Heinrich Heine Universität Düsseldorf

A hitchhiker's guide to active motion

Tobias Plasczyk, Paul A. Monderkamp, Hartmut Löwen & René Wittmann

Article - Version of Record

Suggested Citation:
Plasczyk, T., Monderkamp, P. A., Löwen, H., & Wittmann, R. (2025). A hitchhiker's guide to active motion. The European Physical Journal E, 48(1), Article 1. https://doi.org/10.1140/epje/s10189-024-00465-0

Wissen, wo das Wissen ist.

UNIVERSITÄTS- UND LANDESBIBLIOTHEK DÜSSELDORF

This version is available at:

URN: https://nbn-resolving.org/urn:nbn:de:hbz:061-20250312-105704-4

Terms of Use:

This work is licensed under the Creative Commons Attribution 4.0 International License.

For more information see: https://creativecommons.org/licenses/by/4.0

Eur. Phys. J. E (2025) 48:1
https://doi.org/10.1140/epje/s10189-024-00465-0

THE EUROPEAN PHYSICAL JOURNAL E

Regular Article - Flowing Matter

A hitchhiker's guide to active motion

Tobias Plasczyk[1,a], Paul A. Monderkamp[1], Hartmut Löwen[1,b], and René Wittmann[1,2,c]

[1] Institut für Theoretische Physik II: Weiche Materie, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany
[2] Institut für Sicherheit und Qualität bei Fleisch, Max Rubner-Institut, E.-C.-Baumann-Straße 20, 95326 Kulmbach, Germany

Received 3 October 2024 / Accepted 28 November 2024 / Published online 3 January 2025
© The Author(s) 2024

Abstract Intelligent decisions in response to external informative input can allow organisms to achieve their biological goals while spending very little of their own resources. In this paper, we develop and study a minimal model for a navigational task, performed by an otherwise completely motorless particle that possesses the ability of *hitchhiking* in a bath of active Brownian particles (ABPs). Hitchhiking refers to identifying and attaching to suitable surrounding bath particles. Using a reinforcement learning algorithm, such an agent, which we refer to as intelligent hitchhiking particle (IHP), is enabled to persistently navigate in the desired direction. This relatively simple IHP can also anticipate and react to characteristic motion patterns of their hosts, which we exemplify for a bath of chiral ABPs (cABPs). To demonstrate that the persistent motion of the IHP will outperform that of the bath particles in view of long-time ballistic motion, we calculate the mean-squared displacement and discuss its dependence on the density and persistence time of the bath ABPs by means of an analytic model.

1 Introduction

Living organisms possess the ability to consume energy from the environment and utilise it for developing biologically advantageous strategies. One famous example is swimming organisms which turn this energy into directed self-propelled motion [1,2]. This active behaviour leads to several advantages when it comes to efficiently exploring space, finding new resources or escaping predators. In the usual scenario, however, directed motion alone is barely a sufficient tool for most organisms when it comes to the preservation of one's livelihood. Therefore, the archetypal survival strategy in biology incorporates at least a coupling to a rudimentary form of intelligence, which is characterised by a sensory input and a means of interpretation [3–6]. In the human world, it is utterly common that motorless individuals utilise the means of motorisation of others. This behaviour, popularly termed hitchhiking [7], originates from the desire to optimise one's spending of energy, e.g. for economical or ecological reasons.

There are also a lot of motorless organisms in nature, such as certain species of bacteria, archaea, algae, marine plankton or atmospheric aeroplankton [8], which thrive despite entirely lacking the ability for self-propulsion. For example, airborne spores and pollen rely on wind currents [9] or terrestrial dust storms [10] as their only means of transportation. Even the concept of human hitchhiking is found in more evolved life forms, which developed ingenious navigational strategies that involve attaching themselves to other, usually larger, species [11–14]. Specifically, many marine organisms such as *Barnacles*, *Jellyella* and also microscopic bacteria can move over considerable distances, through layers of the ocean, or travelling upstream along with their hosts [13–17]. Other species elevate the concept of hitchhiking by forming symbiotic relationships that surpass the mere desire for transportation. In addition to being able to swim on its own, the *suckerfish* attaches to larger hosts, profiting from the protection, nutrition and fast incoming water flows, which aids its respiration, while, conversely, the host profits from the provided skincare [18,19].

Due to the interdisciplinary importance of active motion [20], huge efforts were dedicated to theoretically characterise the self-propulsion in both living matter and inanimate systems [21–23]. One standard model is active Brownian particles (ABPs), which propel themselves with a constant velocity along their instantaneous orientation that undergoes rotational diffusion. This toolbox has stimulated theoretical predictions, ranging from analytic results on the single particle level [24–27] to insight into a vast range of collective phenomena [28–31]. For a more realistic description of natural or artificial experimental systems [32–34], several generalisations of the ABP model were devised, including chiral

[a] e-mail: tobias.plasczyk@hhu.de (corresponding author)
[b] e-mail: hartmut.loewen@hhu.de (corresponding author)
[c] e-mail: rene.wittmann@hhu.de (corresponding author)

ABPs (cABPs) which tend to move in circles [35–38], anisotropic self-propelled particles which have different angular dynamics [24,39] and can collectively align their direction of motion [40], or active Langevin particles which incorporate the ubiquitous inertial effects from the macroscopic world [41,42]. Most notably, more recent research avenues evoke machine learning to gain deeper physical insight into active matter systems [43,44] or explore the idea of supplementing the ABP model with machine learning tools to enhance their navigation abilities [45–49]. The persistent motion of intelligent motorless matter has received considerably less attention.

Here, we provide a minimal model for understanding hitchhiking in nature. More specifically, we place a single intelligent hitchhiking particle (IHP) in a bath of ABPs. The IHP is unable to propel itself but can attach itself to the ABPs and follow their path. To model sensory input coupled with intelligent decision-making aiming towards directed motion, this attachment is steered through reinforcement learning [50]. The learning objective is to reach the top of the simulation box as fast as possible, while the perceived information about the surrounding bath is limited. Over the course of the training protocol, the IHP deduces a strategy which allows it to selectively attach to bath ABPs with favourable orientations. The final selection rule depends on the bath particles' degree of persistence, indicating that the IHP can balance between (i) accepting a possibly longer waiting time and only joining for highly promising travel directions (better for a highly persistent bath) or (ii) taking the odds for a larger range of initial travel directions (better for more erratic movement of bath particles). The symmetry of this transport problem is broken by the possibility to let go of the travel partner if its orientation turns out to develop unfavourably. Moreover, the IHP learns to anticipate the circular movement of cABPs by shifting the interval of favourable orientations according to the circular frequency of such bath particles. Once acquired, the IHP's strategy can also be transferred to a system of particles interacting with a soft potential.

This paper is structured as follows. In Sect. 2 we describe the Q-learning process and our different IHP models, which we then analyse in Sect. 3 to discuss how the learning results and persistent motion depend on the properties of the active bath. We conclude in Sect. 4

2 Models for an intelligent hitchhiking particle (IHP)

As illustrated in Fig. 1, our model consists of a two-dimensional environment of non-interacting active Brownian particles (ABPs) constituting potential travel partners of the intelligent hitchhiking particle (IHP). Since it possesses no own means of propelling, its movement is governed by that of the ABP to which it temporarily attaches. The ultimate goal is to learn a strategy, which allows the IHP to travel in the intended direction as fast as possible. The motion pattern resulting from a successful hitchhiking strategy is exemplified in the supplementary video.

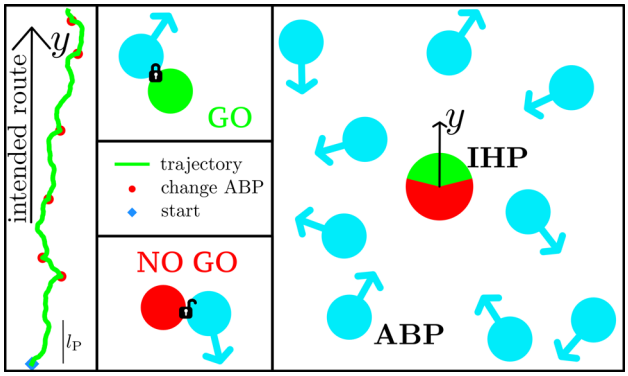

Fig. 1 Hitchhiking strategy to perform persistent Brownian motion. An intelligent hitchhiking particle (IHP, green/red or dark/light grey) is immersed in a bath of non-interacting active Brownian particles (ABPs, cyan). Each ABP moves in a certain direction for their persistence length l_p (see Fig. 2 for details on the model). The IHP learns to fulfil a navigational task by making binary decisions whether to attach and follow the closest ABP (action: GO, green or light grey) or not to attach or let go (action: NO GO, red or dark grey). The fully trained IHP can then persistently move to the top (green trajectory), i.e. in positive y-direction, on a length scale much larger than l_p (see the scale bar at the bottom left), changing its travel partner if necessary (red dots). This decision diagram, resulting from a Q-learning algorithm, is displayed as a pie chart decorating the IHP, whose colour indicates the chosen action depending on the current orientation (cyan arrows) of the nearest ABP (see Fig. 3 for details on the learning process)

The essence of our model is summarised in Fig. 2, and details are given in the remainder of this section. The stochastic dynamics of the bath ABPs is characterised by persistent motion in the direction of their current orientations ϕ_i, which we recapitulate in Sect. 2.1. The training process, in which the IHP learns how to interact with its environment, is described in Sect. 2.2. The central learning objective is to obtain a decision diagram, representing an interval Φ_{GO} of favourable angles on the unit circle, such that the IHP wants to attach to an ABP (denoted by GO) if $\phi_i \in \Phi_{\mathrm{GO}}$ and does not want to attach or let go (denoted by NO GO) if $\phi_i \notin \Phi_{\mathrm{GO}}$. To characterise the IHP's behaviour, we thus introduce a GO angle $\Delta\phi$, which specifies the size of this interval, and an anticipation angle ϕ_0, indicating its location on the unit circle. Thus, we have (always implying a 2π-periodicity of the polar angle)

$$\Phi_{\mathrm{GO}} = \left[\phi_0 - \frac{\Delta\phi}{2}, \phi_0 + \frac{\Delta\phi}{2}\right] . \tag{1}$$

