

Switching Attention in Time-Varying Environments via Bayesian Inference of Abstractions

Meghan Booker and Anirudha Majumdar¹

Abstract—Motivated by the goal of endowing robots with a means for focusing attention in order to operate reliably in complex, uncertain, and time-varying environments, we consider how a robot can (i) determine which portions of its environment to pay attention to at any given point in time, (ii) infer changes in context (e.g., task or environment dynamics), and (iii) switch its attention accordingly. In this work, we tackle these questions by modeling context switches in a time-varying Markov decision process (MDP) framework. We utilize the theory of *bisimulation-based state abstractions* in order to synthesize mechanisms for paying attention to context-relevant information. We then present an algorithm based on Bayesian inference for detecting changes in the robot’s context (task or environment dynamics) as it operates online, and use this to trigger switches between different abstraction-based attention mechanisms. Our approach is demonstrated on two examples: (i) an illustrative discrete-state tracking problem, and (ii) a continuous-state tracking problem implemented on a quadrupedal hardware platform. These examples demonstrate the ability of our approach to detect context switches online and robustly ignore task-irrelevant distractors by paying attention to context-relevant information.

I. INTRODUCTION

Mechanisms for focusing attention play a pivotal role in human cognition [1], [2]. In the field of neuroscience and psychology, attention is defined as the *preferential processing* of the stream of sensory information that an agent receives [2], [3]. This preferential processing is crucial for two reasons. First, *bounded rational* agents have a finite computational capacity and must choose certain portions of sensory information to process in order to perform a given task [4]. Second, by focusing only on *task-relevant* information, an agent can achieve robustness to uncertainty and variations in task-irrelevant distractors in its environment [5], [6], [1]. The motivating goal of this work is to endow robotic systems with similar capacities for focusing attention.

Any mechanism for focusing attention must be *context dependent*. In particular, what aspects of a robot’s state and environment it should attend to depends on two contextual factors: (i) the task that it is trying to accomplish, and (ii) the dynamics of its environment. As an illustrative example, consider a robot dog tasked with following one of two humans that has a “treat” (Fig. 1) that is not always within sight. Intuitively, the robot should pay attention to the state and dynamics of the human that has the treat, and ignore the other human in the environment. Now, if the treat switches from one human to the other, the robot must *infer* this change in context and switch its attention accordingly. Similarly, if the human that does not have the treat changes its behavior



Fig. 1. A robot dog is tasked with following one of two humans that has a “treat” (the AprilTag [7] held by the human wearing the red-shirt on the right). The robot observes the state of the two humans (the x and y distances from the robot), infers which human to pay attention to and therefore which to ignore, and then receives the reward of treats if it moves close enough to the human and maintains proximity. The humans can decide to switch who has the treat at any point, and the robot will subsequently need to switch its attention. Note that the robot is not tracking the AprilTag (which may be invisible at any given point in time), but rather the human who is inferred to have the treat.

and starts interfering with the motion of the human who has the treat, then the robot may need to pay attention to both.

In this work, we consider the following question: how can a robot determine what context-relevant information to attend to, infer when the context (i.e., task or dynamics) has changed, and switch its attention accordingly? We hypothesize that these abilities could improve the robustness of robots operating in complex and time-varying environments with multiple sources of distraction and uncertainty.

Statement of Contributions. The main technical contribution of this work is to formalize a means of switching attention to context-relevant information in an online setting. Specifically, we propose a model of context switches in terms of a time-varying Markov decision process (MDP); a given context then corresponds to a particular MDP, and context switches correspond to changes in reward function (i.e., task) and/or dynamics. We leverage the theory of *bisimulation-based state abstractions* [8], [9], [10] in order to define a mechanism for focusing attention in a given context. We then develop a Bayesian inference approach for determining when the underlying context has switched in order to switch between different abstraction-based attention mechanisms. We demonstrate the efficacy of our approach in terms of automatically inferring context switches and providing robustness to task-irrelevant distractors using two examples: (i) a simulated discrete-state tracking example, and

¹ Mechanical and Aerospace Engineering, Princeton University, NJ, USA. Emails: {mebooker, ani.majumdar}@princeton.edu.

(ii) a continuous-space tracking problem implemented on a quadrupedal hardware platform (Fig. 1).

II. RELATED WORK

Bioinspired attention. Mechanisms for focusing and switching attention have been extensively studied in cognitive science (e.g., [1], [2], [3]). As highlighted in Sec. I, the ability to attend to task-relevant sensory information based on context affords advantages in terms of computational efficiency and robustness to irrelevant distractors. Inspired by these potential benefits, researchers in control theory and robotics have sought to imbue artificial agents with various forms of bioinspired attention, e.g., attention to visual features [11], [12] or to particular agents in multi-agent settings [13], [14]. These works rely on hand-crafted attention mechanisms (e.g., carefully chosen feature representations). In contrast, our goal is to develop a systematic procedure that allows a robot to decide what to pay attention to.

