirch H, Queiroz N, Southall EJ, Sims DW. Foraging success of biological Lévy flights recorded in situ. Proc Natl Acad Sci U S A. 2012;vol. 109, 19: 7169-7174. doi:10.1073/pnas.1121201109.
3. Jansen VA, Mashanova A, Petrovskii S. Comment on "Lévy walks evolve through interaction between movement and environmental complexity". Science. 2012 Feb 24;335(6071): 918; author reply: 918. doi: 10.1126/science.1215747, PMID: 22362991.
4. White EP, Enquist BJ, Green JL. On estimating the exponent of power-law frequency distributions. Ecology. 2008 Apr;89(4): 905-912. doi: 10.1890/07-1288.1. Erratum in: Ecology. 2008 Oct;89(4): 905–912. DOI: 10.1890/07-1288.1, PMID: 18481513.
5. Clauset A, Shalizi CR, Newman MEJ. Power-Law Distributions in Empirical Data. In SIAM Rev. 2009;51(4): 661–703. DOI: 10.1137/070710111.
6. Edwards AM, Phillips RA, Watkins NW, Freeman MP, Murphy EJ, Afanasyev V, et al. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. Nature. 2007 Oct 25;449(7165): 1044-1048. doi: 10.1038/nature06199, PMID: 17960243.