In Sect. 2.3, we translate the decision diagram of a fully trained IHP to a (non-reciprocal) interaction potential,

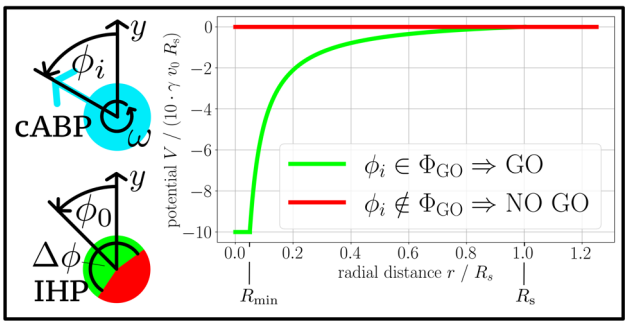

Fig. 2 Interaction of the IHP with the bath ABPs. The ABPs (top left) move according to Eqs. (2) and (3) with constant self-propulsion velocity v_0 in the direction indicated by their instantaneous orientational angle ϕ_i, which is subject to rotational Brownian motion. In addition, chiral ABPs (cABPs) experience an internal torque resulting in a constant circular frequency ω. The fully trained IHP (bottom left) is characterised by the decision diagram, indicating the values of ϕ_i at which it chooses the actions GO or NO GO, compare Fig. 1. This internal memory is quantified by the GO angle $\Delta\phi$, denoting the probability that a certain travel partner is chosen, and the anticipation angle ϕ_0, indicating the mean orientation of the chosen travel partner. Hence, a GO decision is made if $\phi_i \in \Phi_{\mathrm{GO}}$, as defined in Eq. (1). Based on this decision diagram, we consider two models for the IHP motion. The first model is the Q-learning IHP, which shares the path of the chosen ABP according to Eq. (4) and is used for training. As a second model, we transfer the acquired knowledge to the potential IHP, which has its own equation of motion (8) and is attracted to all favourable bath particles within range. The corresponding force is generated by the non-reciprocal interaction potential $V(r, \phi_i)$ plotted on the right

which only generates a force on the IHP towards an ABP if $\phi_i \in \Phi_{\mathrm{GO}}$. Both IHP models are summarised and compared in Sect. 2.4.

2.1 Active Brownian particles (ABPs)

We consider a bath of N non-interacting ABPs (top left drawing in Fig. 2), labelled as $i = 1, \ldots, N$, in a periodic square simulation box with side lengths L, where the bath density is defined as $\rho = N/L^2$. These bath particles move at a constant self-propulsion velocity v_0 in the direction of their orientation vector $\mathbf{u}_i(t) = (\cos(\phi_i(t)), \sin(\phi_i(t))$ in polar coordinates. The orientation angles ϕ_i are defined such that $\phi_i = 0$ corresponds to motion in the preferred travel direction of the IHP, as drawn in Fig. 1. They change over time due to rotational diffusion characterised by Brownian white noise ξ_i with $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_j(t') \rangle = 2\delta_{ij} D_{\mathrm{r}} \delta(t - t')$, where D_{r} is the rotational diffusion constant. Neglecting translational diffusion, the equations of motion for the ABPs' centre-of-mass $\mathbf{r}_i(t)$ read

$$\dot{\mathbf{r}}_i(t) = v_0 \mathbf{u}_i(t). \tag{2}$$

In addition, we consider the angular dynamics

$$\dot{\phi}_i(t) = \xi_i + \omega \tag{3}$$

including a constant circular frequency ω, which leads to ABPs that have the tendency swim on circular trajectories. We use these chiral active Brownian particles (cABPs) to investigate the learning behaviour of the IHP in different environments.

The equations of motion (2) and (3) are integrated, using a forward Euler–Maruyama method with a finite time step Δt. Specifically, we discuss results upon varying the rotational diffusivity D_{r}, which sets the persistence time $\tau = D_{\mathrm{r}}^{-1}$ and length $l_{\mathrm{p}} = \frac{v_0}{D_{\mathrm{r}}}$ of an ABP (with $\omega = 0$), and the additional circular frequency ω of a cABP, while keeping the self-propulsion velocity v_0 fixed.

2.2 Q-learning IHP

The IHP (bottom left drawing in Fig. 2) cannot move through self-propulsion. Its only means of transportation is to attach to the nearest ABP. To mimic realistic behaviour, only bath particles within a certain perception range, indicated by the scan radius R_{s}, can be considered and the IHP requires a perception time τ_Q before being able to make a new decision. In more detail, the IHP makes its $(n+1)$th decision at time $t = n\tau_Q$ with $n \geq 0$. If there are potential travel partners within the scan radius R_{s}, it picks the nearest one. Further, if this selected ABP has a favourable orientation, the IHP will share its path for a time span of at least τ_Q (GO). Otherwise, the IHP will rest at its current position and wait for a time span of τ_Q until the next decision can be made (NO GO). Note that this also applies to a NO GO decision causing the IHP to leave its current travel partner.

We thus model the time evolution of the IHP position $\mathbf{r}(t)$, which starts at $\mathbf{r}(0) = \mathbf{0}$, separately for each perception time span $n\tau_Q < t \leq (n+1)\tau_Q$ between two decisions according to

$$\mathbf{r}(t) = \begin{cases} \mathbf{r}_i(t)\,, & \text{if } (\phi_i(n\tau_Q) \in \Phi_{\mathrm{GO}} \;\&\; |\mathbf{r}(n\tau_Q) - \mathbf{r}_i(n\tau_Q)| \leq R_{\mathrm{s}}) \\ \mathbf{r}(n\tau_Q)\,, & \text{else} \end{cases},$$
$$i = \arg\min_j |\mathbf{r}(n\tau_Q) - \mathbf{r}_j(n\tau_Q)|\,, \tag{4}$$

where i labels the ABP closest to the IHP. The set Φ_{GO} of favourable orientational angles, as given in Eq. (1), follows from the learning process. These equations of motion are universal to the IHP, while only the basis of decision, encoded in Φ_{GO}, evolves while learning.

2.2.1 Learning algorithm

In order to enable navigation in the environment of randomly distributed ABPs, a tabular Q-learning algorithm is used to train the IHP [51,52]. We thus define an action space and a state space for this agent. The action

space holds all possible actions $\nu \in \{0, 1\}$ performed by the IHP, meaning *sticking to an* ABP (GO) if $\nu = 1$ or *not sticking to an* ABP (NO GO) if $\nu = 0$. On the other hand, the state space is given by all polar angles, i.e. the instantaneous travel direction of the selected ABP, such that the state $\mu \in \{1, 2, \ldots, 360\}$ labels 360 discrete intervals, which provide a sufficiently fine numerical resolution for the parameters considered here. This results in a 360×2 matrix $Q_{\mu\nu}$, whose rows represent the 360 states, with each state hosting two actions, represented by the columns. The Q-matrix holds the information collected during the learning process and serves as a basis of decision-making. The resulting decision amounts to performing the action

$$A_\mu = \arg\max_\nu Q_{\mu\nu} \in \{0, 1\} \tag{5}$$

in the respective state μ that has the highest value $Q_{\mu\nu}$. In case of equal values, $Q_{\mu 0} = Q_{\mu 1}$, a random action is performed. If no ABP is within reach, no state can be defined and the agent rests until the next decision (no learning will occur in this step).

A single training cycle consists of 1000 episodes. One episode ends once the IHP reaches the top or bottom of the simulation box of fixed length L or when a maximum time $T_{\max} = 6 \cdot 10^4 \tau_Q$ is reached. The Q-matrix is initialised with zeros. As such, the agent mainly performs random actions due to their equal weight. This reflects the lack of experience in the early stages of the learning process. An independent probability ϵ that the IHP decides randomly decreases linearly from one to zero with increasing number of episodes. This leaves a probability of $1-\epsilon$ to actively decide according to what the agent has learned so far, i.e. through Eq. (5) evaluated for the current values of $Q_{\mu\nu}$. This is to favour an exploratory approach in the beginning and an exploitative approach in the final stages of the training.

The entries $Q_{\mu\nu}$ of the Q-matrix are modified with rewards or punishments. Specifically, after the agent has performed action ν in its old state μ, learning progresses according to the update formula [50]

$$Q^{\text{new}}_{\mu\nu} = (1-\alpha)Q_{\mu\nu} + \alpha(R + \gamma \max_\lambda(Q_{\mu'\lambda})) \tag{6}$$

which shall improve the estimate of how likely it is that this action in that state will lead to rewards in the future. Specifically, each decision eventually leaves the agent in a new state μ' after the perception time τ_Q, when a new decision has to be made. Depending on the value $\max_\lambda(Q_{\mu'\lambda})$, corresponding to the preferred action in this new state, there will be a modification on $Q^{\text{new}}_{\mu\nu}$, as controlled by the *discount factor* $\gamma = 0.9$. This hyperparameter thus determines how much the values of the different states influence each other during a learning episode. Moreover, R provides a global reward (or punishment) when the agent reaches its goal at the top (or hits the bottom), such that a final update is made at the end of each episode, where we use

$$R(y) = \begin{cases} +100\,, & \text{if } y \geq L/2 \\ -100\,, & \text{if } y \leq L/2 \\ 0 & \text{else} \end{cases} . \tag{7}$$

Finally, the *learning rate* $\alpha = 0.01$ regulates how much of the new information is used in updating $Q_{\mu\nu}$, such that the algorithm can converge properly.

2.2.2 Repeated training

To get useful results that will serve for our later analysis of the IHP's motion, the full learning procedure is completed only after repeating the training cycle described in Sect. 2.2.1. We thus consider the IHP to be fully trained after having passed through 303 independent training cycles, each consisting of 1000 episodes. Figure 3 displays this progress from the first learning steps to the final strategy, encoded in the Q-matrix.

