# Synergizing Deep Reinforcement Learning and Biological Pursuit Behavioral Rule for Robust and Interpretable Navigation

**Kazushi Tsutsui** [*1,2] **Kazuya Takeda** [3,1] **Keisuke Fujii** [1,2,4,5]

## Abstract

Integrating theoretical models within machine learning models holds considerable promise for constructing efficient and robust models. In biology, however, integration can be challenging because the behavioral rules described by theoretical models are not necessarily invariant, in contrast to problems in physics. Here, we propose a hybrid architecture that hierarchically integrates a biological pursuit model into deep reinforcement learning. Our approach facilitates seamless agent mode switching and rule-based action selection, demonstrating efficient navigation in a predator-prey environment. Interestingly, our results parallel the hunting behavior observed in nature, offering novel insights into biology. As our framework can be integrated with existing hybrid or gray box models, it paves the way for further exploration in this exciting intersection of machine learning and biology.

## 1. Introduction

Hybrid or gray-box modeling, which blends learning-based and theory-based approaches, has demonstrated improved efficiency and robustness, with additional benefits to model interpretability due to the inclusion of domain knowledge (Takeishi & Kalousis, 2021; Zhang et al., 2022; Likmeta et al., 2020).

Moreover, integrating learning-based and theory-based approaches can serve as a scientific tool to deepen our understanding of real-world phenomena (Fujii et al., 2021). This can be especially challenging in biology, where theory-based models are not necessarily invariant, in contrast to physical laws. For instance, behavioral rules of organisms described as theoretical models may alter due to various factors, such as learning through interaction with the environment.

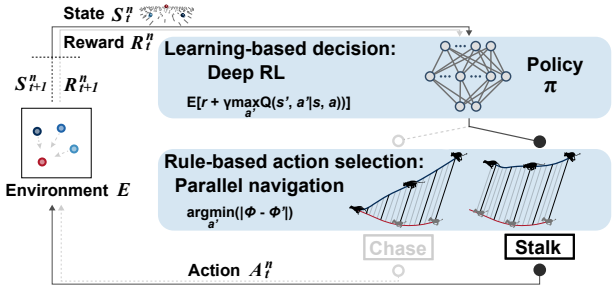


*Figure 1.* Hierarchical agent architecture. This hybrid model composes of deep RL (upper) and a biological pursuit model (lower)

.

Nevertheless, there are commonalities in animal behavior across many species that have been optimized over lengthy evolutionary processes (Collett & Land, 1978; Brighton et al., 2017; Brighton & Taylor, 2019; Ghose et al., 2006; Kane et al., 2015; Tsutsui et al., 2020). Therefore, constructing hybrid models, in which commonalities are given by theory-based models and differences are described by learning-based models, might enhance our understanding of biodiversity.

In this paper, we propose a hybrid architecture that hierarchically integrates a biological pursuit model into deep RL. This allows for seamless agent mode switching and rule-based action selection, facilitating efficient and robust navigation. We demonstrate that our hierarchical agents can successfully balance reward acquisition and travel costs in a predator-prey environment. Interestingly, our results are reminiscent of hunting behaviors observed in large terrestrial mammals like lions, suggesting that our proposed model could offer novel insights into biology.

The main contributions of the present study are as follows: (1) We propose a framework for modeling the hierarchical decision-making process of organisms. (2) Methodologically, we achieve navigation with acceleration/deceleration

[1]Graduate School of Informatics, Nagoya University, Nagoya, Japan [2]Institute for Advanced Research, Nagoya University, Nagoya, Japan [3]Institutes of Innovation for Future Society, Nagoya University, Nagoya, Japan [4]RIKEN Center for Advanced Intelligence Project, Tokyo, Japan [5]PRESTO, Japan Science and Technology Agency, Tokyo, Japan. Correspondence to: Kazushi Tsutsui <k.tsutsui6@gmail.com>.