Attention via state abstractions. In this work, we leverage the theory of *state abstractions* [10] for Markov decision processes (MDPs). Given a MDP, state abstraction techniques seek to find a mapping from the original state space (which can be large or high dimensional) to a compressed *abstract* state space. This mapping is constructed such that states which differ in a task-irrelevant manner are mapped to the same abstract state; the mapping thus defines a mechanism for focusing attention only on task-relevant information. This intuition is formalized by several state aggregation techniques based on the theory of bisimulation [8] and its variants (e.g., bisimulation metrics [15], [16], lax bisimulations [17], [18], [19], and policy-based bisimulations metrics [20], [21]). We provide a brief introduction to bisimulation and highlight its relationship to attention in Sec. IV.

Prior work on bisimulation typically assumes that the mapping from original to abstract states is *fixed* (since the agent is assumed to be operating in an environment described by a given MDP); thus, what the agent attends to does not change over time. Motivated by the need for robots to operate in complex and changing environments, our goal is to allow for *context-dependent* attention mechanisms. We formalize this in Sec. III by allowing the underlying environment MDP to change as the robot is operating. The robot must then infer when such switches occur and correspondingly switch what it is paying attention to by choosing an appropriate state abstraction.

Learning attention mechanisms end-to-end. An alternative to bisimulation-based attention is to learn the attention mechanism in an end-to-end manner. Attention mechanisms in transformer architectures [22] have demonstrated remarkable empirical success in language [22], [23] and vision tasks [24], [25]. Inspired by these successes, researchers in reinforcement learning and robotics have sought to end-to-end learn similar attention mechanisms [26], [27], [28]. Visual attention mechanisms have also been learned via imitation learning [29], [30] (e.g., from human gaze observations). These approaches could feasibly learn context-dependent attention mechanisms automatically. However, sample efficiency remains a significant challenge for such end-to-end

learning approaches. In contrast, we pursue a model-based approach in this work.

III. PROBLEM FORMULATION

Our goal is to provide a robot with a systematic means of focusing and *switching* its attention online. In other words, the robot needs to be able to determine what abstract state information is relevant, infer when the context it is operating in changes, and then correspondingly switch its attention mechanism (i.e., focus attention on new relevant abstract state information in the new context). To formalize this, we assume that the true underlying environment the robot is operating in is described by a time-varying Markov decision process (TVMDP) [31], $\mathcal{M}_t = (\mathcal{S}, \mathcal{A}, \mathcal{P}_t, \mathcal{R}_t)$, where \mathcal{S} is the discrete or continuous state space of the robot and its environment, \mathcal{A} is the action space, $\mathcal{P}_t(s_{t+1}|s_t, a_t)$ is the time-varying probability of transitioning from state s_t to s_{t+1} with action a_t , and $\mathcal{R}_t(s_t) \in \mathbb{R}$ is the time-varying function that assigns a reward for being in state s_t .¹ In this environment, the robot chooses actions according to a time-varying policy, $\pi_t(s_t)$, with the goal of maximizing the expected cumulative reward, $\max_{\pi_t} \mathbb{E}_{\mathcal{P}_t} [\sum_{t=0}^T \mathcal{R}_t(s_t)]$. A challenge, then, for determining where to focus attention lies in the fact that the state transition probabilities and reward functions are time-varying; as a consequence, what the robot should attend to is also time-varying (i.e., the relevant abstract state information is time-varying).

We define a *context* to be a time-invariant MDP, $\mathcal{M}^i = (\mathcal{S}, \mathcal{A}, \mathcal{P}^i, \mathcal{R}^i)$, indexed by $i \in \{1, 2, \dots, N\}$, that captures an instance of the true underlying environment \mathcal{M}_t . In other words, we assume that at any time, t , the underlying environment is some context, \mathcal{M}^i . However, the robot does not have knowledge of the context it is operating in. Instead, it can observe an imperfect version of the state and the reward (e.g., noisy depth measurements).

Next, we assume that for each context, \mathcal{M}^i , we are given an *abstraction*, $\phi^i = (\mathcal{Z}^i, \mathcal{A}, \mathcal{P}^{\mathcal{Z}^i}, \mathcal{R}^{\mathcal{Z}^i})$, defined as a MDP obtained from state abstractions for a context. Here, \mathcal{Z}^i is the compressed abstract state space of \mathcal{S} , $\mathcal{P}^{\mathcal{Z}^i} := \mathcal{P}^i(z_{t+1}^i | z_t^i, a)$ is the probability of transitioning from abstract state z_t^i to z_{t+1}^i , and $\mathcal{R}^{\mathcal{Z}^i} := \mathcal{R}^i(z_t^i)$ is the reward function for the i th abstraction. These abstractions provide a mechanism for focusing the robot’s attention to relevant state information. In particular, by constraining the robot’s policy to only depend on the *abstract* state, task-irrelevant differences in the full state are ignored by the robot. We provide a brief introduction to the bisimulation-based state abstractions used in this paper in Sec. IV.