In Fig. 3a, we show three exemplary trajectories throughout the training in a bath of ABPs (with $\omega = 0$), demonstrating how the strategy gradually improves. Initially, the observed behaviour is governed by random actions, resulting in a regular change of travel partner and travel direction. Halfway through the training, every second decision is, on average, made on the basis of the Q-matrix and the IHP increasingly benefits from the persistent ABP motion. Eventually, it has learned to pick only upward moving travel partners and stay with them for up to about one persistence length. The resulting trajectory displays no downward movement at all.

The main outcome of this learning algorithm is encoded in the Q-matrix, which holds values for both actions in each state, where the highest value in each state defines the action, GO or NO GO, according to Eq. (5). Hence, the content of the matrix elements can be condensed into a binary decision diagram only indicating the chosen action in each state. This intrinsic memory of the IHP is represented by a pie chart, as illustrated in Fig. 1. However, due to our randomised setup, the final Q-matrix slightly differs when repeating the training of the IHP. It is also possible to end up with a decision diagram displaying alternating actions for neighbouring states, rather than only two regions with a sharp distinction of GO and NO GO. This is because the matrix elements for the two actions can be quite similar, in particular for travel angles corresponding to sideways ABP movement. To obtain clean decision diagrams, we determine the final GO probability that the IHP will attach to an ABP with orientation ϕ_i from the relative frequency $h_{\text{GO}}(\phi_i)$ that the IHP makes a GO decision at this angle after having completed 303 independent training cycles. As exemplified by the histograms in Fig. 3b, this pointwise probability changes quite smoothly with the ABP orientation. We then determine the decision diagram of the fully trained IHP by associating all angles with $h_{\text{GO}}(\phi_i) \geq 0.5$ and $h_{\text{GO}}(\phi_i) < 0.5$ to GO and NO GO, respectively. This final result can then directly be translated into the set Φ_{GO} of GO angles from Eq. (1). As the average decision

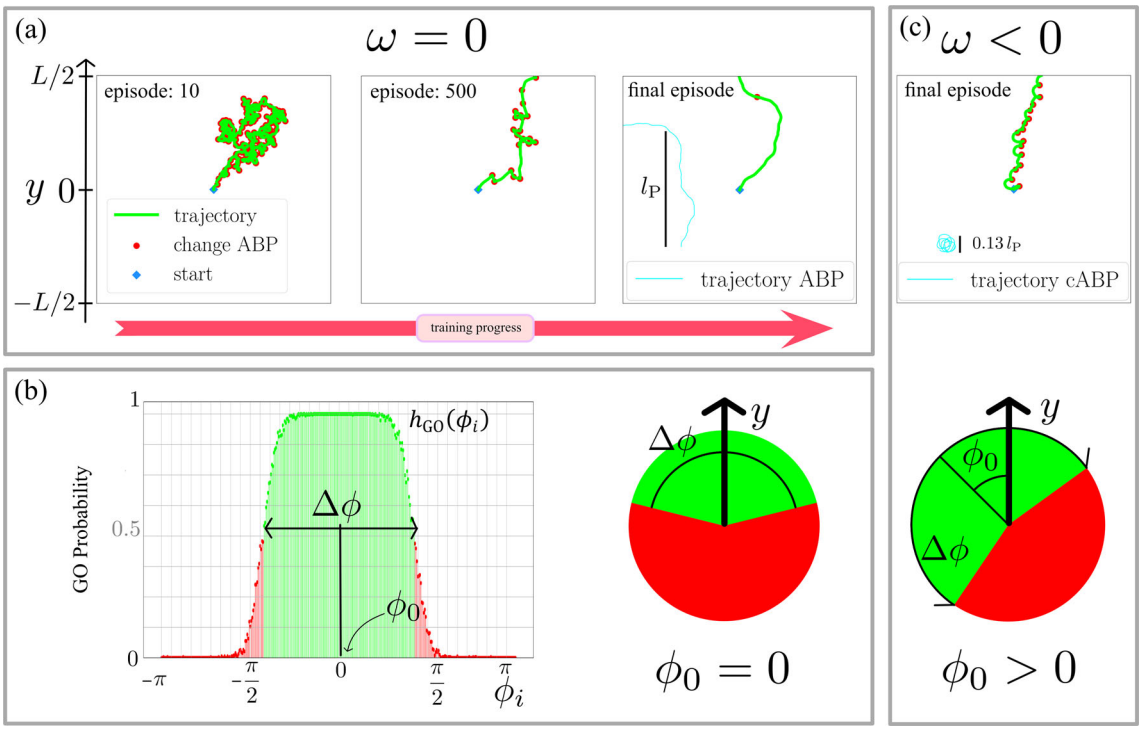

Fig. 3 Full learning process of the IHP. **a** Trajectories (green lines) over the course of one training cycle, consisting of 1000 episodes (see labels), where the IHP successively learns to navigate in a bath of ABPs with density $\rho = 0.6/R_s^2$ and rotational diffusivity $D_r = 6\pi^2 \cdot 10^{-3}/\tau_Q$. The goal is to reach the top, i.e. moving half a box length $L/2 = 25R_s$ in positive y-direction. As in Fig. 1, the red dots indicate the choice of action NO GO followed by a change of travel partner. A typical trajectory of a bath particle with persistence length $l_p = v_0/D_r$ is depicted in cyan. **b** Obtaining the final decision rule after repeating the complete training cycle for 303 times. For each state, i.e. for each orientation ϕ_i of a bath particle, we define the (pointwise) GO probability $h_{GO}(\phi_i)$ as the relative frequency of GO actions, sampled from all training cycles (left). The most likely action GO (green or light grey) or NO GO (red or dark grey) can be directly inferred from the value $h_{GO}(\phi_i) \geq 0.5$ or $h_{GO}(\phi_i) < 0.5$, which defines the decision diagram of the fully trained IHP (right). As annotated, the anticipation angle ϕ_0 and the GO angle $\Delta\phi$ from Eq. (1) can be directly identified in both representations. **c** According IHP and typical bath particle trajectories (top) and decision diagram of the fully trained IHP for a bath of cABPs, which additionally have a nonzero circular frequency $\omega = -36\pi/(100\tau_Q)$ in their angular dynamics

value for the APB bath is typically symmetric around $\phi_i = 0$, there is no anticipation angle, $\phi_0 = 0$, while the GO angle $\Delta\phi$ spans a range of favourable travel directions typically pointing upwards.

The flexibility of the learning procedure is verified by considering $\omega \neq 0$, i.e. a bath of cABPs. Also in this case, the trajectory of the IHP at the end of its training is directed upwards, as shown in the top of Fig. 3c. Since the characteristic radius of the circular trajectories is much smaller than the persistence length of the straight ABPs, this directed hitchhiking requires much more proactive intelligent decision-making and an appropriate timing to change the travel partner. This results in a strategy adjusted to the dynamics of the travel partner, as reflected by the nonzero anticipation angle ϕ_0 in the decision diagram at the bottom of Fig. 3c. In particular, ϕ_0 typically has the opposite sign as the circular frequency ω of the bath particles and thus conveniently quantifies how much the IHP needs to anticipate future changes in travel direction.

2.3 Potential IHP

The motion (4) of the Q-learning IHP, as introduced in Sect. 2.2, follows a simple identification rule, namely travelling along with a bath ABP or not. Its decision strategy derived from the training scheme in Fig. 3 is purely characterised by whether the orientation of the closest bath particle lies within a certain angular interval. This behaviour can be imitated by purely physical interaction via a relatively simple attractive potential. To this end, we introduce the potential IHP as an alternative model by translating the GO action to an attractive force which pushes it towards all sufficiently close ABPs with a promising travel direction. These directions are inferred from the decision diagram obtained in the previously completed learning processes of the Q-learning IHP. Again, the motion of the bath ABPs is not affected by the presence of the potential IHP, which means that we introduce a non-reciprocal hitchhiking interaction.

Specifically, the potential IHP moves according to the overdamped equation of motion

$$\dot{\mathbf{r}}(t) = -\gamma^{-1} \sum_{i=1}^{N} \nabla V\left(|\mathbf{r} - \mathbf{r}_i|, \phi_i\right) \tag{8}$$

induced by each nearby bath particle, where γ is the friction coefficient. The orientation dependence of the non-reciprocal hitchhiking potential $V(r, \phi)$ follows from the binary distinction

$$V(r, \phi) = \begin{cases} V_0(r)\,, & \phi \in \Phi_{\mathrm{GO}} \\ 0\,, & \phi \notin \Phi_{\mathrm{GO}} \end{cases}, \tag{9}$$

selecting all favourable travel directions Φ_{GO} of a bath particle, given by Eq. (1). Moreover, the radial part within $V(r, \phi)$ is described by

$$V_0(r) = \begin{cases} 5\gamma v_0 R_{\mathrm{s}} \left(1 - \frac{R_{\mathrm{s}}}{r}\right), & R_{\min} \leq r \leq R_{\mathrm{s}} \\ 0\,, & \text{else} \end{cases}. \tag{10}$$

To prevent a divergence of the force, we introduce a lower cut-off to the potential, given by the minimal hitchhiking distance $R_{\min} = R_{\mathrm{s}}/20$, while the upper perception limit is given again by the scan radius R_{s}. This hitchhiking potential is shown in Fig. 2 together with an illustration of the IHP's decision diagram. In practice, it can happen that the potential IHP overtakes the ABP when updating the position according to Eq. (8) because of the finite time step Δt in our simulations. To prevent this discretisation artefact, the resulting displacement vector is renormalised to the distance between potential IHP and ABP whenever it is larger (while the direction of the displacement remains the same).

2.4 Model overview

Overall, we devise two different models for an intelligent hitchhiking particle (IHP), each describing individual hitchhiking capabilities.

The Q-learning IHP, on the one hand, is limited by its perception, as it can only make one cognitive operation of a certain complexity at discrete time steps. It also only focuses on a single bath particle, but, if this one turns out to be a suitable hitchhiking partner, will always rigorously follow its path, irrespective of the initial distance (within the scan radius R_{s}). This motion will be maintained until the next decision can be made after the perception time τ_Q, even if the trajectory takes an unfavourable direction before. This possibility must be appropriately anticipated in the learning process.