*Accepted after peer-review at the 1st workshop on Synergy of Scientific and Machine Learning Modeling*, SynS & ML ICML, Honolulu, Hawaii, USA. July, 2023. 

in the form of approximating a biological pursuit model that assumes movement at a constant speed. (3) Our hierarchical model shows the potential to overcome the constraints of existing biological models, which are often limited to describing behavior in ideal environments (Fawcett et al., 2014), and to depict biological behavior in more realistic scenarios.

## 2. Methods

In this section, we provide an overview of the theoretical basis of RL, deep RL, and the biological pursuit model, and describe our proposed agent architecture that integrates them hierarchically.

### 2.1. Reinforcement learning

We consider a sequential decision-making setting in which an agent interacts with an environment $\mathcal{E}$ in a sequence of observations, actions, and rewards. At each time-step $t$, the agent observes a state $s_t \in \mathcal{S}$ and selects an action $a_t$ from a discrete set of actions $\mathcal{A} = \{1, 2, \ldots, |\mathcal{A}|\}$. One time step later, in part as a consequence of its action, the agent receives a reward, $r_{t+1} \in \mathcal{R}$, and moves itself to a new state $s_{t+1}$. In the MDP, the agent learns policies that depend upon these sequences. The goal of the agent is to maximize the expected discounted return (Sutton & Barto, 2018). The discounted return $R_t$ was defined as $\sum_{k=0}^{T} \gamma^k r_{t+k+1}$, where $\gamma \in [0, 1]$ is a parameter called the discount rate that determines the present value of future rewards, and $T$ is the time step at which the task terminates. The Q-function or action-value function is defined as $Q^\pi(s, a) = \mathbb{E}_\pi[R_t | s_t = s, a_t = a]$, where $\pi$ is a policy mapping states to actions. The optimal action-value function $Q^\star(s, a)$ is then defined as the maximum expected discounted return achievable by following any strategy, after observing some state $s$ and then taking some action $a$, $Q^\star(s, a) = \max_\pi \mathbb{E}[R_t | s_t = s, a_t = a, \pi]$. The optimal action-value function can be computed recursively obeying the Bellman equation:

$$Q^\star(s, a) = \mathbb{E}_{s' \sim \mathcal{E}}[r + \gamma \max_{a'} Q^\star(s', a' | s, a)],$$

where $s'$ and $a'$ are the state and action at the next time-step, respectively.

### 2.2. Deep Q-Networks (DQN) and its extensions

Deep Q-Networks (DQN) (Mnih et al., 2015) is a model-free RL algorithm for discrete action spaces. In DQN, deep neural networks and RL were successfully combined to approximate the action values for a given state $s_t$. DQN has been an important milestone, but several limitations of this algorithm are known, and many extensions have been proposed. Here, we briefly introduce three extensions, in order, that have improved overall performance. Double DQN (DDQN) is an algorithm that applies the double Q-learning method to DQN (Van Hasselt et al., 2016). For DDQN, the target network in the DQN architecture was used as the second value function. Prioritized experience replay is a method aimed at enhancing learning efficiency and efficacy (Schaul et al., 2015). For prioritized replay, the probability of sampling from the replay buffer is relative to the absolute temporal-difference (TD) error. Dueling network is a neural network architecture designed for value-based algorithms (Wang et al., 2016). This architecture features two streams of computation, the value and advantage streams, sharing a common encoder, and is merged by an aggregation module that produces an estimate of the state-action value function.

### 2.3. Biological pursuit behavioral rule

Chase and escape behaviors are crucial for survival in many species, and therefore are efficient and robust by necessity (Evans et al., 2019). These behaviors are thought as complex phenomena in which two or more agents interact, yet many studies have shown that the rules of behavior (e.g., which direction to move at each time in a given situation) can be described by relatively simple mathematical models consisting of the current state (e.g., positions and velocities).

