Attentional Shifting and Curiosity: a Reinforcement Learning Approach

Dario di Nocera*, Alberto Finzi†, Silvia Rossi† and Mariacarla Staffa†§

*Dipartimento di Matematica, Università degli Studi di Napoli "Federico II", via Cintia 21, 80125 Napoli, Italy
†Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione (DIETI),
Università degli Studi di Napoli "Federico II", via Claudio 21, 80125 Napoli, Italy
Email: {dario.dinocera, alberto.finzi, silvia.rossi, mariacarla.staffa}@unina.it
§Authors are listed in strict alphabetical order.

Abstract—Attention allocation strategies represent key capabilities of human beings, which are strictly connected with action selection and execution mechanisms, while intrinsic motivations directly affect the allocation of attentional resources. In this paper we propose a model of Reinforcement Learning (RL), where both these capabilities are involved. RL is deployed to learn how to allocate attentional resources in a behavior-based robotic system, while action selection is obtained as a side effect of the resulting motivated attentional behaviors. Moreover, the influence of intrinsic motivations in attention orientation is obtained by introducing internal rewards associated with curiosity drives. In this way, the learning process is affected not only by goal-specific rewards, but also by intrinsic motivations depending on the internal state of the system.

I. INTRODUCTION

THE concepts of attention and motivations are of great interest in adaptive robotic systems. These mechanisms can be exploited to guide, activate, and coordinate multiple concurrent behaviors linking their variability to the circumstances in which they are activated. Beyond their role in perception orientating and filtering, *attentional mechanisms* are considered as key capabilities of human beings, which are strictly connected with action selection and execution [Norman and Shallice, 1986]. On the other hand, intrinsic motivations directly affect the modulation and the control of attentional resources, and affect action selection process. For example, *curiosity* is considered the main drive for human beings to explore novel situation and to learn complex behaviors from experience [Berlyne, 1954], [Litman, 2005]. Moreover, recent studies have shown that both attention and curiosity are strictly related to the dopaminergic system responsible for action driving. It is widely accepted, indeed, that dopamine affects both the reward exciting, fundamental in the learning process, and the demand of more attention by novel stimuli [Nieoullon, 2002], [Jepma et al., 2012].

The Reinforcement Learning [Montague et al., 1996] represents the main technique adopted as a computational model of the dopamine-like reward system. Recent works incorporate models for novelty and curiosity, within Motivated RL algorithms [Baranes et al., 2013], [Singh et al., 2010], [Kaplan and Oudeyer, 2003], to allow behavior adaptation and action selection in response to the changing experiences of a robot. Typically, within these approaches, RL is used to directly model and optimize action selection strategies. Differently from these approaches, starting from a previous work [Di Nocera et al., 2012], we propose a model of action selection, which is indirectly provided by the learning of attentional allocation strategies by means of intrinsically motivated RL, where the states exploration policy is modulated by the curiosity. We refer to the model of attentional mechanisms proposed by [Burattini et al., 2010], while for the curiosity model we take inspiration from the model considered by [Litman, 2005]. As for the Motivated RL we consider an additional internal rewarding system, whose purpose is to drive the agent to progress in learning [Berlyne, 1954]. In the following section we give a brief introduction on models that underlie the mechanisms of attention and curiosity adopted in this work. Whereupon, we present our Motivated RL model, where the learning of the attentional strategies is driven by curiosity. Finally, we detail our approach by describing a case study of a simulated survival domain. Our main aim is to investigate the impact of curiosity on attentional shifting and, consequently, on action selection.

II. BACKGROUND

A. Attentional Shifting System

In this work, we refer to the attentional allocation framework introduced in [Burattini et al., 2010]. In contrast with typical work on visual attention [Itti and Koch, 2001], Burattini et al. approach is not concerned with the orientation of the attention in the space (i.e. the field of view), instead it is about executive attention [Posner et al., 1980] and the temporal distribution of attentional resources needed to monitor and control multiple processes. This model of attention is inspired by [Pashler and Johnston, 1998], where the attentional load due to the accomplishment of a particular task is defined as the quantity of attentional time units devoted to that particular task, and by [Senders, 1964], where attentional allocation mechanisms are related to the sampling rate needed to monitor multiple parallel processes. Burattini et al. [Burattini et al., 2010] propose, in fact, a frequency-based model of attention allocation for regulating the sensors sampling rates and behaviors activations. Such an attentional allocation mechanism is able to increase/decrease the arousal level of each behavior with respect to salient internal/external stimuli, by changing

the activation frequencies of the behaviors and, consequently, their priorities.

