

Autonomous object modeling and exploiting: a new approach based on affordances from continual interaction with environment

Simon Gay, Alain Mille, Amélie Cordier

► To cite this version:

Simon Gay, Alain Mille, Amélie Cordier. Autonomous object modeling and exploiting: a new approach based on affordances from continual interaction with environment. The Seventh Joint IEEE International Conference on Developmental Learning and Epigenetic Robotics (ICDL-Epirob), Sep 2017, Lisbon, Portugal. hal-01632771

HAL Id: hal-01632771

<https://hal.archives-ouvertes.fr/hal-01632771>

Submitted on 10 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Autonomous object modeling and exploiting: a new approach based on affordances from continual interaction with environment

Simon L. Gay

Université de Lyon, CNRS
Université Lyon 1, LIRIS, UMR5205
Villeurbanne, F-69622, France
Email: simon.gay@liris.cnrs.fr

Alain Mille

Université de Lyon, CNRS
Université Lyon 1, LIRIS, UMR5205
Villeurbanne, F-69622, France
Email: alain.mille@liris.cnrs.fr

Amlie Cordier

Université de Lyon, CNRS
Université Lyon 1, LIRIS, UMR5205
Villeurbanne, F-69622, France
Email: amelie.cordier@univ-lyon1.fr

Abstract—We present an architecture for self-motivated agents to generate behaviors in an environment that is continuous, both in space and time, through a continual interaction process. The long-term goal is to design agents that construct their own knowledge of objects and space through experience of the environment, rather than exploiting pre-coded knowledge. The agent exploits this constructed knowledge to exhibit behaviors satisfying its self-motivated principles, based on valences attributed to interactions, that specify inborn behavioral preferences. Over time, the agent learns the relation between its perception of objects and the interactions that they afford, in the form of data structures, called signatures of interaction. The agent keeps track of enacted interactions in a spatial memory in which it can use signatures to recognize and localize distant possibilities of interactions, and exhibits behaviors that satisfy its motivation principles, accordingly to this approach. In this paper, we propose a continual decision cycle between an agent and its environment to cope with the constraints that an artificial agent would meet in a physical environment.

I. INTRODUCTION

In this paper, we address the problem of the construction, interpretation and exploitation of a short-term memory representation of an environment offering a continual interaction process between an artificial agent and this environment. Our work is based on a model described in Georgeon and Aha [4], called Radical Interactionism (RI), in which action-perception couples are considered as *atomic* and kept embedded within data structures called interactions, modeling Piaget’s notion of sensorimotor scheme [13]. This approach assumes that new agents come up with a predefined set of uninterpreted interactions associated with predefined valences, and seek to enact interactions with positive valence and to avoid interactions with negative valences. This motivation principle is called *interaction motivation* [4], and is related to the problem of *intrinsic motivation* [10]. The agent perceives and *interprets* its environment by identifying affordances rather than recognizing objects on the basis of predefined features. This approach addresses the knowledge grounding problem [7] by letting knowledge of objects arise from experience.

Our long term goal is to implement artificial agents in robots to make them able to interact with initially unknown envi-

ronments. Such an agent must overcome severe constraints. Let’s take an example of a planetary exploration rover that must autonomously face environmental situations that can be partially or completely unknown, implying several constraints:

1) The agent cannot use a predefined set of object models. Instead, the agent must learn to recognize interactional opportunities (or *affordances*) offered by the environment.

2) The agent cannot rely on a predefined internal model, as the rover may move on unknown surfaces. Thus, the rover must construct an internal representation of space based on its possibilities of interactions.

Instead of trying to solve a specific class of problems, we try to make an agent able to learn to integrate its environment in order to generate behaviors satisfying its motivational principles. Point 1) is well studied in literature. For example, Pfeifer and Scheier [12], Ugur *et al.* [16] and Baleia [1] proposed approaches where an agent learns to define and classify objects of its environment through possibilities of interaction (such as “walkthroughable” or not). However, these approaches do not let the possibility to generate spatial behaviors. Pisokas and Nehmzow [11] and Ugur *et al.* [15] proposed models that learn relations between actions and affordances to generate plans to reach affordances, but rely on predefined goals and cannot be used to generate exploitable models of surrounding space. Point 2) implies that the agent can construct a peripersonal[8] and extrapersonal [14] representations of its environment.

