Interacting with Other Agents Without a-priori Knowledge: a Radical Interactionist Architecture Developing From the Ground-Up the Ability to Infer Motivations and Predict Behaviour

Simon L. Gay[1], François Suro[1] and Jean-Paul Jamont[1]

Abstract— **In the field of developmental agent learning, Radical Interactionism introduces a novel formulation of the problem of self-discovery of sensorimotor possibilities, environmental properties and the development of behaviours without external rewards.**

Agent architectures following this approach have proven capable of inferring the properties of space, building models of local environments and developing behaviours satisfying inborn motivational drives, without *a priori* knowledge. However, new challenges are expected when extending this approach to multi-agent systems, where behaviours involve interactions with other agents driven by their own, unknown, decision systems.

In this article, we propose an architecture based on interactionist principles, capable of inferring the motives of other agents whose motivational systems are not known *a priori*. The interactionist agent thus learns to predict the behaviour of other agents, improving its chances of interacting with them. We illustrate our proposal with an example of a working system implemented on an interactionist agent in a prey-predator multi-agent context.

Index Terms—developmental learning, interactionism, affordance, autonomous mental development, spatial awareness.

I. Introduction

Developmental learning [1] (DevL [2]) encompasses a set of bio-inspired approaches drawing on discoveries in fields ranging from neuroscience to psychology to propose new ways of designing autonomous artificial agents. It is fundamental research on the principles of learning, from animal to human-like intelligence. Unlike other approaches seeking optimal policies to accomplish specific tasks (e.g. Reinforcement Learning [3], Deep RL [4], MARL [5]) , a particular focus of developmental learning is the development of agents capable of interacting with their environment without any *a priori* knowledge about it or their own abilities.

According to the developmental approach, an agent constructs an emergent model of its sensorimotor systems and environment by interacting with it, to enable the generation of behaviours driven by intrinsic motivation [6], thus remaining independent from any external cause or supervision.

The interactionist approach [7] assumes that perception is the result of an agent's action, its experience providing a greater information content than passive sensory perception [8]. Optical flow or tactile feedback are examples of this phenomenon. The agent is therefore proactive in the perception process, which becomes an internal construct resulting from the active experience. This approach is in line with constructivism [9] and enactive learning [10], which assume that learning about the environment and the emergence of behaviours are built on the basis of *sensory-motor schemes* [11]. These schemes, which we will call *interactions*, comprise the atomic elements with which the agent must discover its environment, then build its model and finally develop behaviours of increasing complexity.

As the complexity of behaviours increases, the agent will soon face problems involving other agents. In such multi-agent perspectives, the interactionist approach must be extended to the development of collective, adversarial or collaborative behaviours. This requires the ability to anticipate the behaviour of other agents, always without any *a priori* knowledge about them or their motivations.

In this paper, we propose an architecture based on interactionist principles to infer the behaviour of agents whose motivational systems are not known *a priori*, on the assumption that their movements are motivated by the presence of entities in their local environment. The paper is organized as follows: Section II provides a state of the art of developmental and interactionist approaches, and describes the basic principles of our architecture. Section III describes the proposed architecture and Section IV shows a working system demonstrating this architecture. Finally, Section V summarizes this research's main contributions and discusses on future development perspectives.

II. Background and Radical Interactionism framework

One of the first challenges in developing an agent with no *a priori* knowledge is building a model of its sensorimotor system (or Body Schema). Sensorimotor learning models abound in the literature [12], including intrinsic motivation models such as artificial curiosity [6]. These models can reach high-level sensorimotor patterns, such as walking gaits, they are however not ideally suited for longer-term behaviors.

The Radical Interactionism (RI) model [7], illustrated in Fig. 1 considers that perception is the feedback generated by actions, and describes sensorimotor patterns as couples (action, result) called *interactions*. Hierarchical sequences of interactions can be formed to generate behaviours of increasing complexity. This model introduces the principle of interactional motivation [13], which guides both the development and subsequent behaviour of the RI agent according

*This work is supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02).

[1] LCIS, Université Grenoble Alpes, Valence, France
simon.gay/francois.suro/Jean-Paul.Jamont @lcis.grenoble-inp.fr

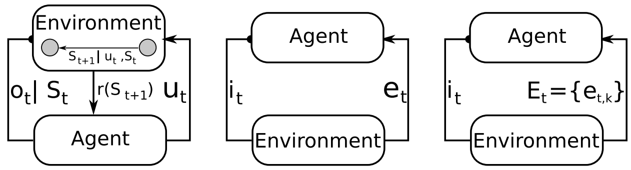

Fig. 1. Comparison of Reinforcement Learning (left) and Radical Interactionism (middle and right). In RL, an observation o_t is analysed by the agent to select an action u_t. The resulting change of state generates a reward r_{t+1} which the agent seeks to maximize. In RI, the cycle begins with an intention of the agent, who enacts a sensorimotor scheme (interaction) i_t, and receives as feedback the sensorimotor scheme e_t that was actually experienced. the agent has no direct access to the environment. The learning model focuses on the relationships between interactions and tries to predict their results. Parallel RI (right) states that the feedback experienced from enacting an interaction is a set of interactions E_t.

to in-born numerical values associated with each interaction, called *valences*. These valences are used to model the agent's *drives* (e.g. eating food 'feels good') and the environment's constraints (e.g. a collision is 'painful') without relying on an external interpreter, states, or predefined goals to evaluate the agent's behaviour. Furthermore, valences steer decisions *a priori* as an operating drive, but are not taken into account in learning as parameters to be optimized, thereby dissociating the generated sensorimotor model from the decision-making model. Thus, the learning model of a RI agent focuses more on the reliability of the constructed model and predictions than on efficiency in solving a predefined task.

