A Biologically Inspired Computational Model of Time Perception

Inês Lourenço, *Student Member, IEEE*, Robert Mattila, *Student Member, IEEE*, Rodrigo Ventura, *Member, IEEE*, and Bo Wahlberg, *Fellow, IEEE*

***Abstract*—Time perception—how humans and animals perceive the passage of time—forms the basis for important cognitive skills, such as decision making, planning, and communication. In this work, we propose a framework for examining the mechanisms responsible for time perception. We first model neural time perception as a combination of two known timing sources: internal neuronal mechanisms and external (environmental) stimuli, and design a decision-making framework to replicate them. We then implement this framework in a simulated robot. We measure the robot's success on a temporal discrimination task originally performed by mice to evaluate their capacity to exploit temporal knowledge. We conclude that the robot is able to perceive time similarly to animals when it comes to their intrinsic mechanisms of interpreting time and performing time-aware actions. Next, by analyzing the behavior of agents equipped with the framework, we propose an estimator to infer characteristics of the timing mechanisms intrinsic to the agents. In particular, we show that from their empirical action probability distribution, we are able to estimate parameters used for perceiving time. Overall, our work shows promising results when it comes to drawing conclusions regarding some of the characteristics present in biological timing mechanisms.**

***Index Terms*—Cognitive modeling, microstimuli, reinforcement learning, robotics, time perception.**

I. Introduction

UNDERSTANDING different aspects and characteristics of humans and animals has been a driving force in research for centuries (e.g., Skinner's rats [1], Pavlov's dogs [2], or Harlow's monkeys [3]). Analyzing behavior can bring insight into how bodies and minds function, from motor impulses to neural mechanisms. This can contribute, on the one hand, to obtaining plausible hypotheses about biological mechanisms and understanding variations to the baseline. Such insights might shed light on, e.g., the causes as well as treatments for diseases. On the other hand, replicating these mechanisms in biologically inspired intelligent agents (e.g., robots) can enhance their cognitive skills and interactions with humans [4]. Moreover, biologically inspired mathematical models also have the potential to push the boundaries of artificial intelligence. For example, the perceptron [5], which is a mathematical model inspired by the brain's neurons, has been key in the recent advances in deep learning [6].

One of the properties of biological systems whose understanding has seen some advances but is yet not fully understood is *time perception* [7], [8]. It concerns the mechanisms responsible for the subjective way time is perceived [9]—it is the origin of such idiomatic expressions as "time flies when you have fun." A variable sense of the perception of time is known to be responsible for adaptive behaviors in animals, and has been shown to vary as a function of body size and metabolic rate [10]. Cyber–physical agents such as robots, however, use a linear clock as the basis for its functions instead of perceiving time according to the surrounding context [11]. Using temporal information in the agents' cognitive processes is considered by many researchers to be one of the milestones in achieving artificial general intelligence [12], [13]. The ability to perceive time would enhance skills, such as planning, recalling of experiences, and communication.

The first step to incorporate time perception in a robot's decision-making process is to reproduce the way biological systems acquire and use such temporal information [14]. Many different theories, such as the internal clock theory [15] or the behavioral theory of timing [16], have sought to explain neural timing mechanisms. Furthermore, many computational and robotic models have been created to study different theories [17], such as the pacemaker-accumulator [18] and memory decay [19] models. Review papers such as [20] have brought these works together. However, the current literature still lacks a complete and explainable framework that exploits temporal information to govern the behavior of agents.

In this work, our main goal can be summarized as follows.

Design an explainable end-to-end framework of cognitive mechanisms of time perception whose characteristics can be inferred from the agent's behavior.

To answer this question, we divide our work in three main parts: 1) the design of a framework to model cognitive time perception mechanisms; 2) testing the ability of the framework to exploit temporal information, by comparing the behavior of agents using it to the behavior of animals performing the same task; and 3) estimation of the characteristics

Manuscript received February 5, 2021; revised August 13, 2021; accepted October 2, 2021. Date of publication October 14, 2021; date of current version June 10, 2022. This work was supported in part by NewLEADS; in part by the Swedish Research Council; in part by the Wallenberg AI, Autonomous Systems and Software Program (WASP); and in part by the FCT Project under Grant UID/EEA/50009/2019. *(Corresponding author: Inês Lourenço.)*

Inês Lourenço, Robert Mattila, and Bo Wahlberg are with the Division of Decision and Control Systems, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden (e-mail: ineslo@kth.se; rmattila@kth.se; bo@kth.se).

Rodrigo Ventura is with the Institute for Systems and Robotics, Instituto Superior Técnico, 1049-001 Lisbon, Portugal (e-mail: rodrigo.ventura@isr.tecnico.ulisboa.pt).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCDS.2021.3120301.

Digital Object Identifier 10.1109/TCDS.2021.3120301

of timing mechanisms intrinsic to agents using the framework.

The main contributions provided in each of these parts are, respectively, as follows.

1) A biologically inspired reinforcement learning framework that replicates neuronal mechanisms of time (the behavior of neurons believed to be responsible for time perception) and combines them with time estimates obtained from the environment. This framework capacitates agents with the ability to perform time-aware actions (Section III).
2) Numerical experiments that validate the ability of the proposed framework to exploit temporal information. A robot performing a temporal-discrimination task using the framework demonstrates internal features known to be present in biological timing mechanisms, and estimates the duration of intervals in a similar manner to that of mice on the same task (Section IV).
3) A method to gain biological insight about the framework used by agents for time perception, based on analyzing their behavior (actions performed) to compute the parameters inherent to their timing frameworks (Section V).

The two first contributions were initially studied in the conference paper [21], written by the authors, and are further developed in the current article. The third contribution is completely novel and exploits the previous framework to approach the desired end goal.

The remainder of this article is structured as follows. Section II formulates the time perception problem. Section III describes the biologically inspired decision-making framework proposed to replicate timing mechanisms, and Section IV validates and evaluates the behavior of an agent using it. Section V presents the method to estimate parameters of the timing framework from the behavioral analysis. Finally, in Section VI, the key conclusions that can be drawn from the article are highlighted and discussed. Moreover, some indications for future work are outlined.

II. Problem Formulation

In this section, we start by defining the notation and then present the two main problems addressed in this article.

A. Notation

The subscript t represents discrete time, and the ith element of vector v_t is $v_t(i)$. A general probability mass function is denoted by $p(\cdot)$. Agents that have time perception are denoted as *timing agents*, and tasks where time perception is needed as *timing tasks*. In the case of interval timing tasks, the variable of interest is the time difference between two events, which we designate τ and define as the *elapsed time* or *interval duration*.

B. Problem Formulation

The first step of this work is to formalize a framework for modeling the mechanisms responsible for the perception of temporal information. This framework needs to be able to replicate and reproduce the biological mechanisms responsible for time perception, and do so in a way that allows agents to perform interval timing tasks. We formalize this goal in the following question.

Problem 1 (Biologically Inspired Timing Framework): How can neural time perception mechanisms be reproduced in a framework that exploits temporal information and produces time-aware behaviors?

By solving Problem 1, we establish a decision-making framework that enables agents to perform interval timing tasks based on the replication of neural timing mechanisms, therefore, similar to the way humans and animals perform them. More specifically, we aim to combine an estimate of the elapsed time obtained from sensory information (external timing) with the agent's biologically inspired decision-making process (internal timing). The latter includes features that are inspired by the mechanisms believed to govern time perception, such as the dopaminergic activity [7]. As a result, we propose a framework that includes multiple facets of timing mechanisms for performing time-aware actions.

The initial work on this time-perception framework for solving Problem 1 was presented recently by Lourenço *et al.* [21]. In the current article, we present the framework in Section III and provide validating experiments in Section IV that also illustrate some of its key components.