The potential IHP, on the other hand, is continuously pulled by all suitable travel partners (within range) at the same time instead of reacting by instantaneously attaching to a certain bath particle. Moreover, it does not feature a discrete perception time, which brings about a strength and a weakness at the same time. While it is less likely to be manoeuvred into an unfavourable travel direction during hitchhiking, detaching too early may increase the time it has to wait at rest. Keeping in mind that the decision diagram of the potential IHP is obtained from training the Q-learning IHP, it is not obvious which model shows the better performance in how far these features are advantageous or disadvantageous.

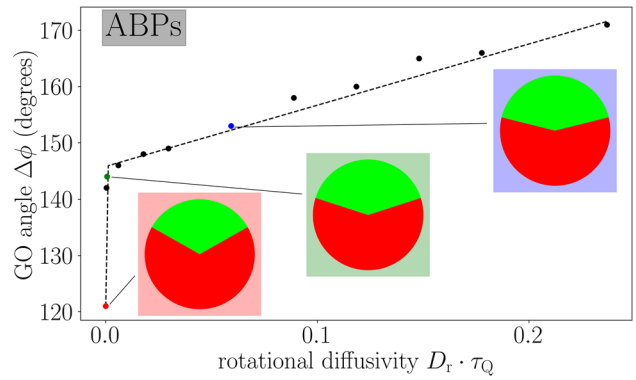

Fig. 4 Learning results in a bath of ABPs. The GO angle $\Delta\phi$ of a fully trained IHP (compare Fig. 3) is shown as a function of the rotational diffusivity D_{r} of the bath particles. We roughly observe a linear trend except for very small D_{r} (highly persistent ABPs), where the learning environment is too small compared to the persistence length. The dashed line thus indicates a fit to $a + bD_{\mathrm{r}} + c/D_{\mathrm{r}}$ as a guide to the eye. The anticipation angle $\phi_0 = 0$ is zero in all cases, as the circular frequency $\omega = 0$ vanishes for this bath. We also show the three decision diagrams corresponding to the values of D_{r} (coloured points) used in our later plots

In the following sections, we use τ_Q and R_{s} as time and length scales, respectively and consider a fixed self-propulsion velocity $v_0 = 3R_{\mathrm{s}}/(2\tau_Q)$ of the bath particles. The numerical time step is given as $\Delta t = \tau_Q/6$ and the length of the simulation box is $L = 25R_{\mathrm{s}}$. The rotational diffusivity D_{r} and the circular frequency ω of the bath particles are variable, while learning takes place at a fixed bath density $\rho = 0.6/R_{\mathrm{s}}^2$, since the choice of ρ does not affect the learning results. When evaluating the performance of the fully trained IHP, we also investigate the effect of the density, as the resulting change in waiting time is crucial for calculating the overall average hitchhiking velocity.

3 Hitchhiking behaviour

Equipped with the decision diagram obtained via Q-learning, the IHP is capable of navigating in different types of baths to persistently move towards its goal (in our case in positive y-direction). Below, we discuss its physical properties, depending on different models and bath parameters. Specifically, we discuss in Sect. 3.1 how the decision diagrams depend on the persistence time and circular frequency of the bath particles, before analysing the Q-learning IHP's mean-squared displace-

ment in Sect. 3.2. Finally, we compare the motion to the potential IHP in Sect. 3.3.

3.1 Learning results

At the end of its training, the IHP has learned to reach its goal by intelligently switching travel partners. As summarised in Fig. 3, the fully trained IHP attaches to an ABP with a promising orientation and holds on until the polar travel angle falls outside a certain range characterised by the anticipation angle ϕ_0 and the GO angle $\Delta\phi$. We thus observe persistent trajectories closely resembling that of the bath particles, while a change of travel partner is observed as soon as the trajectory tends to deviate too much from the intended travel route. In a bath of cABPs, such a change occurs more frequently at the same D_r, as the circular motion reduces the persistence and needs to be anticipated by the IHP. Hence, the IHP's trajectory is characterised by sequential semicircles and its decision diagram is asymmetric due to a nonzero anticipation angle. Below, we quantify these learning results in more detail.

Let us first focus on a bath of ABPs with vanishing circular frequency $\omega = 0$ and investigate the dependence of the GO angle $\Delta\phi$ on the rotational diffusivity D_r or, equivalently, on the persistence time $\tau = 1/D_\mathrm{r}$. As shown in Fig. 4, $\Delta\phi$ increases with increasing D_r, which indicates that the training of the IHP has increased its awareness towards the choice of travel partner by considering a smaller range of favourable angles when the bath particles become more persistent. In this case, the learned strategy to accept a longer waiting time for a bath particle with an optimal travel direction will naturally pay off in the long run. This trend can be intuitively explained by considering the consequences of choosing a travel partner whose orientation initially points sidewards. For a highly persistent bath, the travel direction will remain the same for an extended amount of time, such that the IHP has to stick to its decision without getting an actual benefit, while better alternatives could have been available in the mean time. For lower bath persistence, the travel direction randomises earlier, becoming either more favourable or triggering a NO GO decision allowing to search for a better partner. Over a large range of D_r, the change of GO angle is well described by a linear relation. A significant deviation from this trend is only observed in the extremely persistent case (small D_r), such that $l_\mathrm{p} \gg L$, where a change of a travel partner is almost never required during training.

To underline the flexibility of our Q-learning procedure, we now consider a bath of cABPs, for which we investigate the influence of the circular frequency ω on the decision diagram in Fig. 5. As shown in Fig. 5a, the GO angle $\Delta\phi$ increases approximately linearly for increasing ω, since the circular trajectories of the bath particles lead to less persistent motion (compare Fig. 3c). This observation is thus analogous to that in Fig. 4 for increasing the rotational diffusivity D_r, as discussed in the previous paragraph. Moreover, the slope of $\Delta\phi$ as a function of ω remains nearly independent of D_r, such that we expect the effect of varying D_r for cABPs to be the same as for ABPs. As in Fig. 4, the only exception is for an extremely persistent bath ($D_\mathrm{r} = 6\pi^2 \cdot 10^{-13}\tau_\mathrm{Q}^{-1}$ and $\omega = 0$).

Crucially, the circular frequency of the bath cABPs also challenges the IHP to anticipate the circular movement. This is reflected in our learning results, shown in Fig. 5b, by the increase of the absolute anticipation angle $|\phi_0|$ with increasing absolute circular frequency $|\omega|$, while the sign is the opposite. In other words, the IHP preferably selects travel partners whose orientations are not instantaneously pointing towards the goal, but will turn accordingly during the hike. Quantitatively, the linear fit $\phi_0 \approx -0.64\omega\tau_Q$ (measured in radians) tells us that the magnitude of the anticipation angle corresponds to about half the angle a bath particle turns during one perception time τ_Q. This farsighted

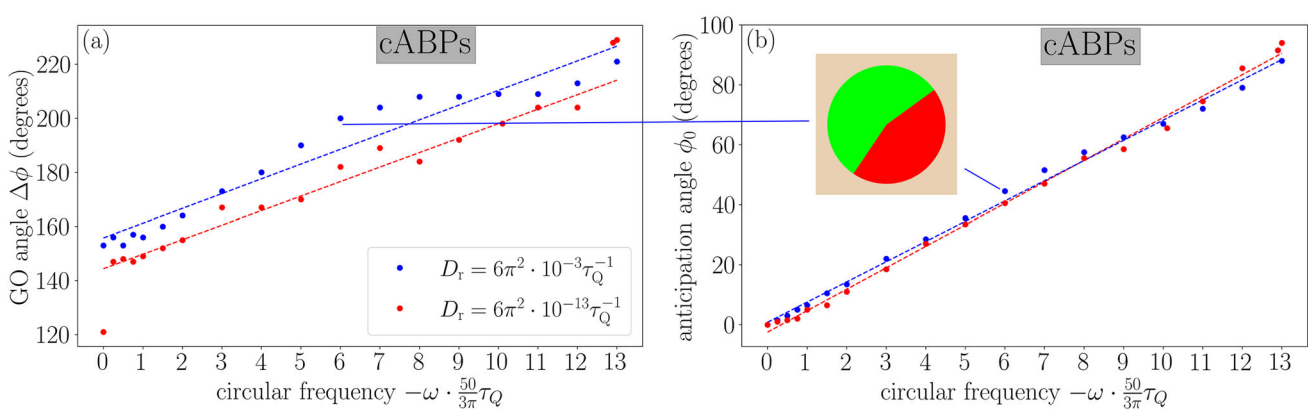

Fig. 5 Learning results in a bath of cABPs. **a** The GO angle $\Delta\phi$ and **b** the anticipation angle ϕ_0 of a fully trained IHP (compare Fig. 3) are shown as functions of the circular frequency ω of the bath particles. We consider two rotational diffusivities D_r following the colour code used in Fig. 4 for achiral ABPs (corresponding to the data for $\omega = 0$). The dashed lines indicate linear fits to guide the eye (the red point for $\omega = 0$ was excluded for this purpose). We also show the decision diagram corresponding to the parameters $D_\mathrm{r} = 6\pi^2 \cdot 10^{-3}\tau_\mathrm{Q}^{-1}$ and $\omega = 36\pi \cdot 10^{-2}\tau_\mathrm{Q}^{-1}$ used in Fig. 7