Specifically, many predators are observed to maintain a constant geographic direction in relation to their prey during approach (Ghose et al., 2006). Because of its geometry, this pursuit strategy is referred to as parallel navigation (sometimes called constant bearing or constant absolute target direction). This can be mathematically described as follows:

$$\beta = \arcsin\left(\frac{|v_t| \sin\alpha}{|v_p|}\right),$$

where $\beta$, $v_p$, $v_t$, and $\alpha$ are the movement angle relative to prey, the velocity of the predator, the velocity of the target, and the angle between the velocity of target and a vector that points from the target to the predator (Ghose et al., 2006; Kane et al., 2015; Tsutsui et al., 2020). This strategy can also be expressed in the form of a guidance law called proportional navigation as follows: $\dot{\gamma} = N\dot{\lambda}$, where $\dot{\gamma}$, $N$, and $\dot{\lambda}$ are turning rate, rotation rate of the line-of-sight (i.e., target direction), and the navigation constant (Brighton et al., 2017). The parallel navigation can lead predators to move along local shortest paths and guarantees eventual interception of a target (Nahin, 2012). The widespread adoption of this strategy by predators is thought to be linked to the use of perceptual invariants. That is, by simply nullifying the rate of change of the visual angle to the target, predators can pursue a moving target (Brighton et al., 2017). However, this simplistic model is not designed as a controllable model, as described below.

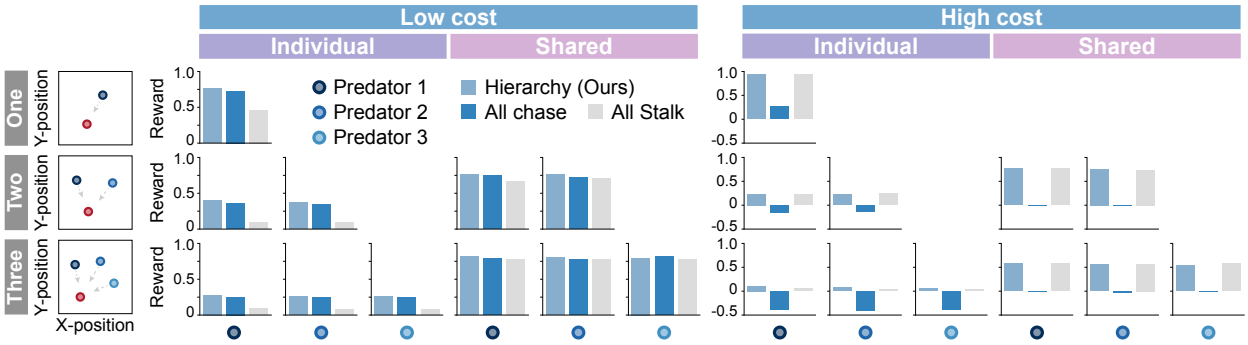


*Figure 2.* Comparison of cumulative rewards of predator agents

### 2.4. Hierarchical agents with integrated architecture

In this paper, we propose the integration of deep RL and parallel navigation. The pursuit model consists of the current information and thus can be smoothly incorporated into standard deep RL methods for a finite MDP in which each sequence is a distinct state. However, the pursuit model has the limitation that it assumes pursuit at approximately constant speed and calculates only the direction of movement. This property does not control the navigation of agents in complex environments with acceleration and deceleration. Also, from a biological perspective, this limitation narrows the scope of analysis, and indeed previous studies on modeling predatory attacks are biased toward aerial predators, which have a relatively constant speed.