Once we have a framework for modeling timing mechanisms, we aim to estimate numerical quantities (that have biological analogues) in the framework from observed behavior of timing agents. In other words, we use our model to estimate information about the intrinsic characteristics of the agents' decision-making process. The focus of Section V is, hence, the question as follows.

Problem 2 (Estimating Timing Aspects From Behavior): How can knowledge about the inner mechanisms of timing agents (such as animals) be gathered from their behavior?

In summary, the solution to these two problems results in insight into: 1) how different characteristics of the dopamine system of an agent change its behavior, as well as 2) which ones are more likely to be the characteristics of the timing mechanisms present in the brain.

III. Biologically Inspired Timing Framework

In this section, we review the framework presented in [21] to address Problem 1. We first discuss background material and then formally introduce the framework. Subsequently, we explain each of the two main component of the framework, which are the internal and external timing components. The section is concluded with a summary of how they are combined to obtain the complete timing framework.

A. Preliminaries

We apply results from neuroscience in a decision-making setup to design a biologically inspired reinforcement learning [22] algorithm that replicates neuronal timing mechanisms. We denote these mechanisms as *internal timing* mechanisms, since they are related to how internal biological neuronal mechanisms are believed to affect, and enable, the perception of time. In this field, one of the most popular theories is that

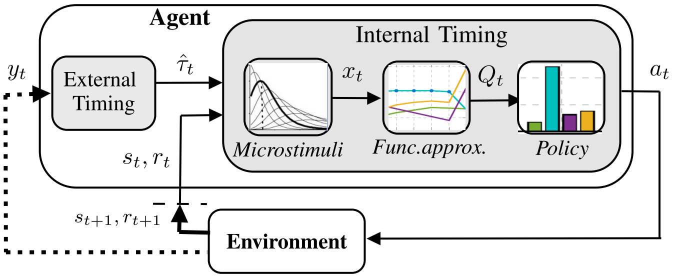

Fig. 1. Twofold time perception framework for replicating time perception mechanisms. Based on the role of *external environmental stimuli* in time perception, the agent receives environmental observations y_t and estimates the elapsed time $\hat{\tau}_t$. A TD learning algorithm replicates *internal timing mechanisms* using this estimate. The algorithm uses microstimuli features (2), which are influenced by the elapsed time estimate $\hat{\tau}_t$ and the state of the environment s_t, to compute the Q-values (1) of each state–action pair. According to its policy (8), it uses the Q-values to select an action a_t to perform, and receives a reward r_t.

it is the spiking activity of neurons, mostly known as firing rate, which encodes the passage of time. This is particularly the case for dopaminergic neurons, evidenced by the change in dopaminergic activity when tasks are carried out at different speeds [7]. To replicate mechanisms of time perception, we, therefore, model the principles of dopaminergic behavior and reproduce them in a robot.

On the other hand, sensory information has been shown to have a direct impact (introduce a bias) on our perception of time [23]. For example, watching a movie in a faster speed than the natural one leads to an overestimation of time intervals [24]. We denote the feature responsible for how the perception of time is influenced by external stimuli as *external timing*.

We thus aim for a framework that reproduces the time estimate that stems from internal neuronal processes (studied in Section III-C), but also that this estimate should be affected by temporal information collected from the environment (explored in Section III-D). By combining temporal information that stems from these two different timing sources, we equip robots with their own structure for time perception and show that context-dependent timing mechanisms can be implemented in nonbiological agents. A reinforcement learning setup is used to evaluate the performance of the biologically inspired decision-making algorithm in timing tasks—the framework is schematically illustrated in Fig. 1, and described in more detail throughout this section.

B. Background

Early models of timing in the brain include the clock accumulator, or pacemaker model [16], and the synchronization of brain wave frequencies [25]. However, later models have been created to more precisely match neural responses. In [26], it was found that time is distributively encoded in the dynamics of neural circuits—the neuronal oscillatory activity or firing rate. It was confirmed in [8] that the perception of interval durations can be explained by the speed of change of certain neuronal populations—particularly, in the dopaminergic system, which can, therefore, be seen as an internal clock of the brain [27].

Due to its intrinsic signal that represents the disparity between the received and the predicted reward, called a reward prediction error, the dopaminergic system has previously been demonstrated to be involved in reward prediction and action selection [28]. To reproduce its ability to predict the importance of future events from patterns of features that encode the agent's experiences, temporal-difference (TD) learning models have been developed. In such models, the reward prediction error signal is referred to as the TD error [29].

To represent these features in our framework, we use one of the currently most plausible theories when it comes to replicating timing mechanisms, which is called *Microstimuli* [30]. Unlike the *Presence* [29] and the *Complete Serial Compound* [29], [31] theories, the microstimuli utilizes a small set of elements per stimulus to accurately replicate timing results [32].

C. Internal Timing

Formally, the TD learning model alluded to in the previous section is modeled in a discrete-time reinforcement learning setting. In such, the interaction between an agent and the environment takes the form of a Markov decision process [33], [34]. Formally, the environment is described by a state $s_t \in \mathcal{S}$ that evolves over time, where $\mathcal{S}$ is the state space and t represents discrete time. At each time step, the agent performs an action, $a_t \in \mathcal{A}$, and receives a reward $r_t \in \mathbb{R}$ based on the action performed and the state of the environment. Its goal is to find the policy (i.e., a strategy) that maximizes the expected sum of future rewards. The rest of this section explains the components of the internal timing component shown in Fig. 1.

We use Q-values [22] in our reinforcement learning setup since we are interested in studying the actions performed by the agent. To generalize the estimate of the value of a state to states that have similar features, we use *function approximation* [22], which has been shown very advantageous in problems with large state–action spaces. In particular, we use a linear weighted combination of D features $x(s, a) \in \mathbb{R}^D$ to compute the Q-values of state–action pairs, $Q(s, a) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, as

$$Q_t(s, a) = w_t^T x_t(s, a) = \sum_{j=1}^{D} w_t(j) x_t(j). \tag{1}$$

These features $x_t(1), \ldots, x_t(D)$ are chosen so as to replicate the internal neuronal mechanisms mentioned in Section III-B, which are based on the behavior of dopaminergic neurons and represented by the *Microstimuli* framework.

In this framework, each cue and reward deploy a set of m microstimuli. This means that a total of $m\zeta = D$ microstimuli are deployed in episodes with ζ cues and rewards. At time t, the level of each microstimulus is represented by a feature x_t (see [30, Fig. 2] for a graphical illustration). Mathematically, this relation is modeled as

$$x_t(j) = h_t f\left(h_t, \frac{j}{m}, \beta\right), \quad \text{for } j = 1, \ldots, D. \tag{2}$$

In this relation, the level x_t of each microstimulus is computed as the product between an exponentially decaying trace height $h_t \in \mathbb{R}$ with decay parameter ξ

$$h_t = \exp\{-(1-\xi)t\} \tag{3}$$

and a Gaussian basis function

$$f(h, \nu, \beta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(h-\nu)^2}{2\beta^2}\right\} \tag{4}$$

with center ν and width β. The amount of decay of each microstimulus can be used as a reference for the elapsed interval due to the steadily decaying memory trace. As can be seen in (1), each microstimuli feature from (2) is multiplied by a weight $w_t \in \mathbb{R}^D$. These weights, $w_t(1), \ldots, w_t(D)$, are the values that the agent wants to learn since they reflect the importance of each feature (such as the strengths of the corticostriatal synapses) for the different state–action pairs [30].

An essential attribute to have in efficient reward learning is eligibility traces, e_t [35]. Eligibility traces are a characteristic of learning that acts as a vector of memory parameters that are susceptible to changes according to the events that they are associated to.