II. DEVELOPMENTAL LEARNING AND RADICAL INTERACTIONISM

Developmental learning [17] makes the assumption that the agent is agnostic of its environment [5]. This assumption prohibits considering that the agent can identify *states*, both for its environment and for itself. The agent can only know its environment through its own interactional experience and uses its experience to construct cognitive structures that it can use to satisfy intrinsic motivational principles. This work proposes to consider interactions as atomic [4], rather than separated actions and perceptions. The interactional cycle thus reviews

the traditional cycle as considered by literature to model agent-environment interactions. Experience is constructed from the emergence of regularities in interactional sequences. This framework is now questioned for studying how objects can emerge from interactions of an agnostic agent with its environment. The theoretical complexity of this question is potentially huge and classical approaches come up against this problem because of the combinational complexity of the environment.

The RI approach only considers the point of view of the agent, which theoretically simplifies this question as the limited ability to experience the environment reduces intrinsically the computational complexity. The issue of objects emergence within the agent's interactional space is critical for proposing a new kind of agent's adaptation process when its environment is largely unknown. This research work addresses the design of an agent architecture satisfying the RI principles, able to build spatial structures and dynamic structures taking into account emerging objects. Several RI models were developed with the long-term aim of implementing them in robots moving in a real environment [3][2]. However, these models used a discrete interaction cycle, which cannot be applied on a physical robot. In this paper, we present a main milestone of the RI model allowing a *continual* interaction cycle between an agent and its environment.

A. Formalization of Continual RI Model

An agent has a set of actions (A) and perceptions (O) characterized by its actuators and sensors. Theories of cognition suggest that perception cannot be separated from the action that generates it. Indeed, the same perception can have different meanings, depending on the action that produced it. Previous models of RI used interactions under the form of couple action-perceptions. Each interaction i is attributed an inborn valence ν_i characterizing the intrinsic *satisfaction* of experiencing i . The decision cycle starts with a decision of the agent, that intends to *enact* an *intended interaction* i_t and experiences, at the end of the decision cycle, the *enacted interaction* e_t that was effectively experienced. An RI agent tries to capture regularities offered by its environment to generate behaviors that allow to enact interactions with positive valences and avoid interactions with negative valences.

In order to make a continual interaction cycle, we propose to define an interaction i as a set $i = \{(\epsilon_k, t_k)\}_{k \in [0; n]}$, where $\epsilon_k \in A \cup O$ and $t_k \in \mathbb{R}^-$. The agent is equipped with a register Φ of length Δt_Φ consisting in a timeline that is continually updated through time and contains actions and perceptions experienced between $t - \Delta t_\Phi$ and t . The length of this register must be defined according to interactional possibilities and must be long enough to help the agent characterize its movement. As an example, an agent with motion sensors (such as a vestibular system) can characterize its movement with a few elements, while an agent with no vestibular system would need a longer record of past actions and perceptions. This doesn't affect functioning of the model, but only affects efficiency and relevance of the generated internal model. At each t , it is thus possible to define e_t , an enacted interaction as the set of

couples (ϵ_k, t_k) in the register ($e_t \equiv \Phi_t$). We note I the set of possible interactions.

We adapted the RI model for a continual decision cycle, using these continual interactions. We define, between two interactions i_{t_1} and i_{t_2} , with $t_2 - t_1 < \Delta t_\Phi$, a variation Δi containing the set of couple (ϵ_k, t_k) that must be added to i_{t_1} to obtain i_{t_2} . We only consider elements that must be added, as couples that move out of Φ will have disappeared at t_2 . When the agent experiences consecutively i_{t_1} and i_{t_2} , we consider that i_{t_1} is a *predecessor* of i_{t_2} and note $i_{t_1} + \Delta i_{1,2} \rightarrow i_{t_2}$, where $\Delta i_{1,2}$ is the *variation* between i_{t_1} and i_{t_2} . We can then consider that the agent is continually intending a variation of interaction di_t and continually experiences enacted interaction e_t through enacted variations de_t .

The Continual RI (CRI) is formalized as follows: the agent begins with a set $I_{t_0} \subseteq I$ of interactions and a set ΔI of possible variations di . At each interaction i of I is attributed an inborn valence ν_i characterizing the satisfaction of experiencing i . The agent continually tries to intend a variation di_t and experiences the enacted variation de_t . The register Φ is updated to define the enacted interaction e_t that the agent experienced at t . The intention of the agent is a success if $di_t = de_t$ and a failure otherwise. If $di_t \neq de_t$ then de_t is considered as an *alternative* of di_t . Figure 1 illustrates this formalism.