The robot has a catalog, $\Phi = \{\phi^0, \phi^1, \dots, \phi^N\}$, of possible abstractions available to it and needs to infer which one corresponds to the context it is operating in. By construction, inferring abstraction ϕ^j at time t and then ϕ^i at $t + 1$ corresponds to switching the robot’s attention since \mathcal{Z}^i now abstracts the state space \mathcal{S} in a different way from \mathcal{Z}^j in

¹The TVMDP is different from a time-dependent MDP that augments the state with time. The time-dependent MDP is interpreted as modelling stochastic travel between states [31].

order to capture differing relevant information. For the scope of this work, we assume that the catalog is finite.

While inferring a particular abstraction, the robot chooses actions $a \in \mathcal{A}$ from a given policy $\pi^i(z_t^i)$, i.e., the robot has a policy maximizing the expected cumulative reward for an abstraction.² The context may switch online, and thus, the robot needs to infer the correct abstraction in order to perform well in the true context. This is the primary objective of this work as this corresponds to determining what the robot is paying attention to and when to switch. We describe our approach for inferring abstractions in Sec. V.

IV. ATTENTION VIA BISIMULATION

In this section, we provide an overview of bisimulation-based state abstraction [8], [9], [10] and its relation to attention. Intuitively, bisimulation is a state abstraction method that groups states into the same “abstract” state if they differ only in a behaviorally irrelevant way. More precisely, bisimulation considers two states to be equivalent if they have (i) equivalent immediate rewards, and (ii) equivalent distributions over the next abstract state. This is formalized in the definition below.

Definition 1 (Bisimulation Relation [9]). *Given a Markov Decision Process \mathcal{M} , a bisimulation relation is an equivalence relation B between states such that for all states $s_i, s_j \in \mathcal{S}$ that are equivalent under B , the following conditions hold:*

$$\mathcal{R}(s_i, a) = \mathcal{R}(s_j, a), \quad \forall a \in \mathcal{A} \quad (1)$$

$$\mathcal{P}(z|s_i, a) = \mathcal{P}(z|s_j, a), \quad \forall a \in \mathcal{A}, \forall z \in \mathcal{S}_B, \quad (2)$$

where \mathcal{S}_B is the partition of \mathcal{S} under the relation B (the set of all groupings z of equivalent states), and $\mathcal{P}(z|s, a) := \sum_{s' \in z} \mathcal{P}(s'|s, a)$.

The bisimulation relation induces a mapping from the full state s to an *abstract state* z (corresponding to the set of states that are equivalent to s under the bisimulation relation). We can interpret this mapping as a form of attention. Specifically, by mapping two states that are behaviorally equivalent (in the formal sense of Definition 1) to the same abstract state, bisimulation focuses only on relevant differences in the state.

As an example, consider the problem of following one of two humans with a treat (described in Sec. I and Fig. 1). Suppose that `human1` has the treat. In this case, any state s (corresponding to the locations of the two humans relative to the robot) that differ *only* in terms of `human2`’s location are equivalent under the bisimulation relation (since `human2`’s location does not impact the immediate reward or the dynamics of `human1`’s relative location). Thus, the full state s is mapped to an abstract state (corresponding to `human1`’s location relative to the robot), which focuses attention on `human1`. We highlight an important benefit of focusing attention in this manner. Any control policy that depends only on the abstract state has a very strong robustness property:

²Depending on the state abstraction method, the optimal policy of the context may be preserved [10].

its performance is not impacted by errors in estimation of the irrelevant portions of the state (e.g., `human2`’s position).

We note that there are a number of variations of the bisimulation framework discussed above which seek to relax the strictness of the equivalence relation, e.g., bisimulation metrics [15], [16], lax bisimulations [17], [18], [19], and policy-based bisimulations metrics [20], [21]. In this work, we will utilize the notion of bisimulation in Definition 1. Extending our framework to other notions of bisimulation is a promising direction for future work.

V. SWITCHING ATTENTION

The robot needs to infer which abstraction, ϕ^i , to apply online for a given time step in order to pay attention to task-relevant state features and perform well in the true underlying context. To do so, we first use bisimulation, discussed in detail in Sec. IV, as the state abstraction method for generating the catalog of abstractions, Φ , available to the robot. We use bisimulation because (i) we can leverage the reward as a signal for inferring the correct abstraction to apply (as elucidated further below), (ii) we obtain policies that are robust to noisy state observations, and (iii) there are existing approaches for synthesizing the abstractions and corresponding policies (e.g., [8], [9]).

We now describe our approach for tackling the problem formulated in Sec. III and inferring which abstraction (and corresponding policy) to apply at any given time. Since the identity of the underlying context is not observed directly, we implement a Bayesian filter to maintain a belief over the abstraction catalog. Specifically, the robot is provided with an abstraction transition model $p(\phi_t^i | \phi_{t-1}^j)$, an abstraction detection model $p(r_t | \phi_t^i, s_t)$, and an initial prior belief over the abstraction catalog $p(\phi_0^i)$. The abstraction transition model captures the uncertainty associated with how the abstractions, and therefore contexts, evolve over time. This and the prior belief provide an opportunity to encode expert or domain knowledge if it is available. The abstraction detection model is pre-generated using the abstractions available to the robot. The probability of a particular reward is one if it matches the abstract state’s reward associated with the given abstraction ϕ^i and observed state s_t . Thus it acts as a “detector” for an abstraction switch; intuitively, if the robot receives rewards that deviate from its current belief over abstractions, it can infer that the underlying context has switched.