In summary, the standard reinforcement learning update equations of the TD error δ_t, the weights w_t, and the eligibility traces e_t are, respectively

$$\delta_t = r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \tag{5}$$

$$w_{t+1}(j) = w_t(j) + \alpha \delta_t e_t(j) \tag{6}$$

$$e_{t+1}(j) = \gamma \eta e_t(j) + x_t(j) \tag{7}$$

where α is the learning rate, γ is the discount rate, and η is a decay parameter that determines the amount of influence of recent stimuli.

Once the Q-values are computed, according to (1), the agent is left with the end-goal task of choosing which action to perform. We apply a standard action selection mechanism, called ε-greedy [22]. The policy of the agent is to, at each time step, either select a random action or the one with the highest Q-value

$$a_t = \begin{cases} \arg\max_a Q(s_t, a), & \text{with probability } 1-\varepsilon_t \\ \text{random action}, & \text{with probability } \varepsilon_t. \end{cases} \tag{8}$$

This corresponds to the exploration–exploitation tradeoff that is the basis for learning, where ε_t is an exploration parameter with decay parameter $\rho \in [0, 1]$ that decreases according to $\varepsilon_t = \rho\varepsilon_{t-1}$ as the agent learns.

This concludes our decision-making framework for replicating internal neuronal mechanisms.

D. External Timing

We now briefly outline how an estimate of the elapsed time can be computed based on environmental stimuli (full details are available in the Appendix). The complete framework we propose then consists in merging the two estimates (internal and external) to form a biologically inspired perception of the elapsed time.

It has been shown that a sensory estimate of the passage of time can be obtained from environmental observations $\mathcal{O}$ [24], [36]. We use a probabilistic expectation of stimulus change in the environment to compute an estimate of the elapsed time as in [23], where a Bayesian framework was proposed to estimate the elapsed time τ from the environmental data $\mathcal{O}$.

Essentially, we model the joint distribution of the sensory data as Gaussian processes with an Ornstein–Uhlenbeck (OU) kernel function. We compute the hyperparameters of the model using Bayesian model selection, and from the maximum-likelihood principle obtain the estimate $\hat{\tau}$.

A more detailed explanation about how the elapsed time estimate can be computed was presented in [21], and a summary can, for completeness, be found in the Appendix.

E. Summary of the Complete Timing Framework

An agent is navigating the environment and receiving stimuli. Using the external timing mechanism, it is able to estimate the perceived elapsed interval length $\hat{\tau}$ between each stimuli. Based on the internal timing mechanisms, sets of m decaying microstimuli features, x_t from (2), are deployed when the agent estimates having received the stimuli (cues and rewards). The perceived interval $\hat{\tau}$ then conditions the microstimuli features x_t that replicate the dopaminergic activity believed to regulate timing mechanisms, which influence the Q-values of the state–action pairs according to (1) and, as can be seen in (8), also the action a_t selected by the agent. In summary, this results in a decision-making framework that incorporates internal and external timing mechanisms.

We assess the complete timing framework in the next section, by analyzing the behavior of the agent when performing a temporal discrimination task.

IV. Simulating the Time Perception Framework

As in [21], we design a robot that following the framework presented in Section III, performs sequences of actions in a temporal discrimination task. We start by presenting the experimental setup used, then detail on how the framework is simulated, and finally, the numerical results that highlight the success of the robot in the task. The success is measured by the similarity of its actions to the actions performed by a biological timing agent (an agent with temporal cognition, which, in this case, is a mouse) in the same task.

A. Experimental Setup

We evaluate the framework proposed in Section III in a temporal discrimination setup, where the ability of a robot to distinguish time intervals (durations) is evaluated and compared to that of mice performing the same task [7].

In the original experiment [7], three buttons are available to a mouse: "Start," "Short," and "Long." The experiment starts when the animal presses the former, and two auditory tones, separated by a certain interval that varies from episode to episode, are played to the mouse. It then has the possibility to press the button that corresponds to its estimated interval length between both tones (Short or Long). A reward of water or food is given to the animal if the correct button is pressed (i.e., if the animal correctly estimated the elapsed time between the two tones).

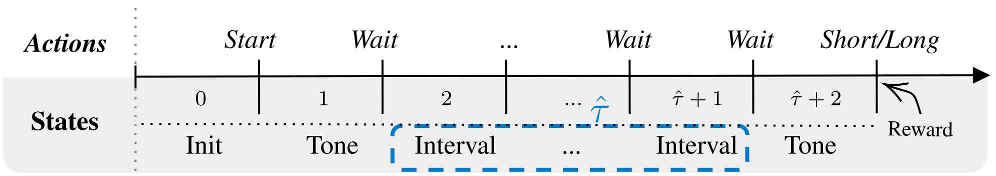

Fig. 2. Optimal sequence of actions (top row) performed during an episode and corresponding state transition (bottom row). Once the *Start* button is pressed, the environment changes to the *Tone* state. From there, the agent must perform the "Wait" action until reaching the second "Tone" state, which happens during a number of *Interval* states sampled is uniformly from the maximum interval length. After the second *Tone* state, the agent then performs the *Short* or *Long* action that corresponds to its estimated number of *Interval* time steps, $\hat{\tau}$. If the correct action is chosen, the agent receives a positive reward.

We use the same setup for our simulated robot: at each time step, the environment can be in one of the states $\mathcal{S} = \{\text{Init, Tone, Interval}\}$, and the robot can perform one of the actions $\mathcal{A} = \{\text{Start, Wait, Short, Long}\}$. Fig. 2 schematically shows the optimal sequence of actions in an episode. The number of "Interval" states is what defines the *interval duration* τ of the episode. If we define $L \in \mathbb{N}$ as the experiment's maximum interval duration between tones, then, in each episode, τ is a realization of a discrete uniform random variable $\tau \sim \text{unif}\{1, L\}$. The temporal discrimination task consists of classifying the interval duration as

$$\begin{cases} \text{"Short"}, & \text{if } \tau \in \left[1, 2, \ldots, \lfloor \frac{L}{2} \rfloor\right] \\ \text{"Long"}, & \text{if } \tau \in \left[\lceil \frac{L}{2} + 1 \rceil, \ldots, L\right] \end{cases} \tag{9}$$

where $\{\lfloor (L/2) \rfloor, \lceil (L/2) + 1 \rceil\}$ is the classification boundary for L and $\lceil (L/2) \rceil$ for L odd. In the numerical experiments below, we selected the maximum interval duration L as in the real experiment [7], where 3 s correspond to $L = 8$ time steps. This results in Short $:= \tau \in \{1, 2, 3, 4\}$ and Long $:= \tau \in \{5, 6, 7, 8\}$ time steps.

The increased complexity of the problem comes from the fact that before learning to distinguish between short and long intervals, the robot first has to compute its estimate of the elapsed time between the two stimuli $\hat{\tau}$ using the external timing mechanism from Section III-D. Only then can it use the estimate to learn the classification task. The robot computes its estimate of the elapsed time between stimuli by navigating around the environment and gathering data from its sensors during the interval τ between which the two stimuli are presented to it.

B. Simulating the Proposed Framework

Following this setup, in the external timing mechanism, we chose the values of i LIDAR angles of the simulated robot (emulating the visual system of a mouse) to be the environmental data $y_t(i)$ at time t. We estimate the elapsed time $\hat{\tau}$ using the maximum-likelihood parameters computed for the model of sensory data (see the Appendix). As a result of its internal timing mechanisms, sets of $m = 8$ microstimuli features x_t are deployed when the agent receives each of the two stimuli, separated by $\hat{\tau}$ time steps. Algorithm 1 summarizes the complete framework from Fig. 1.

In the next sections, we discuss the key aspects of our approach to Problem 1. The results specifically for the external timing estimation can be found in [21].