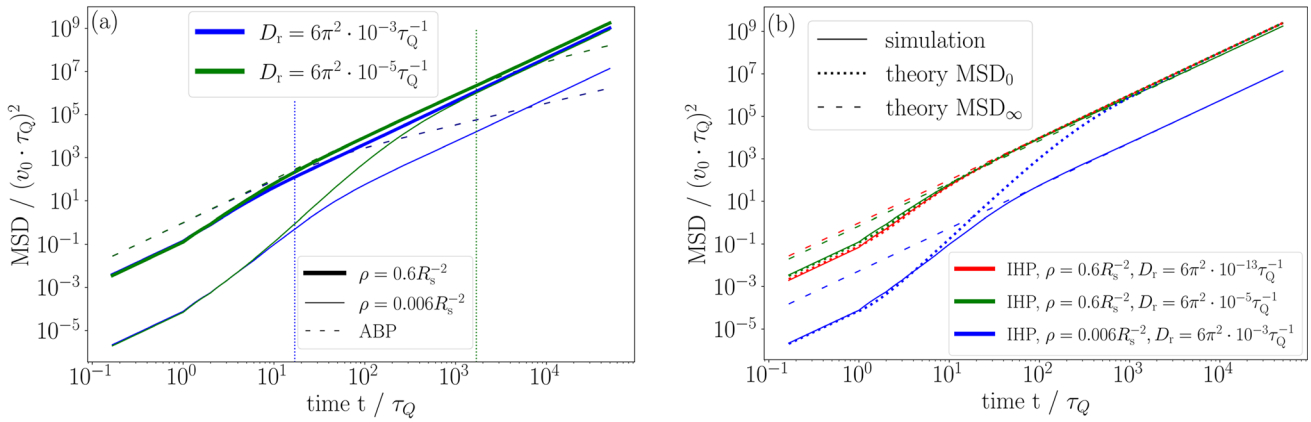

Fig. 6 Persistent motion of the Q-learning IHP. We show the MSD for different bath densities ρ and rotational diffusivities D_r (solid lines with colours given in the respective legends), where each curve is averaged over 1000 simulation runs. The numerical results are compared to **a** the MSD of a single bath ABP (dashed lines), where the vertical dotted lines indicate the persistence times $\tau = 1/D_r$, and **b** the theoretical predictions derived in appendix A, i.e. the short-time approximation MSD_0 (dotted lines) and the long-time asymptotic behaviour MSD_∞ (dashed lines)

anticipation required for high circular frequencies is thus learned successfully. We also find that the rotational diffusion barely affects the anticipation angle, as these orientational fluctuations average out while learning to cope with this deterministic effect.

As learning only takes place if a bath particle is within range, such that a certain action can be consciously chosen, the final decision diagrams are independent of the bath density ρ. Thus, the only effect of changing ρ, so far, is on the slope of the learning curve. Our chosen density for the learning process yields an optimal balance between the expected number of time steps required to complete the training (which decreases for larger ρ) and the computational cost to evaluate each time step (which increases for larger ρ). The performance of a fully trained IHP, however, very well depends on the bath density, as we examine below.

3.2 Ballistic motion of the IHP

To illustrate the IHP's travel characteristics, we examine its mean-squared displacement (MSD)

$$\mathrm{MSD} := \left\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \right\rangle \tag{11}$$

over the course of time. As discussed in appendix A (and described in the theory outlined below), the IHP can actually be well described as performing a drift motion with a time-dependent hitchhiking velocity $v_H(t)$. Nevertheless, we consider the MSD instead of the simpler mean displacement to better assess the IHP's performance in comparison to the bath ABPs as a reference.

As shown in Fig. 6, the IHP has successfully learned to enable directed motion through hitchhiking, which can be seen from the ballistic long-time behaviour. Specifically, for all parameters (bath density and rotational diffusivity), the IHP will eventually outperform a single bath ABP, whose motion becomes diffusive at times $t > \tau$ due to its finite persistence time τ, compare Fig. 6a. Upon closer inspection, we observe up to four different dynamical regimes, starting ballistically at very short times, followed by intermediate super-ballistic and sub-ballistic behaviour, until the IHP eventually enters its final ballistic state with a larger average velocity than in the first ballistic regime. The most important control parameter for this dynamical behaviour and also the overall hitchhiking efficiency is the bath density. In general, the MSD is always larger when the density is higher, as the average waiting time for a new travel partner decreases. This effect is nearly independent of the bath particles' rotational diffusivity at short times. The long-time performance, however, is crucially reduced if the bath is only weakly persistent and its density is low, because of the combined effect of the IHP leaving its travel partner more often and the longer waiting time for a new one. In contrast, hitchhiking in a dilute but persistent bath is only slightly less efficient in the long run, while bath persistence is generally not a relevant factor in a dense bath, as there is practically always an ABP within reach.

Both the qualitative behaviour and the magnitude of the MSD can be quantitatively explained via explicit considerations of the probability that a favourable travel partner is available to the IHP at time t. Our theoretical treatment is detailed in appendix A and outlined below. Let us first recall from Eq. (4) that the IHP will generally either remain at rest or travel with velocity v_0. Therefore, in the early stages, the IHP needs to wait to find its initial travel partner before being able to move at all. The first regime is thus at $t \leq \tau_Q$, for which we observe ballistic motion with the average velocity $p(0)v_0$. Given $N = \rho L^2$ bath particles in a periodic box of side length L, the required probability

$$p(0) = \left(1 - \left(1 - \frac{R_s^2 \pi}{L^2}\right)^{\rho L^2}\right) \frac{\Delta\phi}{2\pi} \tag{12}$$

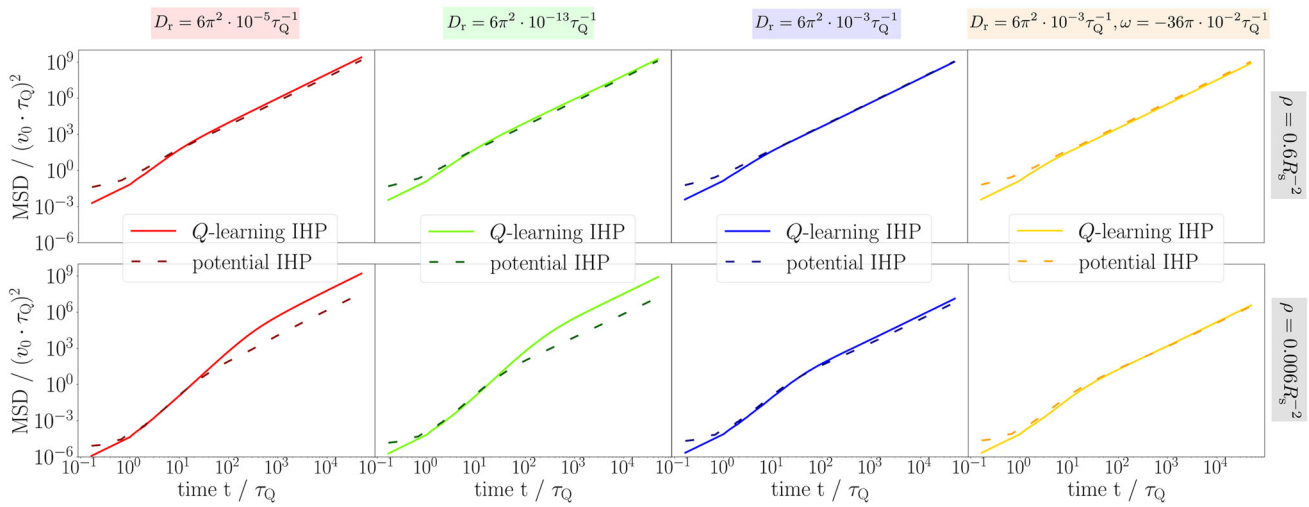

Fig. 7 Q-learning IHP versus potential IHP. We show the MSD for different bath densities ρ, rotational diffusivities D_r and circular frequencies (as labelled) for the Q-learning IHP (solid lines) and the potential IHP (dashed lines). Each curve is averaged over 1000 simulation runs

to find a suitable travel partner in the IHP's observation window $R_s^2\pi$ instantaneously at $t = 0$ can be exactly derived from the binomial distribution and the fraction $\Delta\phi/(2\pi)$ of GO angles. If none is found, the IHP has a new chance after every passing of its perception time τ_Q, which allows us to construct the cumulative probability $p(t)$ to having found the initial travel partner before time t. As, by its nature, $p(t)$ increases as a function of time, the average hitchhiking velocity $v_H(t) = p(t)v_0$ is subject to an effective acceleration, such that the MSD becomes super-ballistic in the second regime $\tau_Q < t \ll \tau$. As the orientation of an ABP randomises after their characteristic persistence time τ, the IHP eventually needs to let go its first companion. This may result in an intermediate deceleration on average, i.e. an effective sub-ballistic MSD regime around $t \approx \tau$ (whether this regime is visible will depend on the form of $p(t)$, which is mainly determined by ρ). Hence, our prediction $p(t)$ provides an upper bound for the probability that the IHP currently has a suitable travel partner at time t, which is the focal point of our theoretical description. To account for the letting go and the subsequent quest for a new travel partner, we notice that, after this procedure has been repeated several times, a good estimate for the long-time hitchhiking velocity in the final ballistic regime for $t \gg \tau$ can be obtained by averaging $p(t)v_0$ over one persistence time of the bath particles. Clearly, this predicted average is larger than $p(0)v_0$ in the first ballistic regime. As indicated in Fig. 6b, the theoretical predictions based on the upper bound and the long-time behaviour provide an excellent representation of the simulated MSD in the corresponding time regimes.

3.3 Comparison to the potential model

Having understood the MSD of the Q-learning IHP, we now turn to the potential IHP and compare the MSD from both models in Fig. 7. In general, we find that the potential IHP has a higher mobility at short times, while the displacement in the initial regime is rather diffusive than ballistic. In later stages, the qualitative behaviour of both models is similar, i.e. the motion is effectively accelerated due to the increasing hitchhiking probability and eventually becomes ballistic, reflecting a successful hitchhiking strategy with persistent trajectories. As the journey proceeds, the Q-learning IHP catches up and prevails over the potential IHP in the long run for most of the parameters.