After updating its belief over abstractions, the robot chooses which abstraction to apply by taking the maximum likelihood estimate. This selection effectively maps the observed, and potentially noisy, state s_t to the abstract state z_t^i and allows the robot to employ the corresponding policy $\pi^i(z_t^i)$. We summarize this attention switching process in Algorithm 1 and refer to it as Online Attention Switching (OAS).

VI. EXAMPLES

Here we illustrate the efficacy of our OAS algorithm described in Sec. V on two examples: (i) an illustrative discrete-space tracking example, and (ii) a continuous-space tracking problem where a robot dog has to follow the human with the “treat”. We use these examples to demonstrate

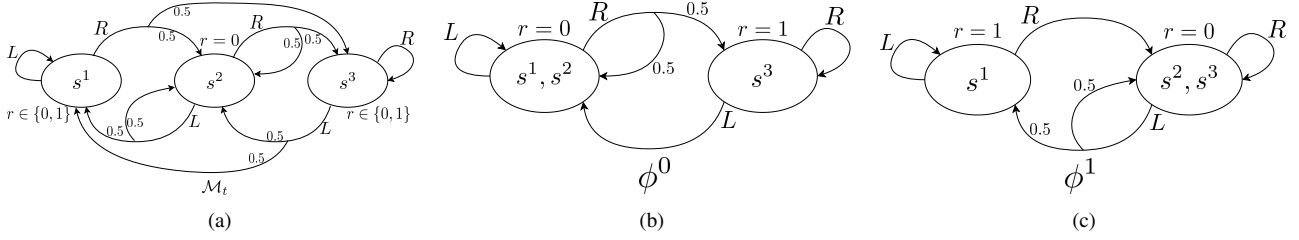


Fig. 2. The problem setup for the discrete tracking example. The robot needs to track the object but the object may appear in states s^1 or s^3 due to unexpected environment disturbances. When the robot is in the same state as the object, it receives a reward of 1 otherwise 0. (a) The TVMDP for this example. States are denoted in the ovals, actions are either L or R, the state transition probabilities are labelled on the arrows, and the rewards for each state are denoted outside the states with r . (b) The first abstraction, ϕ^0 , acquired by performing bisimulation for when the object is in state s^3 . States s^1 and s^2 are mapped to an abstract state together while state s^3 is its own abstract state. This corresponds to the robot paying attention to being with the object or not. (c) The second abstraction, ϕ^1 , acquired by performing bisimulation for when the object is in state s^1 .

Algorithm 1: Online Attention Switching (OAS)

```

Initialize environment state:  $s_0 \sim \mathcal{S}$ 
Initialize maximum likelihood (ML) abstraction:
 $i \leftarrow \arg \max p(\phi_0^i)$ 
for  $t \in \{1, 2, \dots, T\}$  do
  Apply action according to ML abstraction:
   $a_t \leftarrow \pi^i(z_{t-1}^i)$ 
  Robot takes action  $a_t$  and observes  $s_t, r_t$ .
  for all  $\phi^i$  do
    Dynamics update:
     $p(\phi_t^i | r_{1:t-1}, a_{1:t}, s_{t-1}) =$ 
     $\sum_j p(\phi_t^i | \phi_{t-1}^j) p(\phi_{t-1}^j | r_{1:t-1}, a_{1:t-1}, s_{t-1})$ 
    Measurement update:
     $p(\phi_t^i | r_{1:t}, a_{1:t}, s_t) =$ 
     $\frac{p(r_t | \phi_t^i, s_t) p(\phi_t^i | r_{1:t-1}, a_{1:t}, s_{t-1})}{\sum_j p(r_t | \phi_t^j, s_t) p(\phi_t^j | r_{1:t-1}, a_{1:t}, s_{t-1})}$ 
  end
   $i \leftarrow \arg \max_i p(\phi_t^i | r_{1:t}, a_{1:t}, s_t)$ 
end

```

that OAS is able infer the correct abstractions efficiently (and therefore perform well in the true contexts) as well as highlight the robustness benefits from using attention.

A. Discrete-Space Tracking Example

Scenario. In this first example, we consider a robot that has to track an object in a three-state grid-world where the object’s location suddenly changes and can appear in either the first or last state. For example, ocean currents or wind may be pushing the object to the different states. The robot can move to follow the object with actions left or right and if it is in the same state as the object, it receives a reward of 1; otherwise it receives a reward of 0. Concretely, the state space is $\mathcal{S} = \{s^1, s^2, s^3\}$, the action space is $\mathcal{A} = \{L, R\}$, the state transition probabilities \mathcal{P} , illustrated in Figure 2(a), are time-invariant, and the possible reward functions are $\mathcal{R}_t \in \{1_{s_t=s^1}(s_t), 1_{s_t=s^3}(s_t)\}$ where $1_{s_t=x}(s_t)$ is 1 if the robot is in state x and 0 otherwise.