Therefore, we propose a hierarchical agent structure that overcomes these problems. Our hierarchical predator agent consists of an upper layer that determines the magnitude (mode) of movement and a lower layer that determines the direction of movement (Fig. 1). The magnitude and direction are determined by deep RL and the biological pursuit model, respectively. We aimed to construct a biologically plausible (or considered to be more amenable to interpretation) simulation environment, and modeled an agent with independent learning, in which each agent treats the other agents as part of the environment (Tan, 1993). That is, in contrast to previous studies (Silver et al., 2017; Christianos et al., 2020; Lowe et al., 2017), our agents did not have access to models of the environment and observations and policies of other agents. For each agent $n$, the policy $\pi^n$ was represented by a neural network and optimized in the framework of DQN including DDQN, prioritized replay, and dueling architecture. The loss function of each agent takes the form:

$$\mathcal{L}_i(\theta_i, \eta_i, \xi_i) = \mathbb{E}_{s,a,r',s' \sim \mathcal{P}(\mathcal{D})}[(y_i - Q(s, a; \theta_i, \eta_i, \xi_i))^2],$$

where

$$y_i = r + \gamma Q(s', \arg\max_{a'} Q(s', a'; \theta_i, \eta_i, \xi_i | s, a; \theta_i^-, \eta_i^-, \xi_i^-),$$

and $\mathcal{P}(\cdot)$ represents prioritized sampling; $\theta$ denotes the parameters of the common layers, whereas $\eta$ and $\xi$ are the parameters of the layers of the value and advantage streams, respectively; $\theta^-$, $\eta^-$, and $\xi^-$ denote those of the target-network. For simplicity, we omitted the agent index $n$ in these equations. The inputs to the neural network consist of subjectively available information about the environment and the outputs are the mode, namely, chase or stalk.

We then determine the direction of movement using biological pursuit model. However, as noted above, this model assumes pursuit at a nearly constant speed and could be inappropriate for direct application to this study, which involves acceleration and deceleration. Therefore, we designed to choose an action that approximates parallel navigation as follows:

$$a = \arg\min_{a'}(\phi - \phi'),$$

where $\phi$ and $\phi'$ denote absolute target direction at a current and next time step, respectively. This approximation allows the predator to choose actions that reduce the distance while keeping the prey approximately in the same direction, leading to efficient and robust navigation in physics-based environments. The final outputs (actions) are accelerations in 12 directions every 30 degrees in the relative coordinate system to the prey, which was determined based on previous findings on ecology (Wilson et al., 2018). We searched for an action that minimizes the change in the absolute target direction from -90 to 90 degrees on the approaching side.

For the prey agent, the policy $\pi^n$ is represented by a neural network and directly optimized action selection. The actions were selected from accelerations in 12 directions, as for the predators.

## 3. Experiments

### 3.1. Environment

Agents are tasked with a predator-prey interaction in a physics-based environment, which is a two-dimensional world with continuous space and discrete time. This environment was constructed by modifying the predator-prey environment in MAPE (Lowe et al., 2017). The position of each agent was calculated by integrating the accelera-

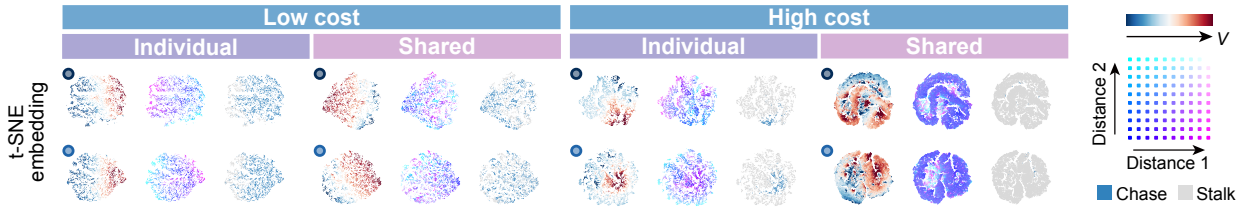


*Figure 3.* Embedding of internal representations underlying efficient navigation

tion (i.e., selected action) twice with the Euler method, and viscous resistance proportional to velocity was considered. Modifications include constraining the movable space to the range of -1 to 1 on the $x$ and $y$ axes, all agent (predator/prey) disk diameters were set to 0.1, landmarks (obstacles) were eliminated, and predator-to-predator contact was ignored for simplicity. The predator(s) was rewarded for capturing the prey (+1), namely for contact between the disks, and punished for moving out of the area (-1). In addition, accelerating in the chase mode (120% of prey) required a cost (negative reward) due to fatigue (see below) and that in the stalk mode (60% of prey) was assumed to be negligible. The prey was punished for being captured by the predator or for moving out of the area (-1), and no cost was considered for acceleration. The predator and prey were represented as a red and blue disk, respectively, and the play area was represented as a black square surrounding them. The time step was 0.1 s and the time limit in each episode was set to 30 s. The initial position for each episode was randomly selected from a range of -0.5 to 0.5 on both the $x$ and $y$ axes.