B. A Parallel CRI Model

Discovering spatial properties of the environment with a unique interaction at a time is unrealistic. The Parallel Radical Interactionism (PRI) model [3] was proposed to overcome this limitation. In PRI model, the agent experiences additional stimuli, in addition to the enacted interaction. However, these stimuli have to be considered with the movement produced by the enacted interaction. As an example, the optic flow on a retina can only carry a spatial information if it is considered with the movement that generates it. The PRI model defines *secondary interactions* as an association between an interaction and an additional stimulus. We can apply this principle to CRI model to define the Continual PRI (CPRI), because we make the assumption that register Φ , and thus e_t , can characterize the movement of the agent at time t . We call a *primary interaction* a set $i_p = \{(\epsilon_k, t_k)\}_k$ and a secondary interaction an association between an interaction and an additional stimulus $i_s = (i_{\{p,s\}}, (\epsilon_s, 0))$, with $i_{\{p,s\}}$ the *associated interaction* of i_s .

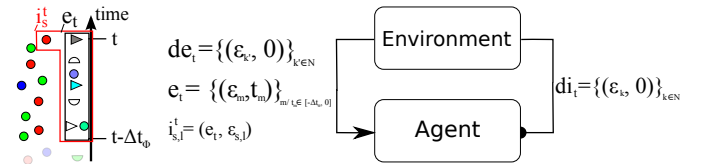


Fig. 1. Model of the Continuous Radical Interactionism. The agent continually tries to enact a variation of interaction $di \in \Delta I \subseteq \mathcal{P}((A \cup O) \times \mathbb{R}^-)$, and experiences the variation $de \in \Delta I$ that is actually *enacted*. Experienced variations are stored on a timeline of length Δt_Φ that is continually updated, defining the enacted interaction e_t . In the parallel CRI model, the agent also *experiences* several additional stimuli, that are associated with e_t to form secondary interactions.

Formally, the CPRI model differs from CRI in that, at time t , the agent experiences a set of enacted interactions $\{e_k\}_t$, containing a unique primary interaction and a set of enacted secondary interactions associated with this primary interaction. A secondary interaction i_s , like a primary interaction, succeeds when $i_s \in \{e_k\}_t$, and fails when its associated interaction i_s is in $\{e_k\}_t$, but $i_s \notin \{e_k\}_t$. When the associated interaction is not enacted, the result of i_s is ignored.

III. A CONTINUAL SPACE MEMORY

Space memory is a set of mechanisms dedicated to the integration and exploitation of environmental properties observed through interaction with the environment. This section presents concepts and principles of the space memory in a continual interaction process.

A. Signatures of interactions

This mechanism is based on the assumption that the result of enacting an interaction depends on a limited context of elements in the environment. We expect such contexts to define *objects* with which the agent can interact. This definition of objects relates to the concept of affordances proposed by Gibson [6]. A RI agent can only perceive its environment through enacted interactions. We define the interactional context E_t , characterizing the environment experienced by the agent, as the continuous sets of enacted interactions that were enacted between $[t-\Delta t_E, t]$, where Δt_E is the length of the interactional context. Δt_E is independent from Δt_Φ , and must be long enough to let the agent *detect* the causes of its interactions. CRI model allows an interesting simplification: two primary interactions i_1 and i_2 such that $i_1 + \Delta i_{1,3} \rightarrow i_3$ and $i_2 + \Delta i_{2,3} \rightarrow i_3$ only differ by their oldest elements, as $\Delta i_{1,3} = \Delta i_{2,3}$. If we assume that oldest elements have a lower influence on the movement, it is possible to consider that i_1 and i_2 characterize similar movements. Then, it is possible to consider secondary interactions $i_{s,1,k} = (i_1, (\epsilon_k, 0))$ and $i_{s,2,k} = (i_2, (\epsilon_k, 0))$ as the same input in context E .

We define a *signature* of an interaction i as a structure learned through experience that can characterize the presence or the absence of the object affording i in a context E , and thus, the enactability of i . We formalize a signature S_i of an interaction i as a function $S_i : \mathcal{P}(E) \rightarrow [-1; 1]$, where $\mathcal{P}(E)$ denotes the set of possible interactional contexts, that gives a numerical value in $[-1, 1]$ that reflects the possibility of successfully enacting i in an interactional context E . $S_i(E) = 1$ means an absolute certainty of success and $S_i(E) = -1$ an absolute certainty of failure. S_i is learned and reinforced when i succeeds or fails to generate accurate predictions. A signature must be reversible: it must be possible to estimate the set of *minimal* interactional contexts (noted \hat{S}_i) that should be present when an interaction is considered as enactable. We use S_i to predict the enaction result of i and \hat{S}_i to extract and exploit information about the object affording i .