Contexts and abstractions. The possible contexts for this TVMDP are induced by the two possible reward functions. Performing bisimulation on these two contexts, as defined in Definition 1, results in the abstractions illustrated in Figure 2(b-c), where the abstract states are capturing whether the robot is paying attention to if it is with the target object or

not. Notice that with these abstractions, the robot is robust to noisy state observations. For example, in ϕ^0 , states s^1 and s^2 are mapped to the same abstract state. Thus, if the true state is s^2 but the observed state is s^1 , the robot will still perform the same actions since ϕ^0 considers these to be the same state, i.e., the robot pays attention to the fact that it is not with the object not whether it is in exactly s^1 or s^2 .

Results. Next we simulated different switching patterns for the true underlying contexts including *step* (where the context remains constant for a number of time steps and then switches to the other context), *periodic* (where the context switches back and forth periodically), and *uniformly random* (where a context is chosen randomly at every time step) for 500 time steps and allowed the robot to randomly choose its actions. We also ran these simulations for two different abstraction transition models where (i) the probability of maintaining the same abstraction is high ($p(\phi_t^i | \phi_{t-1}^i) = 0.8$), and (ii) the probability of maintaining the same abstraction or changing is uniform ($p(\phi_t^i | \phi_{t-1}^i) = 0.5$). Additionally, we provided the robot with noisy states under the same patterns. Specifically, we assign a probability, σ , that the observed state is another possible state rather than the true state. Then we averaged across five seeds the accuracy,³ average lag,⁴ and maximum lag for the maximum likelihood abstraction determined by OAS. These results are summarized in Table I. Example trials for a step, periodic, and random pattern for the first abstraction transition model are illustrated in Fig. 3.

In general, the high accuracy values and low average lags for the first abstraction transition model demonstrate that OAS captured the attention switches well and quickly, with an average lag of 0.5 even for a challenging noisy periodic scenario. The expected robustness benefits of using attention are also demonstrated by the fact that there is virtually no change in performance on these statistics when a high state estimation noise value, σ , is introduced. The second abstraction transition model also performed well; however it experienced more bouncing behavior as seen in the step results where the lag is 0 (i.e., switched at the correct time) but lower accuracy than the first transition model.

³Accuracy is calculated as the number of time steps the robot correctly inferred the abstraction over the time horizon.

⁴Lag is calculated as the number of time steps it takes for OAS to infer the correct abstraction after a context switch and before the next. Average lag is over the number of switches in the trial.

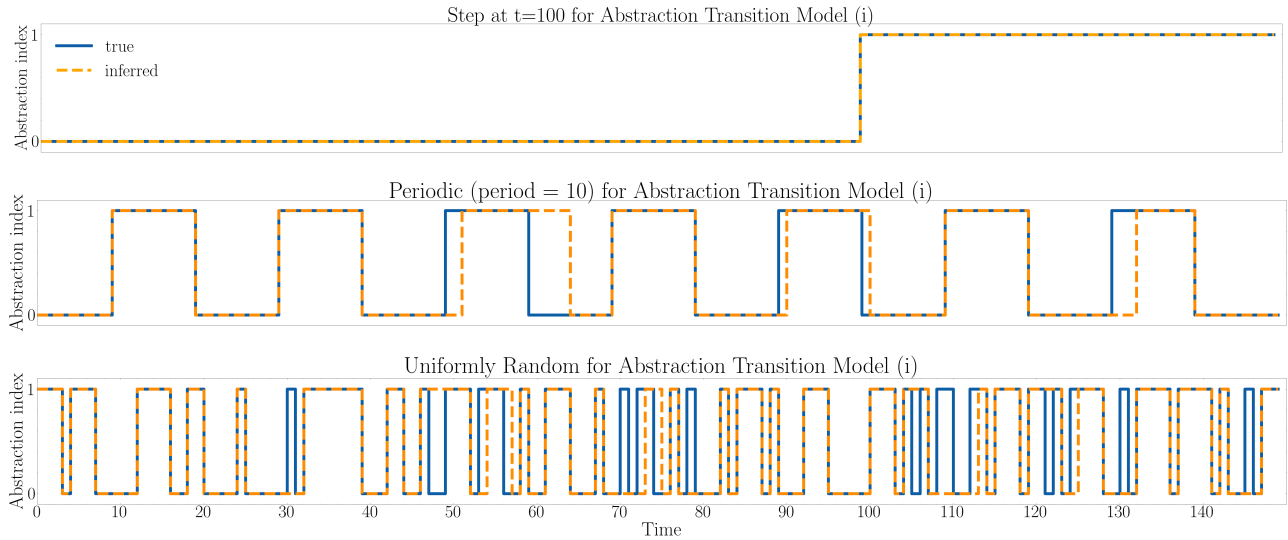


Fig. 3. Visualizations of attention switching for the first 150 time steps of the different switching patterns using the first abstraction transition model. The true abstraction index is shown in solid blue while the inferred abstraction is shown as a dotted orange line. **Top:** Step occurs at time step 100. **Middle:** Periodic switching with a period of 10. **Bottom:** Uniformly random switching.