Algorithm 1 Temporal Discrimination Task

```
Initialize $Q(s, a) = 0$, for all $s \in \mathcal{S}$, $a \in \mathcal{A}$, and $w(1), \ldots, w(D)$ randomly (e.g., $w(j) \in [0, 1]$)
for each episode do
    Initialize the state to "Init"
    for each time step $t$ do
        Update $x_t(1), \ldots, x_t(D)$ according to (2)
        if only one "Tone" state has passed then
            Collect data $y_t(1), \ldots, y_t(M)$
        else if both "Tone" states have passed then
            Estimate the elapsed time, $\hat{\tau}$, by maximizing (15)
            Update $x_{\hat{\tau}}(1), \ldots, x_{\hat{\tau}}(D)$, according to (2)
        end if
        Compute the Q-values according to (1) and choose $a_t$ according to (8). Take action $a_t$, observe $r_t$, $s_{t+1}$
        $\delta_t \leftarrow r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)$
        $w_{t+1}(j) \leftarrow w_t(j) + \alpha \delta_t e_t(j)$,    for $j = 1, \ldots, D$
        $e_{t+1}(j) \leftarrow \gamma \eta e_t(j) + x_t(j)$,    for $j = 1, \ldots, D$
        $s_t \leftarrow s_{t+1}$
    end for
    Until $s_t$ is terminal
end for
```

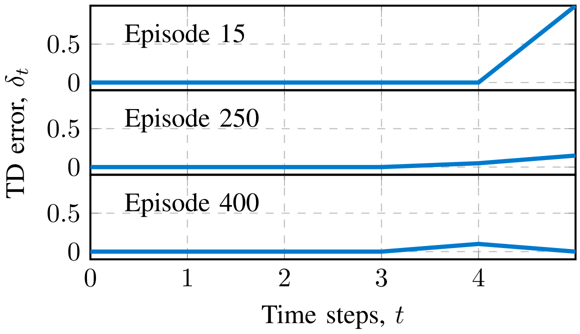

Fig. 3. Evolution of the TD error throughout three episodes with $\tau = 2$ time steps in different phases of learning. Unsurprisingly, the more the agent has learned, the more it expects a reward after correctly classifying the interval ($t = 4$), leading to a decreasing TD error at the end of the episode and an increasing one upon receiving the second tone.

C. Results on the Inner Core Mechanisms of the Agent

We begin by presenting results that bring insight into the robot's inner mechanisms when using our framework, such as the TD error and the Q-values. The time step values on the x-axis correspond to the state numbers from Fig. 2.

Fig. 3 shows the evolution of the TD error throughout three successful episodes with the same interval duration, but each

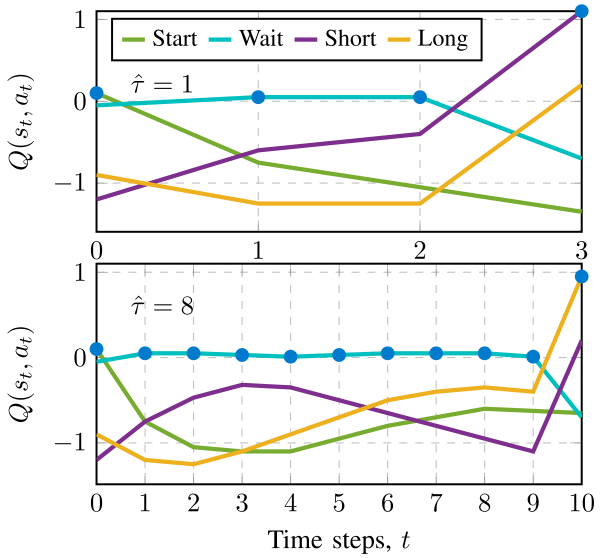

Fig. 4. Evolution of the Q-values throughout two episodes with different interval durations between tones. In the top figure, $\hat{\tau} = 1$ (short interval), and in the bottom one, $\hat{\tau} = 8$ time steps (long interval). The chosen action at each time step is the one with the highest Q-value at that moment. At the end of the episode, different actions have the highest Q-value, in agreement with the corresponding duration.

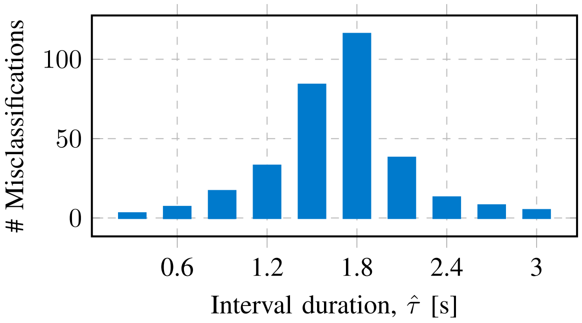

Fig. 5. Number of misclassified episodes according to the interval duration. The total number of misclassifications is 327, the average is 1.65 s, and the median is 1.8 s. As for humans and animals, the intervals in the boundary between classes are the ones most commonly misclassified.

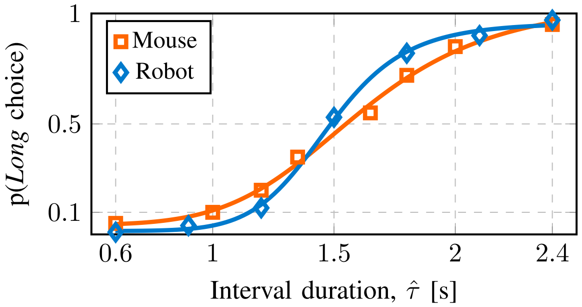

Fig. 6. Psychometric curves corresponding to the empirical probability of intervals being classified as *Long*. The psychometric curve of the agent (in blue) using our framework closely matches that of the mouse (in orange). The latter was averaged over ten experiments and fitted using a logistic function.

at a different training phase. The represented episodes have $\tau = 2$ time steps, which means that the second tone takes place at $t = 4$ where it is followed by the reward. The figure shows that the TD error from (5): 1) decreases over time as the agent learns the optimal policy when the reward is delivered and 2) increases when the second tone is played. This means that previous (conditional) stimuli teach the agent to predict reward delivery. In other words, the second tone functions as a conditional stimulus from *classical conditioning* [28].

Fig. 4 shows a visual representation of the evolution of the Q-values computed from (1) during one episode with $\hat{\tau} = 1$ (short interval), and one with $\hat{\tau} = 8$ time steps (long interval), after the agent has learned the optimal policy. At each time step, the action with the highest Q-value is marked with a circle and is the one selected without exploration. The resulting sequence of actions is the optimal one shown in Fig. 2 for both episodes, which means that based on its time estimate, the agent learns to act appropriately.

D. Results on the Replication of the Behavior

The following results illustrate the behavior of the agent in the task. To simplify the comparison with the performance of mice in the original experiment, we use the seconds elapsed between tones in the x-axis to present the results.

The agent learns to always act correctly until receiving the second tone. However, even once the training phase is over, it is not always able to accurately classify the interval length of the episode since it does not always select the correct action at that moment. Fig. 5 shows the number of episodes in which it misclassified the interval duration. This happens more often for intervals closer to the boundary between Short and Long (around $\hat{\tau} = \{1.5, 1.8\}$ s). These results show a trend also exhibited by humans and animals [8], which is verified for any maximum interval length.

Fig. 6 shows the empirical probability (psychometric curve) of different agents classifying different intervals as Long. The average performance of a mouse (in orange, from [7]) is qualitatively similar to that of an agent using our time perception framework (in blue).

E. Summary of the Results

The above numerical experiments point to the following conclusion.

Result 1: Our robot demonstrates the ability to classify interval durations similar to that of mice, which validates the capability of our framework to use temporal information for performing time-aware actions.

More specifically, Result 1 is supported by the similarity between the timing mechanisms of the robot and mice.