Although defining objects by learning to recognize affordances they provide is abundant in literature (e.g. [16], [9]), the *signatures* approach differs by the use of interactions, which

allows implicit relations between interactions, and thus, spatial properties of the environment, to be discovered.

We use the implementation of signatures proposed by Gay *et al.* [3], based on formal neurons. This implementation characterizes a signature S_i with a vector of weights $W_i = [w_{i,1}, \dots, w_{i,n}]$ that is reinforced each time i is enacted as a success or a failure through *delta rule*. This implementation is interesting as the weight vector of a signature allows observing the average context affording an interaction, and thus, the *objects* as defined by the agent.

B. Object instances

A signature S_i of an interaction i characterizes a context, at a certain position relative to the agent, in the form of sets of couples $(j_k, t_k) \in \hat{S}_i \setminus t_k \in [-\Delta t_E, 0]$. To detect such affordances in space, we exploit a property of signatures: each interaction j_k may have its own signature S_{j_k} . It is thus possible to replace elements j_k of S_i that are related to the same primary interaction j with their own signatures S_{j_k} , resulting contexts $\hat{S}_i^{j, \Delta t_j, i}$ that, after enacting j then $\Delta j, i / j + \Delta j, i \rightarrow i$ (with $\Delta t_{j,i}$ duration of $\Delta j, i$), are expected to afford i . The main difference with implementations on RI/PRI agents is that we need to consider delay between interactions. These contexts characterize objects affording i that the agent can reach through enaction of j and $\Delta j, i$. When such a context is present in the interactional context E_t (i.e. $S_i^{j, \Delta t_j, i} > 0$), we consider that an *instance* of the object affording i is present at *position* $(j, \Delta t_{j,i})$, with a certainty of $S_i^{j, \Delta t_j, i}(E_t)$. Note that these contexts do not characterize reachability of the object instance, but only characterize the presence of the instance and the *position* relative to the agent.

This process can be recursively applied, allowing to back-move an interaction through an increasingly long sequence $\sigma = [(j_l, \Delta t_l)]_{l \in \mathbb{N}}$. An instance of the object affording i is then considered as present at *position* σ , at a distance $d \equiv \Delta t_1 + \dots + \Delta t_n$, ($n = \text{Card}(\sigma)$), with a certainty of $S_i^\sigma(E_t)$.

In an implementation based on formal neurons, predecessors are computed as follows: we note $W_{i,(j, \Delta t_j)}$ the subset of weights of signature S_i limited to weights related to interaction j and its associated secondary interactions $\{j_k\}$ at time Δt_j . The set of weights $W_i^{(j, \Delta t_j)}$ of $S_i^{(j, \Delta t_j)}$ is computed using sets W_{j_k} of j and $\{j_k\}$:

$$W_i^{(j, \Delta t_j)} = \sum_{w_k \in W_{i,(j, \Delta t_j)}} \frac{w_k}{\max_{w_k \in W_{i,(j, \Delta t_j)}} (w_k)} \times W_{j_k} \quad (1)$$

C. Spatial structure

This structure stores and tracks detected object instances, even when they escape from the agent's sensory system. In a previous work[3], we proposed a structure that can characterize the surrounding context of the agent without preconception and without using the notion of space. This work showed that only two types of information are required to localize an object instance affording an interaction i_a : the interaction i_m that allows to move closest to this instance, and an estimation of its distance as the number of interactions needed to reach

this instance. The spatial structure stores and tracks object instances under the form of triplets $(i_a, i_m, d) \in I \times I \times \mathbb{N}^+$. In CRI model, object instance positions are characterized with triplets $(i_a, (i_m, \Delta t), d) \in I \times (I \times \mathbb{R}^+) \times \mathbb{R}^+$, with $(i_m, \Delta t)$ the first element of a sequence σ . However, these changes do not require modifications in the mechanisms of the spatial structure.

D. Decisional systems

We propose three decisional mechanisms to define the next variations to enact. The first mechanism is dedicated to sensorimotor learning: at time t , if there are variations di that were not tested before in the context of e_t , the mechanism selects one of them, allowing to construct the set of possible interactions.