TABLE I
KEY STATISTICS FOR THE DISCRETE-SPACE TRACKING EXAMPLE

Switching Pattern	Accuracy	Avg. Lag	Max. Lag
Abstraction Transition Model (i)			
Step (at $t = 100$)	0.99 ± 0.00	0.10 ± 0.20	0.20 ± 0.40
Noisy step (at $t = 100, \sigma = 0.7$)	0.99 ± 0.00	0.70 ± 0.98	0.80 ± 1.17
Periodic (period = 10)	0.93 ± 0.001	0.70 ± 0.1	6.80 ± 1.72
Noisy periodic (period = 10, $\sigma = 0.7$)	0.95 ± 0.01	0.48 ± 0.05	4.00 ± 0.63
Uniformly random	0.84 ± 0.01	0.31 ± 0.03	3.80 ± 1.17
Abstraction Transition Model (ii)			
Step (at $t = 100$)	0.83 ± 0.01	0.00 ± 0.00	0.00 ± 0.00
Noisy step (at $t = 100, \sigma = 0.7$)	0.83 ± 0.01	0.50 ± 0.32	1.00 ± 0.63
Periodic (period = 10)	0.84 ± 0.01	0.22 ± 0.06	2.60 ± 0.49
Noisy periodic (period = 10, $\sigma = 0.7$)	0.83 ± 0.02	0.24 ± 0.11	2.40 ± 1.02
Uniformly random	0.84 ± 0.01	0.19 ± 0.03	2.40 ± 0.49

B. Continuous-Space Tracking Example

Scenario. Here, we implement the scenario where a robot dog needs to pay attention to one of the two humans with the “treat” in order to receive it (Fig. 1). In this scenario, the humans may walk around arbitrarily as long as the human with the treat stays within the robot’s field of view. Meanwhile, the other human may move out of the field of view of the robot but may not try to crash into the robot. Additionally, the humans may hand off the treat to the other human at any time. As a result, the robot needs to switch its attention to the new human to continue receiving treats.

Hardware implementation. For hardware implementation (Fig. 1), we use the Unitree Go1 quadruped robot with an Intel Realsense RGB-D camera. The camera provides noisy observations of the state, $s_t = [x_1, y_1, x_2, y_2]$, corresponding to the positions of human1 and human2

relative to the robot. We utilize a pre-trained neural network (implemented as part of the TensorFlow Object Detection API [32]) in order to detect humans using RGB images; the depth image is then used to estimate the relative locations of the humans. We also assume the ability to identify each human individually (since the state s_t corresponds to the locations of each individual human); in our hardware setup, we implement this by distinguishing shirt colors. The actions are constant velocity commands corresponding to left, right, straight, right and straight, and left and straight. Additionally the robot may choose to stop. The “treat” corresponds to an AprilTag [7] that may be passed from human to human.

Contexts. We model the scenario described with different contexts arising from the reward function changing over time. Specifically, the reward changes from 1 to 0 when the robot is following a particular human who then passes, or potentially tosses, the treat to the other human. We highlight that this corresponds to a change in the reward *function*, since the same state is assigned different reward values after the treat is passed. We note that the robot does not utilize the AprilTag treat for the purpose of tracking; the AprilTag is simply used to assign rewards (specifically, a reward of 1 if the robot is close enough to the human that is holding the treat). Thus, the attention switch that the robot must perform is intuitively similar to a dog’s: if it stops receiving treats, this serves as a signal to pay attention to and follow the other human instead.

Abstractions. Bisimulation on the two different contexts generates abstractions that capture the aforementioned intuition. In the context where human1 has the treat, the state abstraction considers states $[x_1, y_1, x_2, y_2]$ and $[x_1, y_1, x'_2, y'_2]$ to be equivalent. In other words, the state abstraction focuses attention on the location of human1 and ignores the location of human2. The robot’s control policy is a function of this abstract state. By focusing only on context-relevant information (i.e., the location of the human who has the treat), the robot’s policy is robust to noise in the state estimate of the