1) The elapsed time is successfully computed from environmental data, and uncertainty in the estimation of the interval duration increases with the length of the interval (scalar property [37], [38]).
2) The behavior of the TD error of our model replicates the reward prediction error of the neural dopaminergic system (it decreases with reward expectancy).
3) The uncertainty in the classification of intervals is higher on the boundary between classes, as for humans and animals.

4) The psychometric curves of our agent closely match the one of mice.

V. Estimating Timing Aspects From Behavior

In the previous sections, we: 1) proposed a framework for replicating mechanisms of time perception and 2) showed that it provides results consistent with biological principles, answering Problem 1. Our next goal is to exploit this framework to gain more insight into characteristics of the brains of humans and animals, answering Problem 2.

We start by presenting characteristics that might be of interest to estimate and other works that studied them. We then detail our proposed method and conclude with numerical experiments.

A. Preliminaries

In the previous sections of the article, we have demonstrated a good correspondence between real-world experiments on biological agents (i.e., mice) and that of artificial simulated agents. In other words, our mathematical framework can arguably model such behavior well. Since there are a number of free parameters in the reinforcement learning model, we could use the framework "in reverse" to deduce important biological characteristics of the mice.

An example of a parameter of interest to estimate in the framework presented in Section III is the number m of microstimuli. The number of microstimuli present in each agent conditions its ability to distinguish time intervals. This is the case since with too few microstimuli, the agent has a too scarce perception of time and is not able to distinguish different time intervals well. This is discussed further as follows.

Apart from the number of microstimuli, other parameters of the TD learning framework dictate the agent's behavior and can be desired to estimate—e.g., the learning rate α, the discount rate γ, exploration rate ε, and the exploration decay ρ. Other parameters of the microstimuli framework can also be estimated, such as the microstimuli decay ξ, and the basis functions center ν and width β. To simplify, we focus on the estimation of one or more of these parameters, under the assumption that they are static and the others are known.

Many works have studied the question of how to infer biological parameters from the agent's behavior [39]–[41], including in reinforcement learning models, but not in a way that allows temporal information to be taken into account.

B. Proposed Method for Parameter Estimation

To answer Problem 2, we reformulate the question posed to: given a certain behavior, find the parameter values (broadly denoted θ here) that best explain it.

Here, the behavior corresponds to the policy followed by the agent—that is, the actions taken at different states. Since different parameters induce different behaviors, the most straight-forward approach would consist of analyzing the correspondence between the state–action probabilities induced by the parameters and the agent's behavior. This approach is, however, not possible for time-dependent problems since the states do not encode temporal information. The approach we propose consists of adding to this formulation the variable τ that encodes time when analyzing the behavior of the agent through the actions a performed. Our model is

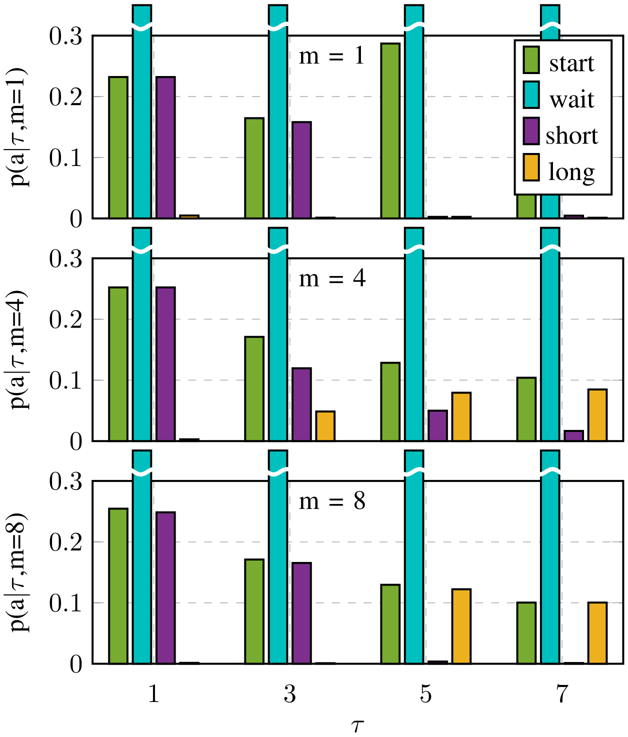

Fig. 7. Effect of the number of microstimuli m on the behavior a of the agent for different time intervals between two stimuli, τ. These statistics are the base of our model (10) and were computed over one training simulation with 2000 episodes.

$$p(a|\tau, \theta) \tag{10}$$

where the actions a are modeled as a stochastic variable described by this probability function. It represents the distribution of actions performed with parameters θ (e.g., $\theta = [m, \alpha, \varepsilon, \dots]$) on an episode with time interval τ.

This model can be obtained from the empirical probability distribution using simulation data, and is shown in Fig. 7 for θ being the number of microstimuli $\theta = m$. The figure shows how the ratio between certain actions changes according to the parameter (namely, $m = 1, 4,$ and 8 microstimuli). For example, the proportion of short and long actions is only correct for $m = 8$ microstimuli. This change can also be observed in the other variables mentioned in Section V-A, although for some more noticeably than others.

Observing the behavior of agents in this context means then collecting information about the actions they perform at different time intervals. Thus, we collect a *history* of actions

$$\begin{aligned}\mathcal{H} = \big[&\big(\tau_1, a_{1,1}\big), \dots, \big(\tau_1, a_{1,N_1}\big)\\ &\big(\tau_2, a_{2,1}\big), \dots, \big(\tau_2, a_{2,N_2}\big), \dots\big]\end{aligned} \tag{11}$$

where different episodes k correspond to different time intervals τ_k. At episode k, N_k actions are performed, $a_{k,1}, \dots, a_{k,N_k}$.

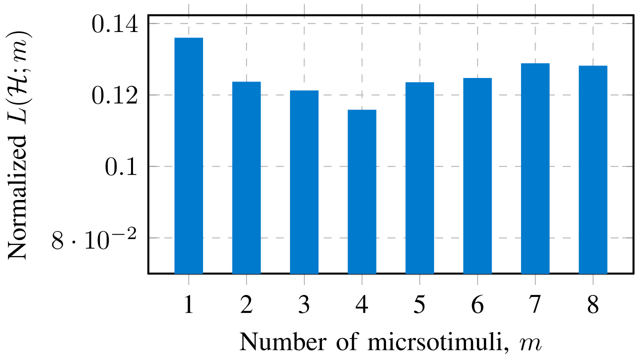

Fig. 8. Normalized likelihood. The likelihood of each number of microstimuli increases as it approaches the true value. The maximum-likelihood estimate $\hat{m}$ (14) coincides with the true value, $m = 4$.

We write the negative log likelihood of history $\mathcal{H}$ given parameters θ as

$$L(\mathcal{H}; \theta) = -\sum_k \sum_{n=1}^{N_k} \log p\big(a_{k,n} \mid \tau_k, \theta\big) \tag{12}$$

since

$$\begin{aligned} p(\mathcal{H} \mid \theta) &\propto p\big(\{a_{k,n}\} \mid \{\tau_k\}, m\big) \\ &= \prod_k \prod_{n=1}^{N_k} p\big(a_{k,n} \mid \tau_k, \theta\big) \end{aligned} \tag{13}$$

assuming conditional independence among individual actions given the parameters and the corresponding episode duration. The maximum-likelihood estimator is then

$$\hat{\theta} = \arg\min_{\theta} L(\mathcal{H}; \theta). \tag{14}$$

This method has a limitation for imbalanced data sets, where, for distinct episodes k, the number of actions performed N_k is very different.

Nevertheless, per the above we can estimate different parameters in the framework from empirically observing action distributions and computing the maximum-likelihood estimate.