Recalling the comparison in Sect. 2.4, we identify two main reasons for the differences between the two models observed in the MSD. First, as detailed in appendix B, the probability $p(0)$ to find a suitable (initial or new) travel partner is always larger in the potential model, as the Q-learning IHP can only focus on a single bath particle. This explains the initial advantage of the potential IHP. Second, while the Q-learning IHP can only make discrete decisions separated by the perception time τ_Q, a suitable bath particle always interacts with the potential IHP, which leads to a continuous update of the instantaneous hitchhiking velocity $-\gamma^{-1}\nabla V$. More specifically, the potential IHP gets pulled towards the current position of the ABP, which itself will move with v_0 in a possibly different direction. Assuming that this ABP will still have a favourable orientation, there are three possible scenarios for the IHP motion in the next time step: (i) the pulling force was not directed properly, such that the ABP is out of range and the IHP needs to wait in rest for another travel partner; (ii) its distance to the ABP has changed and the next pull results in a different hitchhiking velocity, or (iii) the IHP ends up within a distance smaller than $R_{\min}$ and thus does not need a kick to remain attached to the ABP, such that it remains at rest to be kicked again in the next time step. The more random nature of the individual hitchhiking velocity (ii,iii) could be responsible for the rather diffusive motion in the initial hitchhiking phase. Moreover, the higher susceptibility of the

potential IHP to let go of its travel partner prematurely explains the catching up of the Q-learning IHP. A specific scenario which could trigger this situation (i) is due to the joint attraction of several suitable travel partners pushing the IHP in an average direction in between such that it will lose connection to all of them. Such a premature detachment (compared to the Q-learning IHP) from a single travel partner can also be caused by the condition $\phi_i \in \Phi_{\mathrm{GO}}$ being evaluated at each time step (instead of each τ_Q).

In how far the initial advantages of the potential model will remain in the long run depends on different parameters, as compared in Fig. 7. We see that the relative long-time performance, i.e. the final average hitchhiking velocity, of the Q-learning IHP surpasses that of the potential IHP for a more persistent and dilute bath. In contrast, the premature detachment of the potential IHP is less disadvantageous both in a bath with low persistence, where the travel partner needs to be changed more frequently anyway (compare the two trajectories in Fig. 3a and c), and in a dense bath, where many suitable new travel partners are present. To better understand the crucial role of density, let us mimic the effects described above by assuming a smaller effective scan radius of the potential IHP (mimicking the effectively looser connection). Following appendix B, the probability $p(0)$ to find a suitable travel partner is then, indeed, smaller for the potential IHP in the limit of a dilute bath, while we find $p(0) \to 1$ for very high densities, irrespective of the (effective) scan radius. In contrast, the limit $p(0) \to \Delta\phi/2\pi$ of Eq. (12) is set solely by the GO angle of the Q-learning IHP, as its focus is limited to a single bath particle.

Finally, let us recall that the Q-learning IHP has been explicitly trained performing its characteristic dynamics, while the potential IHP model requires a transfer of this knowledge into a more realistic physical model. In this light, the performance of the latter is quite remarkable, in particular in a bath of cABPs, where its MSD is larger for all times.

4 Conclusions

In conclusion, we have modelled an intelligent hitchhiking particle (IHP) that gets transported by attaching to and detaching from neighbouring active particles. It learns the rules for optimal transport by a reinforcement learning algorithm, demonstrating that artificial intelligence allows to efficiently generate directed motion from an undirected bath. We have developed a compact analytical framework for describing the drift velocity of the IHP and verified by numerical exploration of its mean-squared displacement (MSD) that the persistence of this motion exceeds that of the bath particles for long times. This ratchet-like behaviour [53–55] of the IHP is reminiscent of the motility-induced drift motion of anisotropic tracers surrounded by ABPs [56–59].

The learned strategy of the IHP depends on three factors. First, the overall bath persistence is reflected by the GO angle. If the bath is highly persistent (small rotational diffusivity and small circular frequency), then it is advantageous to be more selective (small GO angle), as the presumed long travel time with a chosen partner is usually larger than the increase in waiting time. For a weakly persistent bath, it is better to keep the waiting time smaller. This requires the IHP to also take a leap of faith if the travel direction is not ideal (larger GO angle), given the possibility to opt out again with the next decision. Second, the circular motion inherent of cABPs is reflected by the anticipation angle, opposing the circular frequency. Third, a crucial ingredient of our model is the perception time of the Q-learning IHP, which characterises its cognitive limitations. An increase of the perception time at fixed bath parameters requires larger GO and anticipation angles and results in an overall lower performance. These effects can be directly inferred from the scaling implied in Figs. 4, 5 and 6, respectively.

Our minimal model assumes a non-reciprocal coupling between the IHP and the bath particles. While effective non-reciprocal interactions are ubiquitous in nature and are known to generate intriguing non-equilibrium behaviour [60–65], it would also be interesting to model and investigate the reciprocal effects exerted by the IHP on the bath for various reasons. First, such a coupling can also lead to unexpected joint motion of the composite particle. For example, it has been shown that an ABP reacts to its passive cargo by changing its typical dynamics in an activity gradient [66,67]. Second, alternative types of models could describe the IHP as a carrier vesicle [68–70], which learns to selectively uptake and release active particles to get a directional net push in its interior. Third, an intriguing philosophical perspective would be to assess the thermodynamic role played by the IHP. One might argue that an IHP actually extracting its required work from the active bath effectively learns how to act as a Maxwell's demon.

Our model can be generalised in various other ways for future studies. One important step to go beyond the idealised systems considered here is to include more realistic bath particles interacting with each other. The need for anticipating the emerging collective behaviour, which can lead to complex non-equilibrium patterns such as clusters [71] and vortexes [72], would challenge simple learning strategies. One interesting application is the possible transfer of the acquired knowledge to macroscopic bodies such as robots [73,74], which shall perform comparable navigational tasks. As such, an important aspect that should be taken into account is the role of inertia [41], which brings about a memory effect of the past trajectories and might also require more sophisticated learning tools.

Supplementary information

An animation of the agent's hitchhiking behaviour in a bath populated by cABPs is provided as a supplementary video.

4.1 Supplementary video

Below is the link to the electronic supplementary material. Supplementary file 1 (mp4 7858 KB).

Supplementary information The online version contains supplementary material available at https://doi.org/10.1140/epje/s10189-024-00465-0.

Acknowledgements We would like to thank Alexander R. Sprenger, Lorenzo Caprini and Rahul Gupta for stimulating discussions. Funding by the Deutsche Forschungsgemeinschaft (DFG) through the SPP 2265 under grant numbers LO 418/25-2 (HL) and WI 5527/1-2 (RW) is gratefully acknowledged.

Funding Information Open Access funding enabled and organized by Projekt DEAL.

Data availability The data generated in this study are available from the corresponding authors upon reasonable request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix A Theory for the MSD of the Q-learning IHP

While the MSD of the bath particles is exactly known [35], there are also specific rules for hitchhiking. We can thus provide close analytical estimates for the MSD of the IHP, where we focus on the Q-learning IHP in a bath of straightly swimming ABPs. Our considerations rest on two pillars. First, if the IHP has a suitable travel partner, its motion alongside the persistent trajectory of this ABP also occurs with self-propulsion velocity v_0 and is ballistic. Second, as the IHP may also be at rest, the effective hitchhiking velocity $v_{\mathrm{H}}(t)$ becomes an average quantity which depends on time-dependent hitchhiking probability $p_{\mathrm{H}}(t)$, i.e. the probability that *the IHP is currently attached to an ABP*. Such an ABP is chosen for hitchhiking if (i) it is within scan range R_{s} (at a time given by a multiple of τ_Q) when a decision is made, (ii) it is closest to the IHP such that is actually selected as a potential vessel (in case more than one ABPs are within scan range), and (iii) its orientation $\phi_i \in \Phi_{\mathrm{GO}}$ is considered favourable. Hence, the bath density ρ is a crucial control parameter. Moreover, after the IHP has found a suitable ABP, conditions (i) and (ii) will be automatically fulfilled by the same ABP when the next decision is made, while condition (iii) needs to be evaluated again. Because of the possible choice of the IHP to leave a travel partner that runs out of persistence, we eventually also need to take into account the persistence time τ of the bath particles, which will indirectly influence the hitchhiking probability $p_{\mathrm{H}}(t)$.

If we know the hitchhiking probability $p_{\mathrm{H}}(t)$, we can determine the effective hitchhiking velocity

$$v_{\mathrm{H}}(t) = p_{\mathrm{H}}(t) v_0 \tag{A1}$$

right away, such that the resulting displacement follows as

$$x(t) = \int_0^t v_{\mathrm{H}}(t')\mathrm{d}t' \,. \tag{A2}$$

As a directed IHP motion is ensured by its training and the resulting selective choice of travel partners, we conveniently assume here an effective one-dimensional displacement, which directly yields the desired MSD= $x^2(t)$ by squaring the result. In what follows, we discuss the form of $p_{\mathrm{H}}(t)$ in three steps.

As a first step, we want to calculate the probability $p(t)$ that *the IHP has found its first travel partner before or at time t* as an upper bound to $p_{\mathrm{H}}(t)$. The first decision is made at $t = 0$, where the initial probability $p(0)$ can be stated as *there is at least one ABP within range and the selected ABP has a favourable orientation*. In an infinite non-interacting bath of density ρ, the probability that k particles are found within the area $R_{\mathrm{s}}^2\pi$ (surrounding the IHP) is given by the Poisson distribution

$$P_k(\lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \,, \tag{A3}$$

where $\lambda = \rho R_{\mathrm{s}}^2 \pi$ represents the average number of particles. Together with the GO angle $\Delta\phi$ in the definition of Φ_{GO} from Eq. (1), we find the required initial probability as

$$p(0) = \left(1 - P_0(\rho R_{\mathrm{s}}^2 \pi)\right) \frac{\Delta\phi}{2\pi} = \left(1 - e^{-\rho R_{\mathrm{s}}^2 \pi}\right) \frac{\Delta\phi}{2\pi} \,. \tag{A4}$$

For the sake of comparing predictions based on $p(0)$ to our simulation results, we should take into account the finite periodic simulation box of length L, such that the binomial distribution should be used instead of the Poisson distribution. Accordingly, the result analogous to Eq. (A4) is given by Eq. (12), as stated and discussed in the main text and shown in Fig. 6b. For sufficiently large L, both expressions become equivalent. Irrespective of how $p(0)$ is specified, we can obtain the probability

$$p(n) = \sum_{m=0}^{n} (1 - p(0))^m \, p(0) \tag{A5}$$

that *the journey of the IHP has started no later than with the $n+1$st decision* by cumulating the probabilities

$(1-p(0))^m\, p(0)$ that *exactly the $m+1$st decision results in choosing a travel partner for the first time*. Recalling that these decisions are made at $t = n\tau_Q$ with integers $n \geq 0$, we obtain the desired probability $p(t)$ from Eq. (A5) by setting $n = \lfloor t/\tau_Q \rfloor$, where the brackets evaluate to the nearest integer smaller than or equal to the argument.