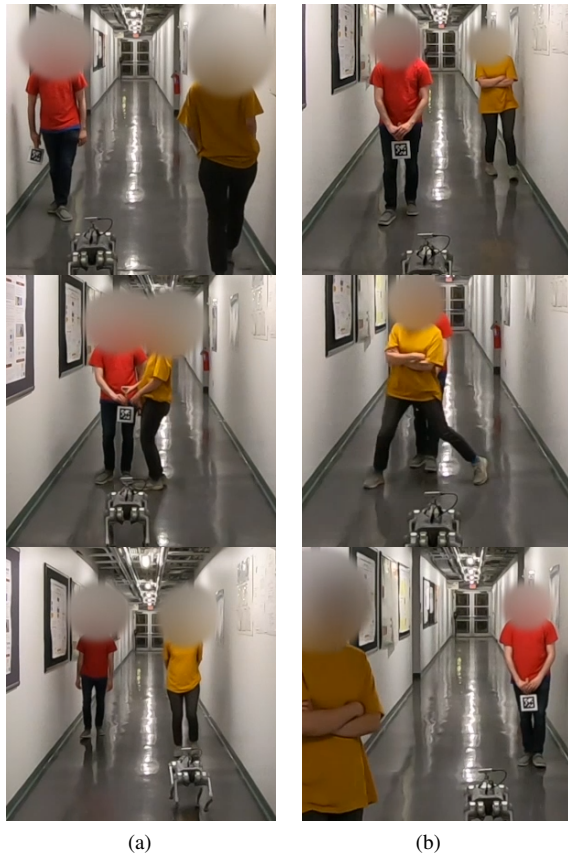


Fig. 4. Instances of the continuous-space tracking example. (a) An example of a treat hand-off (AprilTag) between the two humans followed by an attention switch for the robot dog. Top: the robot dog is following human1 (red) that has the treat. Middle: human1 passes the treat to human2 (yellow). Bottom: the robot dog has switched its attention correctly to human2 and is tracking human2 (not the treat). (b) An example of the robot dog ignoring the human that does not have the treat (human2). The progression of images (top to bottom) shows human2 crossing from right to left in front of human1 while the robot dog continues to follow human1.

human who does not have the treat. Thus, when the robot is paying attention to and following human1, the other human may wander far from the robot (leading to a noisy state estimate) without impacting the robot’s performance. See the Appendix of our extended version for more discussion on the sensor noise experienced for this example [33].

Results. For the trials, we had two humans, human1 wearing red and human2 wearing yellow, walk with the robot down a hallway. When not holding the treat, the human performed behaviors such as walking in and out of the robot’s field of view, zig-zagging behind the human with the treat, and crossing in front of the human with the treat. Meanwhile, the human holding the treat often blocked the treat from view when moving around in order to highlight that the robot was paying attention to and tracking them (rather than the AprilTag). Additionally, each trial contained 1-3 treat hand-offs to the other human and lasted approximately 1 minute (with the robot operating at 5 Hz). The abstraction transition model set the probability of maintaining the same abstraction as 0.8. We refer the reader to our video⁵ to view these behaviors, as well as qualitatively view the robot inferring

context switches using the OAS algorithm and switching its attention and behavior accordingly. Two example interactions are shown with picture progressions in Fig. 4.

TABLE II
KEY STATISTICS FOR THE CONTINUOUS-SPACE TRACKING EXAMPLE

Accuracy	Average Lag (sec)	Max. Lag (sec)	Normalized Reward
0.91 ± 0.07	0.37 ± 0.24	2.31 ± 1.48	0.27 ± 0.11

We summarize key statistics (introduced in Sec. VI-A) of our OAS approach averaged across 12 trials in Table II. In general, the high accuracy indicates that the robot performed well at maintaining its attention on the correct human. The normalized reward is the cumulative reward averaged over the time horizon of a trial. Note that the humans could decide to act adversarially, e.g., walking further away as the robot approached them. The lags (reported in seconds where one time step corresponds to 0.2s each) were very low, indicating that the attention switches occurred in a timely manner.

VII. CONCLUSIONS AND FUTURE WORK

We have presented an approach for allowing a robot to determine what context-relevant information it should attend to, infer when the context (i.e., task or dynamics) has switched (among a catalog of potential contexts), and switch its attention accordingly. To achieve this, we leverage bisimulation-based state abstraction techniques in order to define an attention mechanism in a given context. A Bayesian inference algorithm then operates online in order to detect context switches by maintaining a belief that is updated using reward and state observations. We demonstrated our method on an illustrative discrete-state tracking example and hardware demonstrations of a tracking problem using a quadruped robot.

Challenges and Future Work. There are a number of exciting directions for future work. First, we expect that our approach will generalize to state abstraction techniques beyond the (strict) notion of bisimulation used here (e.g., to metric-based bisimulation notions). Second, it may be of practical interest to relax the finiteness assumption on the catalog of possible state abstractions, and allow for a continuously parameterized set by using Bayesian inference techniques that operate on continuous spaces (e.g., particle filters). Finally, a particularly interesting direction is to synthesize policies that allow the robot to *actively* minimize uncertainty in its belief about the appropriate state abstraction to utilize. This potentially requires a careful tradeoff between actions that reduce uncertainty with those that make progress towards the task.

ACKNOWLEDGEMENTS

The authors were supported by the School of Engineering and Applied Science at Princeton University through the generosity of William Addy ’82, the NSF CAREER award [2044149], and the Office of Naval Research [N00014-23-1-2148, N00014-21-1-2803]. The authors are also thankful for the input provided by Eric Lepowsky and Susan Redmond.

⁵See video of results <https://youtu.be/1cL1Pcf3qmk>.