### 3.2. Experimental conditions

We selected the number of predators, cost (negative reward) of chase, and prey (positive reward) sharing as experimental conditions, based on ecological findings (Bailey et al., 2013; Lang & Farine, 2017). For the number of predators, three conditions were set: 1 (one), 2 (two), and 3 (three). In all these conditions, the number of preys was set as 1. For the cost of chase, two conditions were set: 0.01 (low) and 0.1 (high) for the acceleration exerted by the predator in the chase mode. The cost in the stalk mode was assumed to be negligibly small, namely 0, in both conditions. For the prey sharing, two conditions were set: with sharing (shared), in which all predators were rewarded when a predator catches the prey, and without sharing (individual), in which a predator was rewarded only when it catches prey by itself. In total, there were 10 conditions.

### 3.3. Training details

The neural network is composed of four layers. The inputs to the neural network are the positions of oneself and others in the absolute coordinate system ($x$- and $y$-positions) and the positions and velocities of oneself and others in the relative coordinate system ($u$- and $v$-positions and $u$- and $v$-velocities), which were determined based on findings in ethology (Brighton et al., 2017) and neuroscience (O'Keefe & Dostrovsky, 1971). We assumed that delays in sensory processing are compensated for by estimation of motion of self (Wolpert et al., 1998; Kawato, 1999) and others (Tsutsui et al., 2021) and the current information at each time was used as input as is. The model was trained for $10^6$ episodes, and the network parameters were copied to the target-network every 2000 episodes. The replay memory size was $10^4$, the minibatch size was 32, the learning rate was $10^{-6}$, and the discount factor $\gamma$ was 0.9. We used an $\varepsilon$-greedy policy as the behavior policy $\pi^n$, which chooses a random action with probability $\varepsilon$ or an action according to the optimal $Q$ function $\arg\max_{a\in\mathcal{A}} Q^\star(s,a)$ with probability $1-\varepsilon$. In this study, $\varepsilon$ was annealed linearly from 1 to 0.1 over the first $10^4$ time steps and fixed at 0.1 thereafter.

### 3.4. Evaluation

The model performance was evaluated using the trained model. During the evaluation, $\varepsilon$ was set to 0 in the predator agents, and each predator agent took greedy actions. Since the focus of this study is on predator agents, $\varepsilon$ was left at 0.1 in the prey agent for behavioral consistency with the training. In the evaluation, we simulated 1000 episodes in each condition.

We first show the mean cumulative rewards of predators in each episode for each condition (Fig. 2). This figure includes the mean cumulative rewards when a predator agent always selected to chase or to stalk, as baselines for comparison. At this time, we fixed for the other predators' and prey's behaviors. As shown in this figure, our hierarchical model outperformed the baselines in many cases (16 of 22 panels), and seems to be fairly good performance considering that the prey agent was optimized for the hierarchical predator agents. In addition, even in cases where the value was not the highest, it was close to the highest value (96% of the highest value on average). Furthermore, each predator agent learned independently, yet the trend in results was consistent among the predators within conditions. This indicates that the hierarchical model learned stably and behaved efficiently depending on the task, regardless of the experimental conditions.