The model consists in a Behavior-based architecture [Arkin, 1998], [Brooks, 1986], where each behavior is endowed with an attentional mechanism represented by an internal adaptive clock [Burattini and Rossi, 2008].
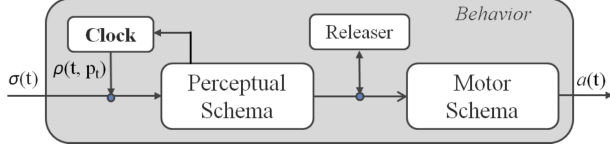

Fig. 1. Schema theory representation of an attentional behavior.

The schema theory representation of an attentional behavior is illustrated in Figure 1. This is characterized by a Perceptual Schema (PS), which elaborates sensor data, a Motor Schema (MS), producing the pattern of motor actions, and an attentive control mechanism, called Adaptive Innate Releasing Mechanism (AIRM), based on a combination of a clock and a releaser. The releasing mechanism works as a trigger for the MS activation (e.g., the view of a predator releases the escape behavior), while the clock regulates sensors sampling rate and, consequently, the rate of possible behaviors activations. The period p_t^b of each clock ranges in an interval $[p_{min}^b, p_{max}^b]$ specific for each behavior b. The clock enables the data flow σ_t^b from sensors to PS by means of a *trigger function* $\rho(t, p_t^b) \in {0, 1}$ every p_t^b time units. A *monitoring function* $f(\sigma_t^b, p_{t-1}^b) : \mathbb{R}^n \to \mathbb{N}$ adjusts the current clock period p_t^b, according to both internal states and environmental changes.

In Section III we will detail our approach, showing how we can learn the attentive monitoring function by means of a RL technique, taking also into account the influence of the curiosity intrinsic motivation.

B. Intrinsic Motivations: Curitosity

Curiosity is an appetitive state involving the recognition, pursuit, and intense desire to investigate novel information and experiences that demand ones attention. In literature, we find two main theoretical accounts of curiosity: the *optimal arousal* [Litman and Jimerson, 2004] and *curiosity-drive* [Berlyne, 1950]. The curiosity-drive model assumes that the main drive of curiosity is the reduction of uncertainty: novel and ambiguous stimuli cause a need for a coherence restore that reduces the uncertainty. This reduction is considered as rewarding. This model is supported by studies showing that unusual situations are associated with approaching behaviors and attentional states (e.g., see the Loewenstein's knowledge gap/approach gradient [Loewenstein, 1994]). However, the curiosity-drive model cannot explain why biological organisms initiate exploratory behaviors without any stimuli. These situations are instead well explained by the optimal-arousal model (e.g., see the Spielberger and Starr model [Spielberger and Starr, 1994]). Following this model, the biological systems are associated with an homeostatic regulation of their arousal

level: when the arousal level is understimulated, the organism is motivated to increase the arousal and look for novel situations; in contrast, when the organisms is overstimulated additional stimuli are evaluated as negative and associated with an avoidance behavior. While in the curiosity-drive model the reward is associated with uncertainty reduction, in the optimal arousal model, the induction of curiosity is directly rewarding. Also this model is not completely satisfactory, because the gain of new knowledge could reduce the feeling of interest. This could be considered counter-productive if we assume that the arousal state should be maintained (see [Litman, 2005]).

A combination of these two approaches is proposed by [Litman, 2005] with *interest/deprivation* model of curiosity. Here, both the satiation and the activation of curiosity can be rewarding: the interest-based curiosity is driven by novel stimuli and opportunity of learning, whereas deprivation-based curiosity is driven by the uncertainty and lack of knowledge. The interest/deprivation model of curiosity is related to the neuroscience of the wanting and liking systems [Berridge, 2003], which are hypothesized to underlie motivation and affective experience for a broad class of appetites. In this case, wanting is associated with deprivation and need of knowledge, while liking is associated with the expected pleasure due to learning and knowledge acquisition.