C. Numerical Experiments

We use the method described in Section V-B to estimate the number of microstimuli ($\theta = m$) of an agent performing the same task from Section IV-A.

Fig. 8 shows the normalized average likelihood (12) computed over ten Monte Carlo simulations of 2000 episodes for the different numbers of microstimuli, when the true value is $m = 4$. It can be seen that the likelihood of each number of microstimuli increases as it approaches the true value, and that the maximum-likelihood estimate is correct.

Since we are using a discrete model, we define *accuracy* as the percentage of times that the maximum-likelihood estimate is the correct solution ($\hat{\theta} = \theta$). We compute the likelihood over a number of microstimuli m in the set $m = \{1, \dots, 8\}$ in all results presented. We averaged the training of the model in (10) over 10 train simulations of $k = 2000$ episodes. When testing it in 30 test simulations, a correct maximum-likelihood estimate (14) was obtained 100% of the time ($\hat{m} = m$).

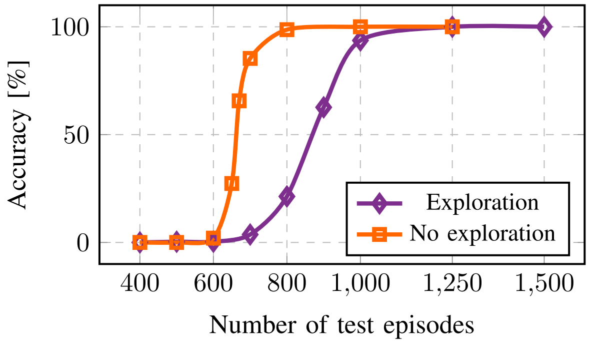

Fig. 9. Evolution of the percentage of successes (accuracy) of the estimator as the number of test episodes increases. After around 1000 test episodes, the original estimator (that admits exploratory actions, in purple) has a 100% accuracy while the estimator that ignores exploratory actions (in orange) has the same accuracy after only 800 episodes.

The above numerical experiments point to the following.

Result 2: Using the maximum-likelihood estimator, one can correctly use the behavior of the agent to infer parameters intrinsic to the agent's internal timing mechanisms.

Let us now analyze some properties of the proposed method.

1) Number of Testing Episodes: Fig. 9 shows the importance of the number of test episodes for the accuracy of the estimator. More specifically, the purple line shows how the accuracy (the percentage of times that $\hat{m} = m$) increases (averaged over 300 test simulations to improve the precision) as the number of test episodes increases. This result is related to the level of exploration of the agent studied next, since the reason for the low accuracy in low numbers of test episodes is the exploratory (i.e., random) behavior of the agent before learning the optimal policy.

2) Exploration: Since our model uses the statistics of all actions performed throughout all episodes, its accuracy is affected by the agent following the optimal policy or performing exploratory actions, as previously shown. If the exploration parameters ε and ρ of the agent are known (or estimated using the method in Section V-B with $\theta = \varepsilon$, or directly from the agent's behavior), we can train our model only with knowledge-exploiting behaviors. This corresponds to only observing the actions performed for small values of ε (e.g., $\varepsilon < 0.01$).

The orange line in Fig. 9 shows that, as expected, observing only knowledge-exploiting actions is advantageous for the accuracy of the estimator, increasing for a comparative number of testing episodes.

3) Sensitivity: The parameters inherent to the agents (that we assume static and known) can be estimated but are typically not known with certainty. To explore the sensitivity of our estimator to the uncertainty in the parameters, we perturb them when estimating the number of microstimuli.

Let us choose α and γ as our parameters with uncertainty. We perturb the true values with noise when training the model (over five train simulations), and compute the average accuracy between the noisy values (e.g., $\alpha \pm$ noise) over five test simulations. Fig. 10 shows the average accuracy for percentages of noise between 0% and 80% in the parameters. The large discrete steps seen on the accuracy are due to the average

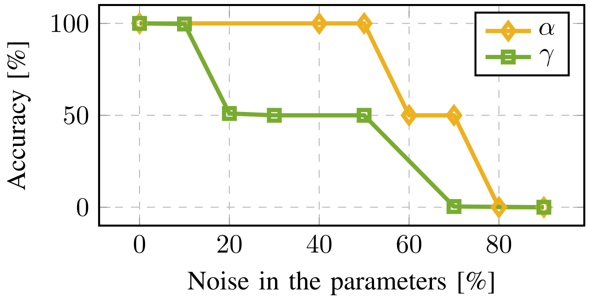

Fig. 10. Evolution of the accuracy of the estimator as the uncertainty (noise) in the parameters α and γ increases.

TABLE I
VARIATION OF THE ACCURACY WITH THE UNOBSERVED PARAMETERS

Unobserved parameters		Accuracy
a	τ	
-	-	100 %
-	2 to 8	78 %
Long	2 to 8	40 %
Long	-	100 %
Wait, Short, Long	-	2 %
Wait, Short	4	0 %

being made between two very different values of the parameter (± noise). To obtain a smoother curve, the noise could instead be sampled from a uniform or Gaussian distribution.

We conclude that the method is robust to small perturbations in both of the parameters studied, and a similar analysis can be done for other parameters of interest.

4) Scalability: Analyzing the behavior of the agent for many different parameter values and different time intervals can be computationally heavy and impractical to implement in real-life experiments. This problem is aggravated for large action spaces. In this section, we evaluate the scalability of the estimator by exemplifying its behavior for situations where only specific actions or certain time intervals between tones are observed. For example, it can either be: 1) a whole action (e.g., the wait action is not counted); or 2) a whole time interval (e.g., intervals of $\tau = 3$ s are not tested); or even 3) random combinations of both, which are not observed. Whichever the situation, some intervals and actions have a higher impact for the accuracy of the estimation than others. For example, for this particular task, the short and long actions present more information about the knowledge of the agent than others. As can be seen in Fig. 7, their probability distributions change significantly according to the number of microstimuli.

Table I shows the accuracy of the estimator computed when different sets of parameters are removed over 50 test simulations. For example, the second row corresponds to not observing the actions of the agent in time intervals between 2 and 8 time steps. An accurate estimate of the number of microstimuli is still obtained 78% of the time, showing that the method is flexible with the testing situations.

VI. CONCLUSION

This article investigated characteristics of neural mechanisms involved in time perception, taking into consideration multiple aspects of these timing mechanisms to design an end-to-end decision-making framework, whose parameters can be inferred from the behavior of an agent.

A. Summary

The first main focus of this work was to create a time perception framework capable of producing time-aware actions (Section III). This framework consists of a combination of two known timing sources: 1) internal neuronal mechanisms and 2) external stimuli. For the former, we replicate dopaminergic behavior by means of a TD learning algorithm with a feature representation called *Microstimuli*. For the later, we estimated of the passage of time from environmental data by exploiting results from Gaussian processes.

We validated the framework in a simulated robot, in Section IV, and compared its behavior to that of real-world mice. The ability of an agent using the proposed algorithm to perceive the passage of time and succeed in time-dependent tasks was validated in numerical results by the comparison with the behavior of real animals (see Result 1). We presented coherent results in both sources of time perception: both in its instrinsic mechanisms of interpreting time, as well as in its performed actions. We concluded the former by identifying features known to be present in humans and animals, and the latter by comparing its actions with those of mice in the same task.

The second main focus of this work was a method presented in Section V to infer biological insight about the framework used by agents for time perception. We used a maximum-likelihood estimator to deduce biological characteristics of the timing functions of agents by estimating parameters they use for perceiving time. In particular, we showed that we are able to estimate the number of microstimuli of an agent given its empirical action probability distribution, using the maximum-likelihood estimator (see Result 2). We further showed: 1) how the estimate improves with the number of testing episodes; 2) how knowing the exploration rate of the agent improves the estimation process; 3) that uncertainty in the other parameters does not affect greatly the estimate; and 4) that the method is scalable and works even with reduced amounts of testing observations.