REFERENCES

- [1] I. C. Fiebelkorn and S. Kastner, "Functional specialization in the attention network," *Annual Review of Psychology*, vol. 71, p. 221, 2020.
- [2] E. Styles, *The Psychology of Attention*. Psychology Press, 2006.
- [3] T. J. Buschman and S. Kastner, "From behavior to neural dynamics: An integrated theory of attention," *Neuron*, vol. 88, no. 1, pp. 127–144, 2015.
- [4] H. A. Simon, "Theories of bounded rationality," *Decision and organization*, vol. 1, no. 1, pp. 161–176, 1972.
- [5] G. Gigerenzer and H. Brighton, "Homo heuristics: Why biased minds make better inferences," *Topics in cognitive science*, vol. 1, no. 1, pp. 107–143, 2009.
- [6] G. Gigerenzer and W. Gaissmaier, "Heuristic decision making," *Annual Review of Psychology*, vol. 62, pp. 451–482, 2011.
- [7] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, May 2011, pp. 3400–3407.
- [8] K. G. Larsen and A. Skou, "Bisimulation through probabilistic testing," *Information and Computation*, vol. 94, no. 1, pp. 1–28, 1991.
- [9] R. Givan, T. Dean, and M. Greig, "Equivalence notions and model minimization in Markov decision processes," *Artificial Intelligence*, vol. 147, no. 1-2, pp. 163–223, 2003.
- [10] L. Li, T. J. Walsh, and M. L. Littman, "Towards a unified theory of state abstraction for MDPs," in *AI&M*, 2006.
- [11] S. Vijayakumar, J. Conradt, T. Shibata, and S. Schaal, "Overt visual attention for a humanoid robot," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 4, 2001, pp. 2332–2337.
- [12] F. Rea, G. Metta, and C. Bartolozzi, "Event-driven visual attention for the humanoid robot iCub," *Frontiers in neuroscience*, vol. 7, p. 234, 2013.
- [13] A. Bizyaeva, A. Matthews, A. Franci, and N. E. Leonard, "Patterns of nonlinear opinion formation on networks," in *Proceedings of the IEEE American Control Conference*. IEEE, 2021, pp. 2739–2744.
- [14] A. Bizyaeva, A. Franci, and N. E. Leonard, "Nonlinear opinion dynamics with tunable sensitivity," *IEEE Transactions on Automatic Control*, 2022.
- [15] N. Ferns, P. Panangaden, and D. Precup, "Bisimulation metrics for continuous Markov decision processes," *SIAM Journal on Computing*, vol. 40, no. 6, pp. 1662–1714, 2011.
- [16] A. Zhang, R. McAllister, R. Calandra, Y. Gal, and S. Levine, "Learning invariant representations for reinforcement learning without reconstruction," *arXiv preprint arXiv:2006.10742*, 2020.
- [17] J. Taylor, D. Precup, and P. Panangaden, "Bounding performance loss in approximate MDP homomorphisms," *Advances in Neural Information Processing Systems*, vol. 21, 2008.
- [18] B. Ravindran and A. G. Barto, "Relativized options: Choosing the right transformation," in *International Conference on Machine Learning*, 2003, pp. 608–615.
- [19] E. van der Pol, D. Worrall, H. van Hoof, F. Oliehoek, and M. Welling, "MDP homomorphic networks: Group symmetries in reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4199–4210, 2020.
- [20] P. S. Castro, "Scalable methods for computing state similarity in deterministic Markov decision processes," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 06, 2020, pp. 10069–10076.
- [21] R. Agarwal, M. C. Machado, P. S. Castro, and M. G. Bellemare, "Contrastive behavioral similarity embeddings for generalization in reinforcement learning," *arXiv preprint arXiv:2101.05265*, 2021.
- [22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [23] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.
- [24] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [25] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8821–8831.
- [26] Y. Tang, D. Nguyen, and D. Ha, "Neuroevolution of self-interpretable agents," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2020, pp. 414–424.
- [27] L. Chen, K. Lu, A. Rajeswaran, K. Lee, A. Grover, M. Laskin, P. Abbeel, A. Srinivas, and I. Mordatch, "Decision transformer: Reinforcement learning via sequence modeling," *Advances in neural information processing systems*, vol. 34, pp. 15 084–15 097, 2021.
- [28] S. James and A. J. Davison, "Q-attention: Enabling efficient learning for vision-based robotic manipulation," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1612–1619, 2022.
- [29] C. Pfeiffer, S. Wengeler, A. Loquercio, and D. Scaramuzza, "Visual attention prediction improves performance of autonomous drone racing agents," *PLoS one*, vol. 17, no. 3, p. e0264471, 2022.
- [30] A. Makrigiorgos, A. Shafti, A. Harston, J. Gerard, and A. A. Faisal, "Human visual attention prediction boosts learning & performance of autonomous driving agents," *arXiv preprint arXiv:1909.05003*, 2019.
- [31] L. Liu and G. S. Sukhatme, "A solution to time-varying Markov decision processes," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1631–1638, 2018.
- [32] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7310–7311.
- [33] M. Booker and A. Majumdar, "Switching attention in time-varying environments via Bayesian inference of abstractions," *arXiv preprint arXiv:2211.05865*, 2022.