Liking	Wanting	
	Low	*High*
Low	LL: Ambivalent disinterest	LH: Need for uncertainty clarification
High	HL: Curiosity as a feeling of "interest"	HH: Curiosity as a feeling of "deprivation"

TABLE I
LITMAN'S CLASSIFICATION OF CURIOSITY STATES WITH RESPECT TO HIGH AND LOW LEVELS OF LIKING AND WANTING [LITMAN, 2005].

In Table I we show the Litman classification [Litman, 2005]. In this paper, we exploit a model of curiosity that is inspired by the Litman's combination of wanting and liking. In our approach, the interpretation of these two are different and related to our behavior-based architecture (see Section III-c).

III. MOTIVATED RL FOR ATTENTIONAL SHIFTING

Following the approach presented in [Di Nocera et al., 2012], we design a new method for learning the attentional shifting strategies by exploiting a RL algorithm [Sutton and Barto, 1998], where also intrinsic motivations are considered. Our aim is to understand how intrinsic motivations, such as curiosity, could affect the learning process. In particular, a Q-learning with a softmax (Boltzman) exploratory policy is used to tune and adapt the sensors sampling frequencies, while action selection is obtained as a side effect.

a) The Action Space: AS^b is the set of possible periods of activation $P = \{p_1, \ldots, p_n\}$, which can be assigned to the behavior b. Thus, an action a^b is an assignment of the clock period p^b for each behavior b in every state, determining the behavior activation frequency. The idea is that the system does not directly learn the action to be executed, but it learns

attentional policies for adapting the activation frequencies of each behavior, providing an action selection that indirectly emerges from a particular configuration of the behaviors attentional values. Furthermore, different attentional shifting strategies will be learned depending on the level of curiosity of the agent.

b) The State Space: SS^b is defined as follows. The state s^b is determined by a triple (c^b, p^b, σ^b), where c^b represents the level of curiosity of the behavior, p^b is for the current clock period, and σ^b is the current perceptive state of a behavior b. In particular the behaviors attentional monitoring periods p^b range in a predefined set of four possible values $[p_1^b, p_2^b, p_3^b, p_4^b]$ for each behavior. The perceptive state σ^b is considered to be a partitioning of the total perceptive domain in equidimensional intervals (i.e., each perceptive state is a sub-range of the input signal). Finally, c^b ranges in $[LL, LH, HL, HH]$ meta-spaces defining the relation between *wanting* and *liking* values inspired by the curiosity model definition of Litman [Litman, 2005] (see section II-B). Therefore, the attentional allocation policy $\pi^b : SS^b \rightarrow AS^b$ represents a mapping between the current state s^b and the next value for the attentional period p^b, and should be learned by means of the QL algorithm.

c) Extrinsic and Intrinsic Rewards: Once defined the Action and State spaces, we define the reward as follows:

$$R^b = (1 - w) * R_e^b + (w) * R_i^b \qquad (1)$$

where R_e^b is the *extrinsic* reward computed considering the observed state, and R_i^b represents the *intrinsic reward* evaluated considering the satisfaction of an observation with respect to a particular curiosity state. The value of R_i^b is thus computed as level of *liking*. The w value represents the level of *wanting*, an internal unmotivated desire to explore something (nothing particular or something specific depending on liking level). Our assumption is that the level of wanting depends on a sort of (global) energy state of the agent (see section IV). The idea is that the robotic agent can explore new situations, guided by curiosity, only when the system is in a wellness state, while, when the system is under a certain wellness threshold, attentional shifting strategies maintain attention focused on priority behaviors (e.g., EAT and DRINK) rather than on secondary ones (unaware exploration of new states).

Liking	Wanting	
	Low	High
Low	$R_e^b \gg R_i^b$	$R_e^b < R_i^b$
High	$R_e^b > R_i^b$	$R_e^b \ll R_i^b$

TABLE II
WANTING AND LIKING RELATIONS AND THE ASSOCIATED RELATIONS BETWEEN INTRINSIC AND EXTRINSIC REWARDS.

These mechanisms will imply that, when the situation is critical the R_i^b intrinsic reward value will be neglected with respect to the R_e^b extrinsic reward value, while R_i^b will gain more influence in determining the reward value as much as the agent will be in a wellness state. Relations between R_i^b, R_e^b and *wanting*, *liking* values are synthesized in Table II.