As a second step, we notice that $p(t)$ is not exactly the hitchhiking probability $p_{\mathrm{H}}(t)$ required in Eq. (A1), because it does not account for the need of the IHP to eventually leave its initial (or current) travel partner and wait again until the next suitable ABP is chosen to continue the trip in the desired direction. Instead, we see from the geometric series in Eq. (A5) that $p(t) \to 1$ for large times, which merely describes that the IHP will always find a suitable travel partner if it just waits long enough. While it is indeed possible to derive an explicit formula for $p_{\mathrm{H}}(t)$, this result will be impracticable for evaluation and we only briefly state the basic idea. The conditional staying probability $p_{\mathrm{S}}(t-t_0)$ is that the IHP that has attached to a bath particle at time t_0 will still make a GO decision at time t can be determined from the Gaussian solution of the angular diffusion process in Eq. (3) with $\omega = 0$, i.e. when the diffusing angle exceeds the limits described by Φ_{GO}. The most notable insight of this exercise is that the relevant change in this probability occurs on a time scale given by τ, which we will refer to below. The appropriate calculation of $p_{\mathrm{H}}(t)$ would then require for an iterative correction of Eq. (A5) by introducing factors of the form $p_{\mathrm{S}}(m\tau_Q)$, accounting for the staying probability after m decisions, which remains nonzero for all decisions after first choosing a travel partner. Likewise, with probability $1-p_{\mathrm{s}}$, the following decision must already account for the possibility of attaching to a new travel partner and so on. This means that a new term of the form of Eq. (A5) must be added after each two decisions, eventually leading to an infinite number of sums.

As a third step, we rather seek here for compact expressions to approximate $p_{\mathrm{H}}(t)$ only using the available result for $p(t)$, which certainly captures the correct behaviour in the early stages of the IHP dynamics. For later times, we utilise the insight that the wait–GO–NO GO sequence described in the previous paragraph will roughly repeat itself after a time span given by τ. Thus, we average our result for $p(t)$ over one persistence time to estimate long-time limit $\bar{p}_\infty := p_{\mathrm{H}}$ the hitchhiking probability as

$$\bar{p}_\infty \approx \frac{1}{\tau}\int_0^\tau p(t)\mathrm{d}t\,. \tag{A6}$$

This integral can be expressed as an analytic series by summation over $p(n)$ Eq. (A5). In summary, we approximate the overall hitchhiking probability as

$$p_{\mathrm{H}}(t) \approx \begin{cases} p(t)\,, & t \lessapprox \tau \\ \bar{p}_\infty\,, & t \gg \tau \end{cases} \tag{A7}$$

and, accordingly, the effective hitchhiking velocity $v_{\mathrm{H}}(t)$ using Eq. (A1). These two cases provide an excellent description of the IHP's MSD in the respective limits. Specifically, we show in Fig. 6 the short-time approximation

$$\mathrm{MSD}_0 = \left(v_0 \int_0^t p(t')\mathrm{d}t'\right)^2 \tag{A8}$$

and the long-time asymptote

$$\mathrm{MSD}_\infty = v_0^2 \bar{p}_\infty^2 t^2\,. \tag{A9}$$

Appendix B Initial hitchhiking probability of the potential IHP

To get a feeling for the performance of the potential IHP relative to the Q-learning IHP, we compare the initial hitchhiking probabilities of both models. For the potential IHP this probability, which we denote by $\tilde{p}(0)$, can be stated as *there is at least one ABP within range and at least one of them has a favourable orientation*. For the reason of obtaining a compact representation, we work with the Poisson distribution from Eq. (A3), appropriate for a sufficiently large system. Doing so, we find

$$\tilde{p}(0) = 1 - \sum_{k=0}^{\infty} P_k(\rho R_{\mathrm{s}}^2\pi)\left(1-\frac{\Delta\phi}{2\pi}\right)^k = 1 - e^{-\rho R_{\mathrm{s}}^2 \pi \frac{\Delta\phi}{2\pi}} \tag{B10}$$

as the complement of having each k bath particles within scan radius, while all of them have an unfavourable orientation. Comparing this result to Eq. (A4), we see that both expressions have the same low-density limit

$$p(0) = \rho R_{\mathrm{s}}^2 \pi \frac{\Delta\phi}{2\pi} + \mathcal{O}(\rho^2) = \tilde{p}(0)\,, \tag{B11}$$

which explicitly depends on the scan radius R_{s}. In general, we see that $\tilde{p}(0)$ is larger than $p(0)$, as can be shown explicitly from the ratio

$$\frac{p(0)}{\tilde{p}(0)} = 1 - \frac{\rho R_{\mathrm{s}}^2 \pi}{2}\left(1-\frac{\Delta\phi}{2\pi}\right) + \mathcal{O}(\rho^2)\,. \tag{B12}$$

Specifically, taking the high-density limits

$$\lim_{\rho\to\infty} p(0) = \frac{\Delta\phi}{2\pi} < 1\,, \quad \lim_{\rho\to\infty} \tilde{p}(0) = 1\,, \tag{B13}$$

the difference between the two models becomes obvious. As the presence of one bath particle with favourable orientation is sufficient for the potential IHP, there surely is a hitchhiking possibility in the high-density limit. In contrast, the Q-learning IHP only has the chance to select a favourable bath particle with a probability determined by the fraction of available GO angles, despite the plethora of suitable travel partners.

References

1. M.C. Marchetti, J.-F. Joanny, S. Ramaswamy, T.B. Liverpool, J. Prost, M. Rao, R.A. Simha, Hydrodynamics of soft active matter. Rev. Mod. Phys. **85**(3), 1143–1189 (2013)
2. J. Elgeti, R.G. Winkler, G. Gompper, Physics of microswimmers-single particle motion and collective behavior: a review. Rep. Prog. Phys. **78**(5), 056601 (2015)
3. E.B. Jacob, I. Becker, Y. Shapira, H. Levine, Bacterial linguistic communication and social intelligence. Trends Microbiol. **12**(8), 366–372 (2004)