To better understand how such efficient navigation is ac-

complished, we analyzed predators' internal representations using t-SNE (Van der Maaten & Hinton, 2008). We visualized the last hidden layers of the state streams in the policy network of predator agents in the two-predator condition (Fig. 3). Coloring the embedding, we found that each state value corresponds with the distance between predators and prey. For example, in the individual condition, the representation of predator 1 showed a high state value for the cyan-colored state. Conversely, the representation of predator 2 showed a high state value for the magenta-colored state. This result is consistent with intuition, since under the condition the two predators could be considered rivals competing with each other for prey. Additionally, the difference in the mode selection tendencies under the low and high cost conditions resonates with predators in nature, where predators that rapidly exhaust their metabolic resources during a chase tend to first stalk their prey, slowly approaching their prey to decrease chase distance and time (Mech, 1970).

## 4. Conclusion

In this study, we introduced a hierarchical agent that integrates deep RL with a biological mathematical model, and demonstrated its usefulness in a multi-agent interactive environment, as well as its potential for understanding real-world organisms such as wild animals. While incorporating functions such as obstacle avoidance presents challenges, the introduction of deep RL integration for each behavioral module (e.g., (Johannink et al., 2019)) could be beneficial. Our framework is compatible with existing hybrid or gray box models, underscoring the value of further research.

## Acknowledgements

This work was supported by JSPS KAKENHI (Grant Numbers 21H04892, 21H05300, and 22K17673), JST PRESTO (JPMJPR20CA), and the Program for Promoting the Enhancement of Research Universities.

## Broader Impact Statement

The findings from this study hold significant potential for broader impacts across both scientific and societal realms. From a scientific perspective, the integration of deep reinforcement learning with biological mathematical models as introduced in our research provides a novel approach in understanding complex multi-agent environments. This can revolutionize how we study not only artificial intelligence systems, but also biological organisms and their interactions. From a societal standpoint, the implications of this study extend to practical applications like crowd management, traffic systems, or robotic coordination. Implementing this approach can lead to more efficient systems, benefiting society as a whole. However, there may also be negative implications. For example, the misuse of such technologies may potentially lead to intrusive surveillance systems or autonomous weapons. Hence, it is crucial to accompany further development in this field with strict ethical guidelines and regulations. This statement is a brief overview, and the actual effects can be more extensive and varied. Therefore, we emphasize the importance of continual dialogue between researchers, ethicists, policymakers, and society to anticipate, evaluate, and respond to the broader impacts of this technology.

## References

Bailey, I., Myatt, J. P., and Wilson, A. M. Group hunting within the carnivora: physiological, cognitive and environmental influences on strategy and cooperation. *Behavioral ecology and sociobiology*, 67(1):1–17, 2013.

Brighton, C. H. and Taylor, G. K. Hawks steer attacks using a guidance system tuned for close pursuit of erratically manoeuvring targets. *Nature communications*, 10(1):1–10, 2019.

Brighton, C. H., Thomas, A. L., and Taylor, G. K. Terminal attack trajectories of peregrine falcons are described by the proportional navigation guidance law of missiles. *Proceedings of the National Academy of Sciences*, 114 (51):13495–13500, 2017.

Christianos, F., Schäfer, L., and Albrecht, S. Shared experience actor-critic for multi-agent reinforcement learning. *Advances in neural information processing systems*, 33: 10707–10717, 2020.

Collett, T. S. and Land, M. How hoverflies compute interception courses. *Journal of comparative physiology*, 125 (3):191–204, 1978.

Evans, D. A., Stempel, A. V., Vale, R., and Branco, T. Cognitive control of escape behaviour. *Trends in cognitive sciences*, 23(4):334–348, 2019.

Fawcett, T. W., Fallenstein, B., Higginson, A. D., Houston, A. I., Mallpress, D. E., Trimmer, P. C., and McNamara, J. M. The evolution of decision rules in complex environments. *Trends in cognitive sciences*, 18(3):153–161, 2014.

Fujii, K., Takeishi, N., Tsutsui, K., Fujioka, E., Nishiumi, N., Tanaka, R., Fukushiro, M., Ide, K., Kohno, H., Yoda, K., et al. Learning interaction rules from multi-animal trajectories via augmented behavioral models. *Advances in Neural Information Processing Systems*, 34:11108–11122, 2021.