IV. CASE STUDY

To test our approach we introduce the following *Survival Domain*. The robot must survive for a predefined amount of time within an environment avoiding obstacles and recharging energy by eating and drinking. We consider simulated environments of $16m^2$. Obstacles, water, and food locations are cubes of size $0.5m \times 0.5m \times 0.5m$, respectively of black, blue, and green color (see Figure 7-left). An experiment ends in a positive way if the robot is able to survive till the end of the test (max_time), while it fails in three cases: the robot collides with an obstacle or dies of hunger or thirst. We test our approach using a simulated *Pioneer3-DX* mobile robot endowed with a blob camera and 16 sonar sensors (using the Player/Stage tool http://playerstage.sourceforge.net/).

A. Internal Needs Functions

We assume that the robot is endowed with internal drives. In our case study, we consider three internal needs: huger, thirst, and sensation of safety. The *Hunger function* is to compute the need for food:

$$Hunger(t) = Hunger(t-1) + k \times (nb\_act) \\ - (e_f \times food\_consumed). \qquad (2)$$

Here, the hunger increases the need for food at each machine cycle by a k value for each active behavior (nb_act), and decreases it when a quantity of food is ingested ($food\_consumed$), depending on the energy power of the food (e_f). An analogous *Thirst function* is used to compute the need for water:

$$Thirst(t) = Thirst(t-1) + k \times (nb\_act) \\ - (e_w \times water\_consumed). \qquad (3)$$

Finally, a *Safety function* computes the need of safety and depends on the current degree of danger:

$$Safety(t) = \frac{min(\sigma_t^a)}{\sigma_{max}^a} \times \frac{p_{max}^a - p_t^a}{p_{max}^a - p_{min}^a}, \qquad (4)$$

where the level of safety is calculated with respect to the minimum distance between the robot current position and an obstacle (σ_t^a). When the distance decreases and the AVOID activation period (p_t^a) is relaxed the safety decreases; viceversa, the safety increases when the activation period of the avoid is suitably balanced with respect to the distance from an obstacle.

B. Attentional Behaviour-Based Architecture

We consider three behaviors to model the robotic system: AVOID, EAT and DRINK (see Figure 2). The domain for AVOID spans in the interval $[0, \sigma_{max}^a]$; the domain of DRINK is $[0, \sigma_{max}^d]$, where σ_{max}^d represents the maximum value for the thirst function; the EAT domain is in $[0, \sigma_{max}^e]$, where σ_{max}^e is the maximum state of hunger the robotic system can assume. The perceptive state of each behavior is obtained as a discretization of the perceptive domain using 4 equidimensional intervals $[\sigma_1^b, \sigma_2^b, \sigma_3^b, \sigma_4^b]$ used for indexing the learning state space by means of the triple (c^b, σ^b, p^b). In the case

study, assuming the minimum clock period as 1 machine cycle, the possible period set for the AVOID, EAT and DRINK is: $p^a, p^e, p^d = \{1, 4, 8, 12\}$.
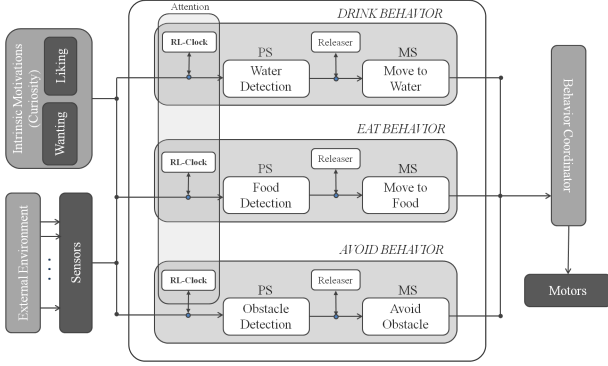

Fig. 2. Attentional Behavior-Based Architecture with intrinsic motivations.

d) Extrinsic Rewarding function: We assume the reward always positive except for a strong *penalty* if the system is not able to survive. For the other cases the reward is computed, for each behavior, considering two additive components. The first evaluates the impact of frequent activation of a specific behavior. The higher the frequency, the smaller is the obtained reward. This component is equal to zero if $p_t^b = p_{min}^b$. The second component is related to the specific behavior. In particular, concerning AVOID, each activation is rewarded directly proportional to the distance from the obstacle.