4. K.J. Hellingwerf, Bacterial observations: a rudimentary form of intelligence? Trends Microbiol. **13**(4), 152–158 (2005)
5. S. Majumdar, S. Pal, Bacterial intelligence: imitation games, time-sharing, and long-range quantum coherence. J. Cell Commun. Signal. **11**, 281–284 (2017)
6. C. Kaspar, B.J. Ravoo, W.G. van der Wiel, S.V. Wegner, W.H.P. Pernice, The rise of intelligent matter. Nature **594**(7863), 345–355 (2021)
7. J.T. Schlebecker, An informal history of hitchhiking. Historian **20**, 305 (1958)
8. D.J. Smith, Aeroplankton and the need for a global monitoring network. Bioscience **63**, 515 (2013)
9. V. Markgraf, Pollen dispersal in a mountain area. Grana **19**, 127 (1980)
10. E.A. Shinn, D.W. Griffin, D.B. Seba, Atmospheric transport of mold spores in clouds of desert dust. Arch. Environ. Occup. Health **58**, 498 (2003)
11. J. Favet, A. Lapanje, A. Giongo, S. Kennedy, Y.-Y. Aung, A. Cattaneo, A.G. Davis-Richardson, C.T. Brown, R. Kort, H.-J. Brumsack et al., Microbial hitchhikers on intercontinental dust: catching a lift in chad. ISME J. **7**, 850 (2013)
12. A. Muok, A. Briegel, Intermicrobial hitchhiking: how nonmotile microbes leverage communal motility. Trends Microbiol. **29**, 542 (2021)
13. H.-P. Grossart, C. Dziallas, F. Leunert, K.W. Tang, Bacteria dispersal by hitchhiking on zooplankton. Proc. Natl. Acad. Sci. (PNAS) **107**, 11959 (2010)
14. A. Siddique, A. Bhowal, J. Purushothaman, R. Madhusoodhanan, C. Raghunathan, K. Chandra, Seasonal oceanographic phenomenon promotes hitchhiking among the plankton in a coastal marine ecosystem: a tropical perspective. Ecol. Ind. **129**, 107914 (2021)
15. G.K. Howard, H.C. Scott, Predaceous feeding in two common gooseneck barnacles. Science **129**, 717 (1959)
16. Y. Kano, Hitchhiking behaviour in the obligatory upstream migration of amphidromous snails. Biol. Lett. **5**, 465 (2009)
17. L.M. Vieira, A.E. Migotto, *Membraniporopsis tubigera* (osburn, 1940)(bryozoa) on floating substrata: evidence of a dispersal mechanism in the western atlantic. Mar. Biodivers. **45**, 155 (2014)
18. A.E. Nicholson-Jack, J.L. Harris, K. Ballard, K.M. Turner, G.M. Stevens, A hitchhiker guide to manta rays: patterns of association between *Mobula alfredi, m. birostris*, their symbionts, and other fishes in the Maldives. PLoS One **16**, 0253704 (2021)
19. B.M. Norman, S.D. Reynolds, D.L. Morgan, Three-way symbiotic relationships in whale sharks. Pac. Conserv. Biol. **28**, 80 (2021)
20. M. te Vrugt, R. Wittkowski, A review of active matter reviews, (2024), arXiv Preprint at arXiv:2405.15751
21. S. Ramaswamy, The mechanics and statistics of active matter. Annu. Rev. Condens. Matter Phys. **1**(1), 323–345 (2010)
22. C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, G. Volpe, Active particles in complex and crowded environments. Rev. Mod. Phys. **88**(4), 045006 (2016)
23. F. Sagués Mestre, *Colloidal Active Matter: Concepts, Experimental Realizations, and Models* (CRC Press, 2022)
24. B. ten Hagen, S. van Teeffelen, H. Löwen, Brownian motion of a self-propelled particle. J. Phys. Condens. Matter **23**(19), 194119 (2011)
25. C. Kurzthaler, S. Leitmann, T. Franosch, Intermediate scattering function of an anisotropic active Brownian particle. Sci. Rep. **6**(1), 36702 (2016)
26. J. Jeggle, J. Stenhammar, R. Wittkowski, Pair-distribution function of active brownian spheres in two spatial dimensions: simulation results and analytic representation. J. Chem. Phys. **152**(19) (2020)
27. M. Caraglio, T. Franosch, Analytic solution of an active Brownian particle in a harmonic well. Phys. Rev. Lett. **129**(15), 158001 (2022)
28. S.C. Takatori, W. Yan, J.F. Brady, Swim pressure: stress generation in active matter. Phys. Rev. Lett. **113**(2), 028103 (2014)
29. J. O'Byrne, A. Solon, J. Tailleur, Y. Zhao, An introduction to motility-induced phase separation, in Out-of-equilibrium Soft Matter. (The Royal Society of Chemistry, 2023), pp. 107–150
30. L. Caprini, U. Marini Bettolo Marconi, A. Puglisi, Spontaneous velocity alignment in motility-induced phase separation. Phys. Rev. Lett. **124**(7), 078001 (2020)
31. L. Caprini, U. Marini Bettolo Marconi, A. Puglisi, H. Löwen, Entropons as collective excitations in active solids. J. Chem. Phys. **159**(4) (2023)
32. F. Kümmel, B. ten Hagen, R. Wittkowski, I. Buttinoni, R. Eichhorn, G. Volpe, H. Löwen, C. Bechinger, Circular motion of asymmetric self-propelling particles. Phys. Rev. Lett. **110**(19), 198302 (2013)
33. A.I. Campbell, R. Wittkowski, B. ten Hagen, H. Löwen, S.J. Ebbens, Helical paths, gravitaxis, and separation phenomena for mass-anisotropic self-propelling colloids: Experiment versus theory. J. Chem. Phys. **147**(8) (2017)
34. C. Scholz, S. Jahanshahi, A. Ldov, H. Löwen, Inertial delay of self-propelled particles. Nat. Commun. **9**, 5156 (2018)
35. S. van Teeffelen, H. Löwen, Dynamics of a Brownian circle swimmer. Phys. Rev. E **78**, 020101(R) (2008)
36. F.J. Sevilla, Diffusion of active chiral particles. Phys. Rev. E **94**(6), 062120 (2016)
37. S. Jahanshahi, H. Löwen, B. ten Hagen, Brownian motion of a circle swimmer in a harmonic trap. Phys. Rev. E **95**(2), 022606 (2017)
38. O. Chepizhko, T. Franosch, Ideal circle microswimmers in crowded media. Soft Matter **15**(3), 452–461 (2019)
39. R. Wittkowski, H. Löwen, Self-propelled Brownian spinning top: dynamics of a biaxial swimmer at low Reynolds numbers. Phys. Rev. E **85**(2), 021406 (2012)
40. M. Bär, R. Großmann, S. Heidenreich, F. Peruani, Self-propelled rods: insights and perspectives for active matter. Ann. Rev. Conden. Matter Phys. **11**(1), 441–466 (2020)
41. H. Löwen, Inertial effects of self-propelled particles: from active Brownian to active langevin motion. J. Chem. Phys. **152**(4) (2020)
42. A.R. Sprenger, S. Jahanshahi, A.V. Ivlev, H. Löwen, Time-dependent inertia of self-propelled particles: the Langevin rocket. Phys. Rev. E **103**(4), 042601 (2021)
43. A.R. Dulaney, J.F. Brady, Machine learning for phase behavior in active matter systems. Soft Matter **17**(28), 6808–6816 (2021)
44. J. Colen, M. Han, R. Zhang, S.A. Redford, L.M. Lemma, L. Morgan, P.V. Ruijgrok, R. Adkins, Z. Bryant, Z.

Dogic et al., Machine learning active-nematic hydrodynamics. Proc. Natl. Acad. Sci. (PNAS) **118**(10), 2016708118 (2021)
45. E. Schneider, H. Stark, Optimal steering of a smart active particle. Europhys. Lett. **127**(6), 64003 (2019)
46. F. Cichos, K. Gustavsson, B. Mehlig, G. Volpe, Machine learning for active matter. Nat. Mach. Intell. **2**(2), 94–103 (2020)
47. P.A. Monderkamp, F.J. Schwarzendahl, M.A. Klatt, H. Löwen, Active particles using reinforcement learning to navigate in complex motility landscapes. Mach. Learn. Sci. Technol. **3**, 045024 (2022)
48. M. Nasiri, H. Löwen, B. Liebchen, Optimal active particle navigation meets machine learning. Europhys. Lett. **142**(1), 17001 (2023)
49. J. Grauer, F.J. Schwarzendahl, H. Löwen, B. Liebchen, Optimizing collective behavior of communicating active particles with machine learning. Mach. Learn. Sci. Technol. **5**(1), 015014 (2024)
50. R.S. Sutton, *Reinforcement Learning: An Introduction* (A Bradford Book, 2018)
51. B. Jang, M. Kim, G. Harerimana, J.W. Kim, Q-learning algorithms: a comprehensive classification and applications. IEEE Access **7**, 133653 (2019)
52. B. Mahesh, Machine learning algorithms—a review. Int. J. Sci. Res. **9**, 381 (2020)
53. P. Hänggi, F. Marchesoni, F. Nori, Brownian motors. Ann. Phys. **517**(1–3), 51–70 (2005)
54. R. Di Leonardo, L. Angelani, D. Dell'Arciprete, G. Ruocco, V. Iebba, S. Schippa, M.P. Conte, F. Mecarini, F. De Angelis, E. Di Fabrizio, Bacterial ratchet motors. Proc. Natl. Acad. Sci. (PNAS) **107**(21), 9541–9545 (2010)
55. C.J.O. Reichhardt, C. Reichhardt, Ratchet effects in active matter systems. Annu. Rev. Conden. Matter Phys. **8**(1), 51–75 (2017)
56. A. Kaiser, A. Peshkov, A. Sokolov, B. ten Hagen, H. Löwen, I.S. Aranson, Transport powered by bacterial turbulence. Phys. Rev. Lett. **112**(15), 158101 (2014)
57. S.A. Mallory, C. Valeriani, A. Cacciuto, Curvature-induced activation of a passive tracer in an active bath. Phys. Rev. E **90**(3), 032309 (2014)
58. M. Knežević, H. Stark, Effective Langevin equations for a polar tracer in an active bath. New J. Phys. **22**(11), 113025 (2020)
59. O. Granek, Y. Kafri, J. Tailleur, Anomalous transport of tracers in active baths. Phys. Rev. Lett. **129**(3), 038001 (2022)
60. A.V. Ivlev, J. Bartnick, M. Heinen, C.-R. Du, V. Nosenko, H. Löwen, Statistical mechanics where newton's third law is broken. Phys. Rev. X **5**(1), 011035 (2015)
61. J. Agudo-Canalejo, R. Golestanian, Active phase separation in mixtures of chemically interacting particles. Phys. Rev. Lett. **123**(1), 018101 (2019)
62. Z. You, A. Baskaran, M.C. Marchetti, Nonreciprocity as a generic route to traveling states. Proc. Natl. Acad. Sci. (PNAS) **117**(33), 19767–19772 (2020)
63. D. Levis, A. Diaz-Guilera, I. Pagonabarraga, M. Starnini, Flocking-enhanced social contagion. Phys. Rev. Res. **2**(3), 032056 (2020)
64. J. Múgica, J. Torrents, J. Cristín, A. Puy, M.C. Miguel, R. Pastor-Satorras, Scale-free behavioral cascades and effective leadership in schooling fish. Sci. Rep. **12**(1), 10783 (2022)
65. L. Gómez-Nava, R. Bon, F. Peruani, Intermittent collective motion in sheep results from alternating the role of leader and follower. Nat. Phys. **18**(12), 1494–1501 (2022)
66. H.D. Vuijk, H. Merlitz, M. Lang, A. Sharma, J.-U. Sommer, Chemotaxis of cargo-carrying self-propelled particles. Phys. Rev. Lett. **126**(20), 208102 (2021)
67. P.L. Muzzeddu, É. Roldán, A. Gambassi, A. Sharma, Taxis of cargo-carrying microswimmers in traveling activity waves (a). Europhys. Lett. **142**(6), 67001 (2023)
68. G. Kokot, H.A. Faizi, G.E. Pradillo, A. Snezhko, P.M. Vlahovska, Spontaneous self-propulsion and nonequilibrium shape fluctuations of a droplet enclosing active particles. Commun. Phys. **5**(1), 91 (2022)
69. S. Uplap, M.F. Hagan, A. Baskaran, Design principles for transporting vesicles with enclosed active particles (a). Europhys. Lett. **143**(6), 67001 (2023)
70. R. Wittmann, P.A. Monderkamp, H. Löwen, Statistics of carrier-cargo complexes. Phys. Rev. E **107**(6), 064602 (2023)
71. I. Theurkauff, C. Cottin-Bizonne, J. Palacci, C. Ybert, L. Bocquet, Dynamic clustering in active colloidal suspensions with chemical signaling. Phys. Rev. Lett. **108**(26), 268303 (2012)
72. H.H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R.E. Goldstein, H. Löwen, J.M. Yeomans, Meso-scale turbulence in living fluids. Proc. Natl. Acad. Sci. (PNAS) **109**(36), 14308–14313 (2012)
73. Y. Zhao, Y. Chi, Y. Hong, Y. Li, S. Yang, J. Yin, Twisting for soft intelligent autonomous robot in unstructured environments. Proc. Natl. Acad. Sci. (PNAS) **119**(22), 2200265119 (2022)
74. L.M. Kamp, M. Zanaty, A. Zareei, B. Gorissen, R.J. Wood, K. Bertoldi, Reprogrammable sequencing for physically intelligent under-actuated robots, (2024). arXiv Preprint at arXiv:2409.03737