Ghose, K., Horiuchi, T. K., Krishnaprasad, P., and Moss, C. F. Echolocating bats use a nearly time-optimal strategy to intercept prey. *PLoS biology*, 4(5):e108, 2006.

Johannink, T., Bahl, S., Nair, A., Luo, J., Kumar, A., Loskyll, M., Ojea, J. A., Solowjow, E., and Levine, S. Residual reinforcement learning for robot control. In *2019 International Conference on Robotics and Automation (ICRA)*, pp. 6023–6029. IEEE, 2019.

Kane, S. A., Fulton, A. H., and Rosenthal, L. J. When hawks attack: animal-borne video studies of goshawk pursuit and prey-evasion strategies. *Journal of Experimental Biology*, 218(2):212–222, 2015.

Kawato, M. Internal models for motor control and trajectory planning. *Current opinion in neurobiology*, 9(6):718–727, 1999.

Lang, S. D. and Farine, D. R. A multidimensional framework for studying social predation strategies. *Nature ecology & evolution*, 1(9):1230–1239, 2017.

Likmeta, A., Metelli, A. M., Tirinzoni, A., Giol, R., Restelli, M., and Romano, D. Combining reinforcement learning with rule-based controllers for transparent and general decision-making in autonomous driving. *Robotics and Autonomous Systems*, 131:103568, 2020.

Lowe, R., Wu, Y. I., Tamar, A., Harb, J., Pieter Abbeel, O., and Mordatch, I. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30, 2017.

Mech, L. D. *The Wolf: The ecology and behaviour of an endangered species*. University of Minnesota Press, USA, 1970.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540): 529–533, 2015.

Nahin, P. J. *Chases and Escapes: The Mathematics of Pursuit and Evasion*. Princeton University Press, USA, 2012.

O'Keefe, J. and Dostrovsky, J. The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 1971.

Schaul, T., Quan, J., Antonoglou, I., and Silver, D. Prioritized experience replay. *arXiv preprint arXiv:1511.05952*, 2015.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

Sutton, R. S. and Barto, A. G. *Reinforcement learning: An introduction*. MIT press, 2018.

Takeishi, N. and Kalousis, A. Physics-integrated variational autoencoders for robust and interpretable generative modeling. *Advances in Neural Information Processing Systems*, 34:14809–14821, 2021.

Tan, M. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the tenth international conference on machine learning*, pp. 330–337, 1993.

Tsutsui, K., Shinya, M., and Kudo, K. Human navigational strategy for intercepting an erratically moving target in chase and escape interactions. *Journal of motor behavior*, 52(6):750–760, 2020.

Tsutsui, K., Fujii, K., Kudo, K., and Takeda, K. Flexible prediction of opponent motion with internal representation in interception behavior. *Biological cybernetics*, 115 (5):473–485, 2021.

Van der Maaten, L. and Hinton, G. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

Van Hasselt, H., Guez, A., and Silver, D. Deep reinforcement learning with double q-learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., and Freitas, N. Dueling network architectures for deep reinforcement learning. In *International conference on machine learning*, pp. 1995–2003. PMLR, 2016.

Wilson, A. M., Hubel, T. Y., Wilshin, S. D., Lowe, J. C., Lorenc, M., Dewhirst, O. P., Bartlam-Brooks, H. L., Diack, R., Bennitt, E., Golabek, K. A., et al. Biomechanics of predator–prey arms race in lion, zebra, cheetah and impala. *Nature*, 554(7691):183–188, 2018.

Wolpert, D. M., Miall, R. C., and Kawato, M. Internal models in the cerebellum. *Trends in cognitive sciences*, 2 (9):338–347, 1998.

Zhang, G., Yu, Z., Jin, D., and Li, Y. Physics-infused machine learning for crowd simulation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2439–2449, 2022.