$$R_e^a(t) = \begin{cases} \frac{1}{2} * \left[\frac{p_t^a - p_{min}^a}{p_{max}^a - p_{min}^a} + \left(\frac{\sigma_t^a - \sigma_{min}^a}{\sigma_{max}^a - \sigma_{min}^a} \right) \right], & \text{if !crash} \\ penalty, & \text{otherwise} \end{cases}$$
(5)

As for EAT behavior, for each activation, the reward is inversely proportional to the current hunger value. A system that is hungrier takes a smaller reward.

$$R_e^e(t) = \begin{cases} \frac{1}{2} * \left[\frac{p_t^e - p_{min}^e}{p_{max}^e - p_{min}^e} + \left(1 - \frac{\sigma_t^e}{\sigma_{max}^e} \right) \right], & \text{if !crash} \\ penalty, & \text{otherwise} \end{cases}$$
(6)

Finally, each activation of DRINK is rewarded in a way that is inversely proportional to the current value of thirst:

$$R_e^d(t) = \begin{cases} \frac{1}{2} * \left[\frac{p_t^d - p_{min}^d}{p_{max}^d - p_{min}^d} + \left(1 - \frac{\sigma_t^d}{\sigma_{max}^d} \right) \right], & \text{if !crash} \\ penalty, & \text{otherwise} \end{cases}$$
(7)

For our experiments we adopt the following settings:

- *penalty*: $-max\_cycles$ maximum penalty (-1500 units of penalties, where 1500 is the medium number of cycles for episode);
- *max_time*: maximum time allowed to accomplish the task (180 seconds);
- σ_{max}^a: maximum sonar range (1 meter);
- σ_{min}^a: safety minimum distance (0.4 meters);
- σ_{max}^e: maximum value for σ_t^e (1500 units of charge);
- σ_{min}^e: minimum value for σ_t^e over which the robot needs to Eat (300 units of charge);

- σ_{max}^d: maximum value for σ_t^d (1500 units of charge);
- σ_{min}^d: minimum value over which the robot needs to drink (300 units of charge).

We choose as learning rate $\alpha = 0.8$, while for what concerns the SS, we refer to the tests implemented in [Di Nocera et al., 2012] considering a 24 states setting, since it has been shown this represents the best regulation in this kind of environment.

C. Motivated Attentional Framework

Intrinsic rewards are modeled by Curiosity, which, according to the description of above, we divide into two emotional components dealing respectively with the feeling of *wanting* and *liking*. We link the first one to the concept of *residual energy* for the robot body, while the second one to the level of *novelty* in the exploration of the learning states.

We consider a global value $E(t)$ called *Energy*, that takes into account all the variables regulated by the three needs functions defined above.

$$\begin{aligned} E(t) = & E(t-1) - e_u - e_{nb} * (nb\_act) + \\ & e_f * (food\_consumed) + e_w * (water\_consumed) \end{aligned}$$
(8)

where the current value of the energy $E(t)$ is computed starting from the previous level of energy $E(t-1)$, decremented of one unit of energy e_u supposed to be the energy consumed at each machine cycle. Then, we also consider the energy spent to activate each behavior e_{nb}, where nb_act is the number of currently active behaviors. On the other hand, we assume increments of the energy in correspondence of consummatory behaviors, such as EAT or DRINK, where the quantity e_f/e_w of energy, related to a particular food/water object, is assumed when these are consumed (i.e. if boolean conditions related to $food\_consumed$ and $water\_consumed$ are true).
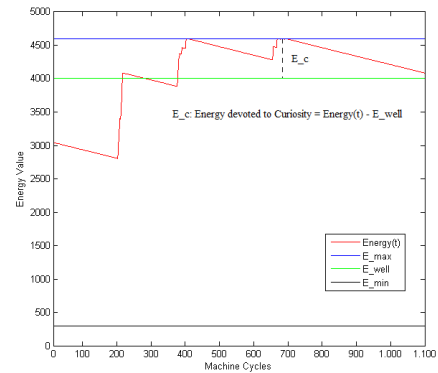

Fig. 3. *Energy(t)* is the curent energy level; *E_well*: is the level of energy corresponding to a wellness state of the system; the *blackline* is for the minimum amount of energy permitting the system to work.