The *exploration mechanism* tests and improves signatures of interactions. This mechanism considers variations $\Delta i_{e_t, i_t + \Delta t}$ considered as enactable, i.e. $\forall \Delta t_i \in [0, \Delta t[, S_{i_t + \Delta t_i}(E_{t + \Delta t_i}) > 0$ where $E_{t + \Delta t_i} = \{(i_t, t_t)\} / (i_t, t_t + \Delta t_i) \in E_t$. The maximum length of variations is thus limited by Δt_E . Using possible variations, the mechanism defines a *learning utility* value lu_{di} to each known variations di_t :

$$lu_{di_t} = \max_{\sigma/di_1, \sigma=di} \frac{1}{|S_{i_t + \Delta t}(E_{t + \Delta t})| \times (nb_{e_t, di} + 1)} \quad (2)$$

where $nb_{e_t, di}$ counts the number of enactations of variation di from e_t ensuring that learning utilities values will decrease through time. Then, if $\max_{di}(lu_{di}) > \tau_{learn}$, the agent selects and tries to enact the variation di that has the maximum lu_{di} . We note $\tau_{learn} \in]0, 1[$ the *learning threshold* that defines the minimum reliability value to consider an interaction as reliable. Otherwise, the agent uses the *exploitation mechanism*. Thus, the more accurate interactions signatures become, the less used the exploration mechanism is.

The *exploitation mechanism* helps generating behaviors that satisfy the agents motivational principles at short and medium terms. The principle is that this mechanism adds a positive utility value to interactions that enable moving closer to object instances affording interactions with high valence ν , and a negative value when the object instances afford interactions with negative valences. The utility value is weighted by the distance of object instances so that far object instances have a lower influence. The exploitation mechanism considers *candidate variations*. A variation di_t is candidate for enaction if $i/e_t + di_t \rightarrow i$ is predicted as a success. A variation di can bring an object instance characterized by triplet $(i_a, (i_m, \Delta t)_k, d_k)$, stored in spatial structure M , *closer* by dt if there exists a variation $\Delta i_{e_t, i_m}$ considered as enactable and that begins with di . The mechanism defines an *exploitation utility* value eu_{di} to each *candidate* variations (3) and proposes a candidate with the greatest exploitation utility.

$$eu_{di_t} = \nu_i + \sum_{(i_a, (i_m, \Delta t)_k, d_k) \in M_t / \exists \Delta i_{e_t, i_m} \text{ enactable} \wedge di_t \in \Delta i_{e_t, i_m}} \nu_{i_a} \times f(d_k) \quad (3)$$

where $f: \mathbb{R}^+ \rightarrow]0, 1]$ is a function that characterizes object influence according to their relative distance. After tests [3], we chose the function $f: x \rightarrow \beta \cdot e^{-\gamma \cdot x}$ (where $\beta = 0.01$ and $\gamma = 0.05$ are *influence* coefficients).

IV. IMPLEMENTATION

We tested our mechanisms on an artificial agent moving in a 2-dimensional continuous and static environment. The test environment is intentionally simple to make it easier to analyze. Indeed, with a 3D environment, or with more than three colors, it would not have been possible to display and analyze signatures as RGB images. However, the complexity of the environment does not affect spatial memory principles: an agent with 3D interactional possibilities would construct a structure reflecting a 3D environment. The sensorimotor possibilities of the agent define a list of five actions and perceptions, listed below:

- ▷ forward impulsion, - ● bump in a solid object,
- ◁ left rotation impulsion, - ● eat something edible.
- ◃ right rotation impulsion,

Bump and eat perceptions can be observed while the agent gives an impulsion. The agent can intend 12 different di : wait ($di = \emptyset$), forward impulsion, left impulsion, right impulsion, bump, eat, forward and bump, forward and eat, left and bump, left and eat, right and bump, right and eat. The *simulation step* of the environment defines the shortest period dt that the agent can experience. A forward impulsion sets the linear speed of the agent to 0.2 grid cell per step. Speed decreases at each simulation step to 0, simulating fluid friction. After a forward impulsion, the agent moves for nearly its length (1 grid cell) in 10 simulation steps. A rotation impulsion sets the angular speed of the agent to 18° per step, then decreases at each step to 0. After giving an impulsion, the agent rotates nearly 90° (left or right) in 10 simulation steps. Bump is enacted when the agent bumps in a solid element. Eat is enacted when the agent moves over an edible element.

The timeline Φ and the length of interactional context E_t are limited to 10 simulation steps. To make easier the results analysis, we only defined interactions composed of at most two non-empty variations, and limited the number of non-empty variations on the timeline to three. When three non-empty variations are already present on the timeline, the agent can only *wait*. An interaction is thus under the form $i = (\{\epsilon_k, t_0\}_k)$ or $i = (\{\epsilon_k, t_0\}_k, \{\epsilon_{k'}, t_1\}_{k'}, t_0, t_1 \in [-10, 0]$. Experiments use valences defined as follows: if $t_0 = 0$, then valence of i is 50 when $\{\epsilon_k\}$ contains *eat*, -10 when it contains *bump*, 2 when it contains *forward* and -2 when it contains *turn*. Otherwise, valence is 0. We thus arbitrarily decided that the *newborn* agent strongly *likes* eating, *dislike* bumping and slightly *likes* moving forward. We added a set of secondary stimuli provided by the agent visual system, that can detect colors among red, green, blue, and measures distances. Visual interactions consist in seeing a colored element while enacting a primary interaction, at a certain, but unknown, position of space. We discretize the visual field as a regular grid of 30×15 positions. These interactions have a predefined valence of 0. We thus define $30 \times 15 \times 3 = 1350$ possible secondary interactions per primary interaction.