B. Future Work

There are several interesting extensions that we would like to investigate in future work. First, implementing the framework on a real robot and having it perform the same real-world task as mice would give valuable experimental data that can be used to validate and modify our proposed framework. Since the mechanisms controlling temporal judgments are believed to vary across time scales, we would like to study the behavior of our framework outside the time scale of seconds and study how it can be adapted.

Second, it would be highly interesting to estimate biological characteristics of various animals (and humans) by using the algorithms discussed in Section V. By observing their actions, we could estimate their intrinsic timing parameters, and validate them by comparing them with the actions of simulated agent with those parameters. In that case, we should also consider the estimation of dynamic parameters. Another relevant question concerns how knowledge about the animal's inherent parameters can be obtained by observing its behavior in different tasks.

Finally, it would also be interesting to investigate how more sophisticated methods for inverse learning problems can be combined with the timing framework discussed in this article. For example, methods from inverse reinforcement learning and revealed preferences (e.g., [42] and [43]) or inverse filtering (e.g., [44] and [45]) could be used to infer the internal beliefs of the agents based on their behavior.

Appendix
External Timing

Ahrens and Sahani [23] proposed a method to estimate the elapsed time τ from environmental data $\mathcal{O}$ using a Bayesian framework. Under uniform prior, the maximum-likelihood estimate of the elapsed time is given by the maximum of $p(\mathcal{O}|\tau)$ [46] and corresponds to the probability of observing $\mathcal{O} = \{y_t(i)\}_{i=1}^M$ during the interval τ. This probability can be modeled as a zero-mean joint Gaussian distribution over the N observations of all M independent sensors [47]. With $t_1 = 0$ and $t_N = \tau$, this is given by

$$p(y_t(i)|\tau) = \mathcal{N}(y_t(i); 0, K_\Omega) = \frac{e^{-\frac{1}{2}y_t(i)K_\Omega^{-1}y_t^T(i)}}{\sqrt{\det(2\pi K_\Omega)}}. \qquad (15)$$

This joint distribution includes a kernel function K_Ω that is parametrized by Ω and expresses the variation of the process between time steps. It has been observed that the power spectrum of the observations (the rhythm of change of the natural environment) resembles that of the OU function [48]

$$K_{\lambda,\sigma}(\tau) = e^{-\lambda|\tau|} + \sigma^2\psi(\tau). \qquad (16)$$

Here, $\psi(0) = 1$ and $\psi(\tau) = 0$ for $\tau \neq 0$, and $\Omega = [\lambda, \sigma]$ are the hyperparameters of the model. We estimate them using Bayesian model selection, by maximizing the logarithm of the likelihood with respect to Ω [46]. This involves computing the respective derivatives

$$\frac{\partial}{\partial \Omega_j}\log p(y_t(i)|\Omega) = -\frac{1}{2}\mathrm{tr}\Big(\phi\phi^T - K_\Omega^{-1}\Big)\frac{\partial K_\Omega}{\partial \Omega_j} \qquad (17)$$

where $\phi = K^{-1}y_t(i)$.

Hence, the external timing problem is solved by identifying the values for λ and σ that make the properties of the Gaussian process most similarly approximate the ones of the environmental data. The robot's estimate of the elapsed time between the two stimuli $\hat{\tau}$ is thus obtained by computing the maximum-likelihood estimate of (15).

References

[1] B. F. Skinner, *The Behavior of Organisms: An Experimental Analysis*. New York, NY, USA: Appleton-Century," 1938.

[2] I. P. Pavlov, *The Work of the Digestive Glands*. London, U.K.: C. Griffin, 1910.

[3] H. F. Harlow, R. O. Dodsworth, and M. K. Harlow, "Total social isolation in monkeys," *Proc. Nat. Acad. Sci. United States Amer.*, vol. 54, no. 1, pp. 90–97, 1965.

[4] B. Webb, "What does robotics offer animal behaviour?" *Animal Behav.*, vol. 60, no. 5, pp. 545–558, 2000.

[5] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.

[6] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.

[7] S. Soares, B. V. Atallah, and J. J. Paton, "Midbrain dopamine neurons control judgment of time," *Science*, vol. 354, no. 6317, pp. 1273–1277, 2016.

[8] T. S. Gouvêa, T. Monteiro, A. Motiwala, S. Soares, C. Machens, and J. J. Paton, "Striatal dynamics explain duration judgments," *Elife*, vol. 4, Dec. 2015, Art. no. e11386.

[9] L. G. Allan, "The perception of time," *Percept. Psychophys.*, vol. 26, no. 5, pp. 340–354, 1979.

[10] K. Healy, L. McNally, G. D. Ruxton, N. Cooper, and A. L. Jackson, "Metabolic rate and body size are linked with perception of temporal information," *Animal Behav.*, vol. 86, no. 4, pp. 685–696, 2013.

[11] M. Maniadakis, P. Trahanias, and J. Tani, "Explorations on artificial time perception," *Neural Netw.*, vol. 22, nos. 5–6, pp. 509–517, 2009.

[12] M. Maniadakis and P. Trahanias, "Time in consciousness, memory and human-robot interaction," in *Proc. Int. Conf. Simul. Adapt. Behav.*, 2014, pp. 11–20.

[13] M. Maniadakis and P. Trahanias, "Temporal cognition: A key ingredient of intelligent systems," *Front. Neurorobot.*, vol. 5, p. 2, Sep. 2011.

[14] B. Deverett, R. Faulkner, M. Fortunato, G. Wayne, and J. Z. Leibo, "Interval timing in deep reinforcement learning agents," in *Proc. Advances Neural Inf. Process. Syst. (NeurIPS)*, Vancouver, BC, Canada, 2019, pp. 6689–6698.

[15] R. M. Church, "Properties of the internal clock," *Ann. New York Acad. Sci.*, vol. 423, no. 1, pp. 566–582, 1984.

[16] P. R. Killeen and J. G. Fetterman, "A behavioral theory of timing," *Psychol. Rev.*, vol. 95, no. 2, pp. 274–295, 1988.

[17] C. Addyman, R. M. French, and E. Thomas, "Computational models of interval timing," *Current Opinion Behav. Sci.*, vol. 8, pp. 140–146, Apr. 2016.

[18] P. Simen, F. Rivest, E. A. Ludvig, F. Balci, and P. Killeen, "Timescale invariance in the pacemaker-accumulator family of timing models," *Timing Time Percept.*, vol. 1, no. 2, pp. 159–188, 2013.

[19] C. Addyman, R. French, D. Mareschal, and E. Thomas, "Learning to perceive time: A connectionist, memory-decay model of the development of interval timing in infants," in *Proc. Annu. Meeting Cogn. Sci. Soc.*, vol. 33, 2011, pp. 354–359.

[20] H. Basgol, I. Ayhan, and E. Ugur, "Time perception: A review on psychological, computational and robotic models," 2020. [Online]. Available: arXiv:2007.11845.

[21] I. Lourenço, R. Ventura, and B. Wahlberg, "Teaching robots to perceive time: A twofold learning approach," in *Proc. Joint IEEE 10th Int. Conf. Develop. Learn. Epigenet. Robot. (ICDL-EpiRob)*, Valparaiso, Chile, 2020, pp. 1–7.

[22] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[23] M. B. Ahrens and M. Sahani, "Observers exploit stochastic models of sensory change to help judge the passage of time," *Current Biol.*, vol. 21, no. 3, pp. 200–206, 2011.

[24] S. W. Brown, "Time, change, and motion: The effects of stimulus movement on temporal perception," *Percept. Psychophys.*, vol. 57, pp. 105–116, Jan. 1995.