We model the *wanting* component of the Curiosity as the residue of the Energy value (3).

$$w = E_c = E(t) - E\_well$$
(9)

That is to say that the robot can show a curious behavior only when the situation is not critic (i.e. only when the global energy exceeds the E_well threshold, indicating a sort of *wellness* state of the system). E_well is supposed to be associated to a state of the system where the regulation of the periods of activation of the different behaviors is well balanced and leads to a suitable scheduling of the actions. We can interpret this residue value E_c as the energy that the system can spent not giving attention to prior behaviors, but just going around searching to reduce its curiosity or to move towards *liking* things. The second component of the curiosity is the *liking*, which we model as quantity related to new situations and represents the internal reward. In particular, since the curiosity in our system is interpreted as the exploration within the learning states space, we can assume that the novelty of a space is computed as follows:

$$l = R_i^b = 1 - \frac{NV(\sigma_t^b)}{NV\_tot} \tag{10}$$

where, NV is for number of visits and $\frac{NV(\sigma_t^b)}{NV\_tot}$ represents the number of times the percept σ_t^b has been observed during the previous NV_tot observations. We, thus, maintain a sort of temporal window of value NV_tot. In this way, on the one hand, we have an idea of how new is the observation; on the other hand, considering a finite temporal window, we simulate a sort of lapsing mechanism of the novelty of the states observations. The model of the temporal window can be compared to the Itti's model of *surprise* [Baldi and Itti], by interpreting the temporal window as the approximation of a statistic on the percept. That is to say that if the system is not observing a percept for NV_tot times, the stimulus becomes likable again. While, if the system observes that particular perceptual state σ_t^b many times (i.e. NV_tot), the stimulus associated becomes boring. The Curiosity modeled as above (depending on both liking and indirectly from global Energy value of the system) will affect the learning progress by providing an internal reward (details will be provided in next paragraphs) based on the wanting and linking combination.

V. Experimental Results

In order to observe how the curiosity affects the learning process we first compare the convergence, in terms of survival time percentage, of the learning process of a robotic system endowed with intrinsic motivation ($CR = CuriousRobot$) with one that does not ($NCR = Not - CuriousRobot$). The plot in Figure 4 shows that during the first 220 episodes the NCR system is more able to survive in the environment, in fact, the survival time percentage starts from a value of 92%. This could be due to the fact that the curiosity, initially, leads the system to prefer the exploration of novel spaces rather than goal-directed ones. After a while the CR system starts to rapidly increase its survival time until it over pass the NCR system and reach the convergence (100% of the survival time) at the episode 278, w.r.t. the NCR system that does not reach the convergence before the episode 382.
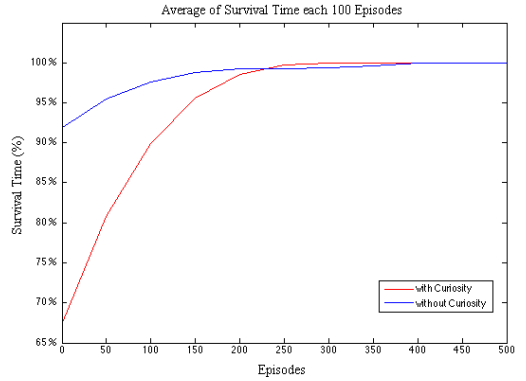

Fig. 4. Survival time percentage (averages each 100 episodes). Comparison between CR and NCR systems.

Figure 5 shows the cumulative rewards for each behavior during the learning process. As expected, during the first
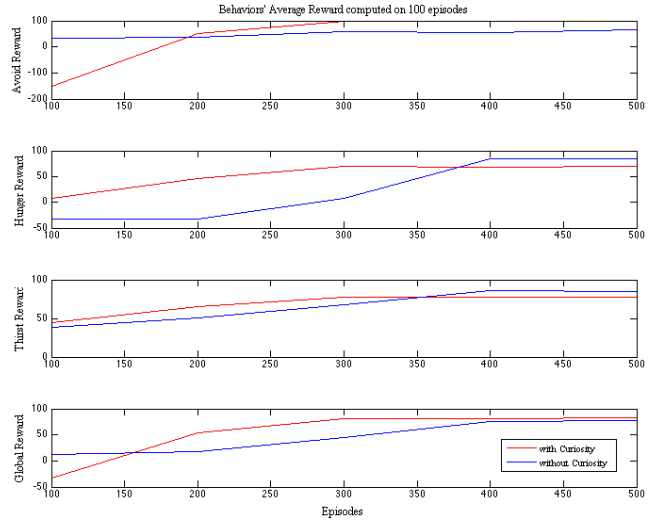