We also simplified the interactional context, using the simplification proposed in Section III-A to gather secondary interactions as the same inputs of the interactional context.

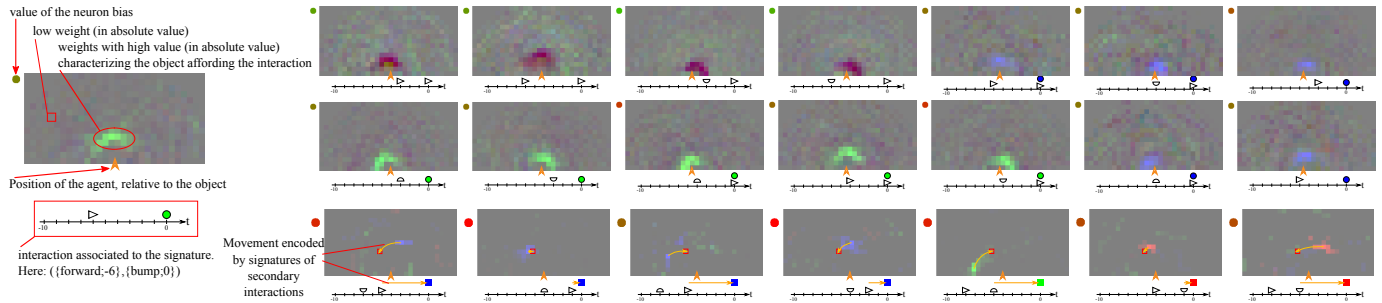


Fig. 2. Sample of signatures obtained after 400,000 simulation steps. Weights of signatures are normalized according to the greatest weight, and organized topographically to match position of the corresponding secondary interaction in space. For each position, the three weights related to *seeing* one of the three colors are displayed using the three color channels to define a RGB image, allowing to observe the color of *objects* considered by the agent. The little disc indicates the neuron bias, green means a normalized value of 1, red a normalized value of -1. Interactions associated with signatures are displayed below signatures. This representation allows an external observer to observe an average context required to successfully enact the displayed interaction. We observe that interactions containing *move forward* are afforded by the absence of green and blue element in front of the agent (mid-red blobs), those containing *bump* (green disc at $t = 0$) are related to a green element, and those containing *eat* event (blue disc) are related to a blue element. While the size of the object nearly correspond to the size of the agent, the exact position of the object depends on elements that compose interactions. The third line shows signatures of some secondary interactions consisting of seeing green, blue or red element at position designated by as a red square. We observe that these interactions are related to the presence of an element of the same color at a position that characterizes the movement produced by primary interaction since the last event (orange arrow): some properties of the environment, such as inertia, are thus encoded in these signatures.

This simplified interactional context can contain at most $30 \times 15 + 1 = 451$ different elements. We implemented a hard-coded spatial structure similar to the structure proposed by Gay and Hassas [2], that implements properties observed in agnostic spatial structure [3].

The environment contains three types of elements characterized by a color that makes them recognizable with visual interactions: walls (green) affording bump, preys (blue) affording eat, and algae (red) that are walkthroughable. These three elements are opaque: the agent cannot detect an element that is behind another one.

V. EXPERIMENTS AND RESULTS

We report on two experiments. The first focuses on the sensorimotor learning and exploration decisional systems and the analysis of properties emerging from signatures of interactions. The second tests the exploitation decisional mechanism and analyses emergent behaviors of the agent.

A. Emergence of signatures

We first let the agent move in its environment, driven by the sensorimotor learning and exploration mechanisms. In the first 10,000 simulation steps, the agent discovers most possible interactions, although new interactions are regularly discovered. Then, the agent's behavior becomes more and more driven by the exploration mechanism. Figure 2 shows examples of signatures obtained after 400,000 simulation steps, displayed in an observer-friendly way. We observe that interactions containing *move forward*, *bump* and *eat* elements are related to elements that are in front of the agent, respectively the absence of wall and prey (dark red blobs), the presence of a wall (green blobs) and the presence of a prey (blue blobs). However, the position of the blobs depends on elements that compose the interaction: past elements change the movement of the agent when the last event is enacted, which changes the relative position of the object that affords the interaction. These

changes in position indicate that properties of the environment, such as inertia of the agent, are encoded in signatures, and can help defining when to enact a variation.