[25] M. S. Matell and W. H. Meck, "Cortico-striatal circuits and interval timing: Coincidence detection of oscillatory processes," *Cogn. Brain Res.*, vol. 21, no. 2, pp. 139–170, 2004.

[26] J. Wang, D. Narain, E. A. Hosseini, and M. Jazayeri, "Flexible timing by temporal scaling of cortical responses," *Nat. Neurosci.*, vol. 21, no. 1, pp. 102–110, 2018.

[27] S. J. Gershman, A. A. Moustafa, and E. A. Ludvig, "Time representation in reinforcement learning models of the basal ganglia," *Front. Comput. Neurosci.*, vol. 7, p. 194, Jan. 2014.

[28] P. W. Glimcher, "Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis," *Proc. Nat. Acad. Sci.*, vol. 108, no. S3, pp. 15647–15654, 2011.

[29] R. S. Sutton and A. G. Barto, "Time-derivative models of Pavlovian reinforcement," in *Learning and Computational Neuroscience: Foundations of Adaptive Networks*. Cambridge, MA, USA: MIT Press, 1990, pp. 497–537.

[30] E. A. Ludvig, R. S. Sutton, and E. J. Kehoe, "Stimulus representation and the timing of reward-prediction errors in models of the dopamine system," *Neural Comput.*, vol. 20, no. 12, pp. 3034–3054, 2008.

[31] P. R. Montague, P. Dayan, and T. J. Sejnowski, "A framework for mesencephalic dopamine systems based on predictive Hebbian learning," *J. Neurosci.*, vol. 16, no. 5, pp. 1936–1947, 1996.

[32] E. A. Ludvig, R. S. Sutton, and E. J. Kehoe, "Evaluating the TD model of classical conditioning," *Learn. Behav.*, vol. 40, no. 3, pp. 305–319, 2012.

[33] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY, USA: Wiley, 2014.

[34] V. Krishnamurthy, *Partially Observed Markov Decision Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[35] S. P. Singh and R. S. Sutton, "Reinforcement learning with replacing eligibility traces," *Mach. Learn.*, vol. 22, nos. 1–3, pp. 123–158, 1996.

[36] D. M. Eagleman, "Time perception is distorted during slow motion sequences in movies," *J. Vis.*, vol. 4, no. 8, p. 491, 2004.

[37] M. Sucala, B. Scheckner, and D. David, "Psychological time: Interval length judgments and subjective passage of time judgments," *Current Psychol. Lett. Behav. Brain Cogn.*, vol. 26, no. 2, pp. 1–9, 2011.

[38] H. Lejeune and J. H. Wearden, "Scalar properties in animal timing: Conformity and violations," *Quart. J. Exp. Psychol.*, vol. 59, no. 11, pp. 1875–1908, 2006.

[39] L. J. Eaves, K. A. Last, P. A. Young, and N. G Martin, "Model-fitting approaches to the analysis of human behaviour," *Heredity*, vol. 41, no. 3, pp. 249–320, 1978.

[40] G. Lillacci and M. Khammash, "Parameter estimation and model selection in computational biology," *PLoS Comput. Biol.*, vol. 6, no. 3, 2010, Art. no. e1000696.

[41] W. D. Bradford, P. Dolan, and M. M. Galizzi, "Looking ahead: Subjective time perception and individual discounting," *J. Risk Uncertainty*, vol. 58, no. 1, pp. 43–69, 2019.

[42] A. Y. Ng and S. J. Russell, "Algorithms for inverse reinforcement learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 1, 2000, pp. 663–670.

[43] A. Mas-Colell *et al.*, *Microeconomic Theory*, vol. 1. New York, NY, USA: Oxford Univ. Press, 1995.

[44] R. Mattila, C. R. Rojas, V. Krishnamurthy, and B. Wahlberg, "Inverse filtering for hidden Markov models," in *Advances in Neural Information Processing Systems (NIPS)*. Red Hook, NY, USA: Curran, 2017, pp. 4207–4216.

[45] R. Mattila, I. Lourenço, C. R. Rojas, V. Krishnamurthy, and B. Wahlberg, "Estimating private beliefs of Bayesian agents based on observed decisions," *IEEE Control Syst. Lett.*, vol. 3, no. 3, pp. 523–528, Jul. 2019.

[46] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ, USA: Prentice-Hall, 1987.

[47] C. B. Do, *Gaussian Processes*, vol. 5, Stanford Univ., Stanford, CA, USA, 2007, p. 2017.

[48] G. E. Uhlenbeck and L. S. Ornstein, "On the theory of the Brownian motion," *Phys. Rev.*, vol. 36, no. 5, p. 823, 1930.

Inês Lourenço (Student Member, IEEE) received the M.Sc. degree in decision systems and control from the Electrical Engineering and Computer Science Programme, Instituto Superior Técnico, Lisbon, Portugal, in 2018. She is currently pursuing the Ph.D. degree with the KTH Royal Institute of Science, Stockholm, Sweden.

She was an intern with the Gulbenkian Science Institute, Oeiras, Portugal, and INESC-ID, Lisboa, and collaborated with Champalimaud Foundation, Lisbon. She received the KTH Electrical Engineering Scholarship of Excellence. Her primary research interests are within control of stochastic dynamical systems and biologically inspired robotics.

Robert Mattila (Student Member, IEEE) received the M.Sc. degree in systems, control, and robotics from the Engineering Physics Programme, KTH Royal Institute of Technology, Stockholm, Sweden, in 2015, and the Ph.D. degree from KTH in 2020, by submitting the thesis *Hidden Markov Models: Identification, Inverse Filtering, and Applications*.

In 2015, he was awarded the KTH Electrical Engineering Scholarship of Excellence. He has been a Visiting Researcher with the California Institute of Technology, Pasadena, CA, USA; the University of British Colombia, Vancouver, BC, Canada; and Cornell University, Ithaca, NY, USA. His primary research interests are within inference and control of stochastic dynamical systems.

Rodrigo Ventura (Member, IEEE) received the Licenciatura, M.Sc., and Ph.D. degrees in electrical and computer engineering from Instituto Superior Técnico (IST), Lisbon, Portugal, in 1996, 2000, and 2008, respectively.

He is an Assistant Professor (tenured) with IST, and a Permanent Member of Institute for Systems and Robotics, ISR-Lisboa, Lisbon. He has published more than 130 publications in peer-reviewed international journals and conferences, on various topics intersecting Robotics and Artificial Intelligence. He is also a co-inventor of several national and international patents on innovative solutions for robotic systems. Broadly, his research is focused on the intersection of artificial intelligence and robotics, with particular interests in human-robot collaboration, mobile manipulation, robot planning and control, cognitive robotics, and biologically inspired cognitive architectures. This research is mostly driven by research questions elicited by application in space robotics, aerial robots, social service robots, and urban search and rescue robots.

Dr. Ventura is a Founding Member of the Biologically-Inspired Cognitive Architecture society and alumni of the International Space University. He participated in several international and national research projects.

Bo Wahlberg (Fellow, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree from Linköping University, Linköping, Sweden, in 1983 and 1987, respectively.

In December 1991, he became a Professor of the Chair of Automatic Control with the KTH Royal Institute of Technology, Stockholm, Sweden, where he was a co-founder of Centre of Autonomous Systems and the Linnaeus Center ACCESS on networked systems. His research interests include system identification, modeling and control of industrial processes, machine learning, and statistical signal processing with applications in automated transportation systems.

Prof. Wahlberg has received several awards, including the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING Best New Application Paper Award in 2016. He is a member of the program management group of the Wallenberg AI, Autonomous Systems, and Software Program (WASP). He was a plenary speaker at the 35th Chinese Control Conference in 2016. He is a Fellow of IFAC for his contributions to system identification.