Fig. 5. Comparison between Curious (red line) and Not-curious (blue line) Robot rewarding values during learning process.

episodes, the CR fails in the sense that it is not able to learn suitable attentional strategies for regulating the behaviors activation. The cumulative rewards related with the CR AVOID behavior are worst w.r.t the NCR one until the episode 200, then they start to increase and to converge from the episode 300. Instead, the NCR shows a trend that seems not satisfactory concerning the cumulative rewards for the EAT and DRINK behaviors. This could be due to the fact that the robot is not guided by curiosity to immediately explore the spaces of the environment where there are food or water. It just learns to eat or drink when the associated needs functions exceed certain thresholds. Despite at the end of the experiments NCR EAT and DRINK rewards converge to better values, the global cumulative rewards show a faster convergence for the CR (see Figure 5). Moreover, if we look at the trend of the needs functions (see Figure 6), we observe a stable rhythmic path for each function for the CR. We
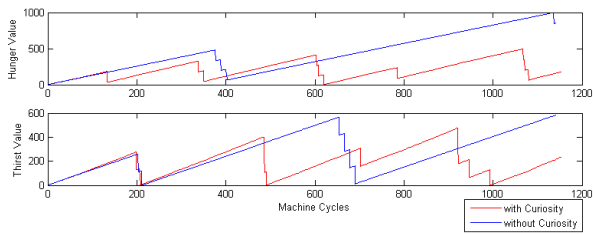
Fig. 6. Vital Functions Comparison.

interpret the plots, in the case of the CR, as an effective learned attentional shifting policy of the behaviors EAT and DRINK. The robot seems to find a rhythmic alternation in its needs for eating and drinking (the decreasing part of the two functions corresponds to the consuming of food/water), while the NCR just waits to become very hungry/thirsty for searching sources of food/water. Finally, an unexpected result is that the motivated RL has lead the learning process not only at exploring new internal learning states, but it has also affected the spatial exploration of the environment. In Figure 7 we observe that the CR is more explorative (the cumulative traces of 500 episodes covered the 50% of the total area) with respect to the NCR (44% of the total area covered).
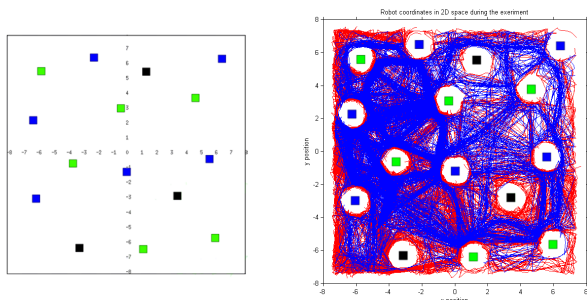

Fig. 7. The experimental environment and the cumulative paths of the CR (red line) and NCR (blue line) during the learning process.

VI. Conclusions

In this work we proposed an intrinsically motivated RL, where the states exploration policy is modulated by the curiosity, while rewards are evaluated not only considering external cues, but also some internal motivations. The collected results show that the approach is feasible and effective and this learning process achieves very good results in orchestrating behaviors. That is, the curiosity-driven RL, applied to attentional shifting, allows to improve behaviors coordination and action selection. It provides the robot with an increased adaptation skill, permitting the robot to survive in the environment. The system learns how to dynamically and suitably adapt the behaviors activation rates in order to balance the regulation of the needs functions and of the global energy of the system.

Acknowledgment

References

[Arkin, 1998] R. C. Arkin. (1998), *Behavior-Based Robotics*, MIT Press, CA.

[Baldi and Itti] P. F. Baldi, L. Itti, Of bits and wows: A Bayesian theory of surprise with applications to attention, *Neural Networks*, Vol. **23**, No. 5, pp. 649-666, Jun 2010.

[Baranes et al., 2013] A. Baranes, P-Y Oudeyer, (2013), Active Learning of Inverse Models with Intrinsically Motivated Goal Exploration in Robots, *Robotics and Autonomous Systems*, **61(1)**, 49-73.

[Berlyne, 1950] D. E. Berlyne (1950), Novelty and curiosity as determinants of exploratory behavior. *British Journal of Psychology*, 41, pp. 68–80.

[Berlyne, 1954] DE. Berlyne (1954), A theory of human curiosity, *British Journal of Psychology*, **45**, 180-191.