We observe, on signatures of secondary interactions, that the result of seeing an element of color c at a position p while enacting a primary interaction i is related to seeing an element of the same color c at a position p' that characterizes the movement produced by enaction of i . Thus, movement produced by a primary interaction, and thus spatial properties

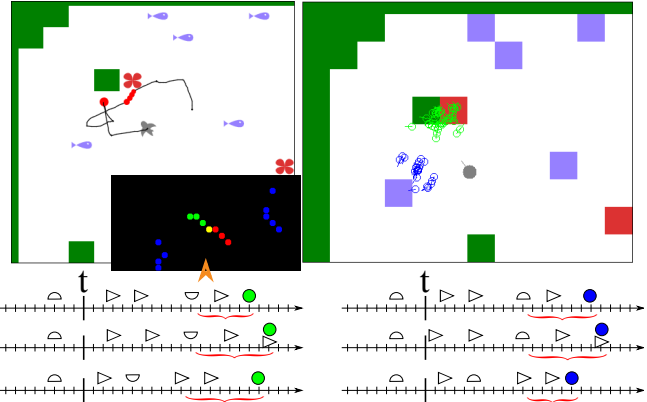


Fig. 3. Detection of object instances. Top left: the agent (gray shark) in its environment. Walls are represented as green blocks, preys as blue fishes and algae as red flowers. A black line shows the agent's path over the last 200 simulation steps. Here, the agent is turning left (left impulsion three simulation steps ago). The inset picture shows enacted visual interactions on a topographic and colored representation that makes them recognizable by an external observer. Top right: Detected instances of objects affording *bump* interactions and of objects affording *eat*. As external observers, we represent object instances with circles and lines (green for bump and blue for eat) indicating positions and orientations characterized by sequences σ of instances. Bottom: examples of sequences characterizing object instances (afforded interactions are indicated in red). Sequences of instances affording *bump* begin with *move forward* and *turn right*, while sequences related to *eat* begin with *move forward* and *turn left*, characterizing that there is a solid object in front-right side of the agent and an edible object in front-left side.

of space, can be characterized by signatures of secondary interactions. We used these signatures to detect object instances around the agent. Figure 3 shows an example of environmental configuration. We can observe positions relative to the agent where objects are detected.

B. Navigation behavior emergence

This experiment tests the exploitation mechanism, using the hard-coded memory described in Section IV. The agent moves in its environment, driven by the three decisional mechanisms. Once most usual interactions are discovered, the exploration mechanism begins to influence the behavior of the agent. Then, through time, signatures become accurate, and the exploitation mechanism is more often used: at simulation step 100,000, the exploitation mechanism provides nearly 30% decisions, and 56% at step 300,000. Once signatures of interactions containing *bump* and *eat* becomes accurate enough, the agent begins to move toward preys and to avoid walls.

We then deactivate sensorimotor learning and exploration mechanisms to study exploitation mechanism and how the agent *interprets* elements of its environment. We observe that the agent is strongly attracted by preys, as they afford interactions with a high valences: the agent moves from a prey to another one. We use this property to observe how the agent interprets elements of the environment, by hiding a prey with an other element and observing how the agent behaves in front of the new element. Figure 4.b) shows the path used to catch a prey. We first add algae, then walls. We observe that with algae, the agent behaves like if they were not present: indeed, algae have the same interactional properties as empty space, and became *interactionally invisible*. With walls, the agent uses another path avoiding the wall. As walls afford interactions with negative valence, they are considered as repulsive. This new path illustrates how the space memory works: as the agent comes closer to the wall, any interaction containing move forward gets a strong negative utility value. When the utility value becomes strong enough to overcome *turn* interactions, the agent rotates. After bypassing the wall, the agent moves again towards the prey.