[Berridge, 2003] K. C. Berridge (2003), Pleasures of the brain. Brain and Cognition, 52, 106-128.

[Brooks, 1986] R. Brooks (1986), A Robust Layered Control System for a Mobile Robot, *IEEE Journal of Robotics and Automation*, Vol. **RA-2**, No. 1, pp. 14-23.

[Burattini and Rossi, 2008] E. Burattini and S. Rossi, (2008), Periodic adaptive activation of behaviors in robotic system, *International Journal of Pattern Recognition and Artificial Intelligence IJPRAI* **22(5)**, pp. 987999.

[Burattini et al., 2010] E. Burattini, A. Finzi, S. Rossi and M. Staffa, (2010), Attentive Monitoring Strategies in a Behavior-based Robotic System: an Evolutionary Approach, In *Proceedings of the 2010 International Conference on Emerging Security Technologies* (EST '10). IEEE Computer Society, Washington, DC, USA, 153-158.

[Di Nocera et al., 2012] D. Di Nocera, A. Finzi, S. Rossi and M. Staffa (2012), Attentional Action Selection using Reinforcement Learning, in *Proceedings of 12th International Conference on Adaptive Behaviour (SAB 2012)*, LNCS **7426**, 371-380.

[Itti and Koch, 2001] L. Itti, and C. Koch, (2001), Computational modeling of visual attention. In *Nature Reviews Neuroscience*, **2(3)**, pp. 194-203.

[Jepma et al., 2012] M. Jepma, R. Verdonschot, H. van Steenbergen, S. Rombouts and S. Nieuwenhuis (2012), Neural mechanisms underlying the induction and relief of perceptual curiosity, *Frontiers In Behavioral Neuroscience*,**6**.

[Kaplan and Oudeyer, 2003] F. Kaplan and P.Y. Oudeyer (2003), Motivational principles for visual know-how development. In *C.G. Prince et al. (Editors), Proceedings of the 3rd international workshop on Epigenetic Robotics*, Lund University Cognitive Studies, pp. 73-80.

[Litman and Jimerson, 2004] J. A. Litman and T. L. Jimerson (2004), The measurement of curiosity as a feeling-of-deprivation, *Journal of Personality Assessment*, 82, pp. 147–157.

[Litman, 2005] J. Litman (2005), Curiosity and the pleasures of learning: Wanting and liking new information, *Cognition and Emotion*, **19(6)**, pp. 793-814.

[Loewenstein, 1994] G. Loewenstein (1994), The psychology of curiosity: A review and reinterpretation. Psychological Bulletin, pp. 75-98.

[Montague et al., 1996] PR. Montague, P. Dayan and TK Sejnowski (1996), A framework for mesencephalic dopamine systems based on predictive Hebbian learning, *Journal of Neuroscience*, **16**, 1936-1947.

[Nieoullon, 2002] A. Nieoullon (2002), Dopamine and the regulation of cognition and attention, *Progress in neurobiology*, **67(1)**, 53-83.

[Norman and Shallice, 1986] D. Norman and T. Shallice (1986), Attention in action: willed and automatic control of behaviour, *Consciousness and Self-regulation: advances in research and theory* **4**, 1-18.

[Pashler and Johnston, 1998] H. Pashler, and J. C. Johnston (1998), Attentional limitations in dual-task performance, In *H. Pashler (Ed.), Attention*, East Essex, UK: Psychology Press, pp. 155-189.

[Posner et al., 1980] M. Posner, C. Snyder and B. Davidson (1980), Attention and the detection of signals, *Journal of Experimental Psychology General*, **109:**160-174.

[Senders, 1964] J. Senders, (1964), The human operator as a monitor and controller of multidegree of freedom systems, pp. 26.

[Singh et al., 2010] S. Singh, R. L. Lewis, A. G. Barto and J. Sorg (2010), Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective, *IEEE Trans. on Auton. Ment. Dev.*, **2**, 70-82.

[Spielberger and Starr, 1994] C. D. Spielberger and L. M. Starr (1994). Curiosity and exploratory behavior. In H. F. O'Neil, Jr. and M. Drillings (Eds.), Motivation: Theory and research, pp. 221-243.

[Sutton and Barto, 1998] R. Sutton and A. Barto (1998), Reinforcement learning: An introduction, vol. 1. Cambridge Univ. Press.