VI. CONCLUSION

We proposed new RI models and mechanisms to enable an artificial agent to continually interact with its environment,

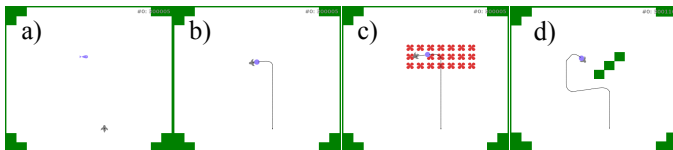


Fig. 4. a) initial configuration for each run. b) the agent is strongly attracted by the prey, as it affords interactions with high valences. c) we hide the prey with algae. The agent uses a similar path than without algae: the agent learned that algae have the same interactional properties as empty space. The path is not exactly the same because the agent uses the noisy position given by the spatial memory. d) we add a wall between the agent and the prey. The agent bypasses the wall while keeping its distance with the wall. Walls thus becomes repulsive as they afford interactions with negative valences.

integrating properties of its environment and its body such as inertia. We presented a formalization and an implementation of a continual RI model that considers continual agent-environment interaction process rather than a discrete cycle action-result, enabling implementation of RI agents in robots. It appeared that these changes do not impact importantly the spatial structure as it was proposed in Gay *et al.* [3], indicating that this structure can be directly used in a CRI implementation.

This work sheds some lights on how a physical agent can interact with its environment, learn properties of its environment and behave according to intrinsic motivational principles, without any preconceptions about its environment or its sensorimotor possibilities. In future works, we intend to implement our mechanisms in more complex systems, and in particular agents using a continuous set I of interactions.

REFERENCES

- [1] J. Baleia, P. Santana, and J. Barata, *On exploiting haptic cues for self-supervised learning of depth-based robot navigation affordances*, Journal of Intelligent & Robotic Systems, 80(3):455–474, 2015.
- [2] S. L. Gay and S. Hassas, *Autonomous object modeling based on affordances in a dynamic environment*, In BICA 2015, volume 71, pages 150–156, 2015.
- [3] S. L. Gay, A. Mille, O. L. Georgeon, and A. Dutech, *Autonomous construction and exploitation of a spatial memory by a self-motivated agent*, Cognitive Systems Research, 41:1–35, 2017.
- [4] O. L. Georgeon and D. W. Aha, *The radical interactionism conceptual commitment*, Journal of Artificial General Intelligence, 4(2):31–36, 2013.
- [5] O. L. Georgeon and I. Sakellariou, *Designing environment-agnostic agents*, In Adaptive Learning Agents workshop, at AAMAS2012, 11th International Conference on Autonomous Agents and Multiagent Systems, pages 25–32, 2012.
- [6] J. J. Gibson, *The theory of affordances*, In Robert Shaw and John Bransford, editors, *Perceiving, Acting, and Knowing: Toward and Ecological Psychology*, pages 62–82. Erlbaum, Hillsdale, NJ, 1977.
- [7] S. Harnad, *The symbol grounding problem*, Physica D: Nonlinear Phenomena, 42:335–346, 1990.
- [8] M. Hoffmann, H. Marques, A. Arieta, H. Sumioka, M. Lungarella, and R. Pfeifer, *Body schema in robotics: A review*, IEEE Transactions on Autonomous Mental Development, 2(4):304–324, 2010.
- [9] H. Min, C. Yi, R. Luo, J. Zhu, and S. Bi, *Affordance research in developmental robotics: A survey*, in IEEE Transactions on Cognitive and Developmental Systems, 8(4):237–255, 2016.
- [10] P.-Y. Oudeyer, F. Kaplan, and V. V. Hafner, *Intrinsic motivation systems for autonomous mental development*, IEEE Transactions on Evolutionary Computation, 11(2):265–286, 2007.
- [11] J. Pisokas, U. Nehmzow, *Experiments in Subsymbolic Action Planning with Mobile Robots*, In: Kudenko D., Kazakov D., Alonso E. (eds) Adaptive Agents and Multi-Agent Systems II. Lecture Notes in Computer Science, vol 3394. Springer, Berlin, Heidelberg, 2005.
- [12] R. Pfeifer and C. Scheier, *Sensory - motor coordination: The metaphor and beyond*, Robotics and Autonomous Systems, 20(2-4):157–178, 1997.
- [13] J. Piaget, *The construction of reality in the child*, New York: Basic Books, 1954.
- [14] F. H. Previc, *The neuropsychology of 3-D space*, Psychological Bulletin, 124(2):123–164, 1998.
- [15] E. Ugur, E. Sahin, and E. Oztop *Unsupervised Learning of Object Affordances for Planning in a Mobile Manipulation Platform*, IEEE International Conference on Robotics and Automation (ICRA), 2011.
- [16] E. Ugur, M. R. Dogar, M. Cakmak, and E. Sahin, *The learning and use of traversability affordance using range images on a mobile robot*, In IEEE International Conference on Robotics and Automation, 2007, pages 1721–1726, 2007.
- [17] D. Vernon, G. Metta, and G. Sandini, *A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents*, Trans. Evol. Comp, 11(2):151–180, 2007.