Parallel Radical Interactionism (PRI) [14] extends RI to handle complex environments. In PRI, intending an interaction may result in experiencing multiple interactions simultaneously. This set of experienced interactions (E_t) has a causal link with the intended interaction. For instance, the optical flow measured by a point on the retina provides in itself little information, but, combined with the agent's movement, can inform it of the relative distance of an entity.

In a realistic environment, sensorimotor learning is hindered by the number of possible actions and outcomes. An approach to this issue is to consider that the result of a sensorimotor pattern is linked to the presence of phenomena in the environment, akin to an affordance [15], defined here as a possibility of interaction offered by the environment to the agent [16], whose presence or absence affects the result of its action. The use of affordances as a means of representing the environment has been extensively studied [17]. Some approaches sense the presence of affordances to determine the crossability of an environment [18], or locate positions from which a robotic arm can grasp an object [19]. However, due to their strong coupling with the sensory system, these models are unable to make longer-term predictions without additional spatial information [20].

In a RI model, an affordance can be defined as the implicit particular configuration of physical elements and/or properties in an egocentric frame of reference whose presence enables its successful enaction. Detecting the affordance of an interaction based on feedback experienced from previously enacted interactions, which are also afforded by their affordances, allows for a recursive process that enables the detection of distant affordances and the emergence of longer-term or compound behaviours [14].

As an agent improves its understanding of the environment, it will find itself confronted with particular affordance phenomena whose prediction error cannot be reduced. Their affordances are linked to mobile entities driven by an unknown motivational system: they are other agents.

In the field of multi-agent systems, abundant research has been devoted to modelling interactions between agents [21]. These models are given by a designer and can be instantiated to fit a particular problem.

Autonomous development requires the agent to learn to infer the intention of another agent without any *a priori* knowledge of their decision systems, based solely on knowledge acquired during sensorimotor learning. Assuming the other agent follows a simple reactive decision model, the prediction of its behaviour comes down to determining which affordances are attractive and which are repulsive. For a RI agent, it amounts to building a model of another agent using the same RI principles (regardless of its internal workings). The RI agent has to locate affordances relevant to the other agent and establish the context from the other agent's point of view. By observing its behaviour, the RI agent gradually estimates how it values the various affordances and eventually predicts its most likely move.

The following section introduces an architecture extending the PRI model to enable an agent to detect, integrate and predict the movement of other agents in its environment.

III. An interactionist learning architecture for predicting behaviour in a multi-agent context

In our architecture, the agent's model initially contains a set of sensorimotor schemas, or interactions, and builds structures from these interactions to characterize the agent's environment and generate behaviour that satisfies its motivational principles. The architecture, shown in Fig. 2 can be divided into 3 main modules:

1) The **decision module** which uses data from the other two modules to select the next interaction to enact, either to improve the agent's model or to generate behaviours satisfying its interactional motivation.
2) The **entity module** learns to define the entities populating the environment on the basis of the interactions they afford (therefore, entities will be called "affordances"), and to predict the results of interactions.
3) The **space module** generates an internal representation of the local environment in ego-centric space. It exploits the properties discovered by the entity module to detect remote affordance instances. This module also manages a structure that maintains the position of affordance instances as the agent moves, even when the agent is no longer able to detect them. Finally, by leveraging this knowledge, the module learns to predict the behaviour of moving entities.

Before going into the details of each of these modules, we will briefly review the notions, notations and intuition of the interactionist concepts involved. Sections 3.1 to 3.4

summarize mechanisms previously described in [14], [22], [23], and Section 3.5 describes the additional mechanisms that finalize the architecture.

A. Interactionist concepts

When a PRI agent [14] intends an interaction i_t at decision cycle t, it experiences the set of interactions E_t which are actually enacted. The enaction of i_t is a success if $i_t \in E_t$ and a failure otherwise. An agent who chooses the interaction $i_t = move$ may experience a movement, the interaction of moving is then considered a success: $i_t \in E_t$. Additionally, it may experience particular variations in its perceptual field (e.g. optical flow, Doppler effect) linked to the interaction i_t to form the other experienced interactions of E_t. The interaction $i_t = move$ may also fail, in which case the agent will not experience movement, $i_t \notin E_t$. In its place, the agent may experience an interaction of $collision$.

Thus, the agent does not have direct access to its environment and 'perceives' it by actively interacting with it. The experienced feedback can be used to predict the outcome of a following intended interaction, which dispenses with notions external to the agent, such as entities and affordances. These terms will henceforth be used to help the reader interpret how the agent works. The notion of affordance to which we will refer is the particular context of elements located to the agent, whose presence makes the successful enaction of its associated interaction possible. The affordances are thus implicitly defined by the agent-environment coupling, following the Stroffegen [24] and Chemero [25] definitions.

The notion of space is initially unknown to the RI agent. However, interactions are related to affordances that can be localized in space, and the enaction of an interaction can be associated with a specific movement (in an egocentric reference frame). Thus, the agent implicitly generates a notion of space through relations between interactions, which differs from a Euclidean geometry, but incorporates the agent's sensorimotor properties. Section III-D shows how this principle is used to generate a model of space.

B. Decision module

The decision module is tasked with planning the agent's behaviour in accordance with its own motivational principles, based on interactional motivation (Sec. II). The agent's behaviour is oriented by assigning a numerical valence ν_i to each interaction i, defining the immediate utility of successfully enacting i.

When no interactions with high valences are immediately possible, the exploitation mechanism tries to lead the agent towards distant affordance instances that afford interactions with a 'positive' valence (and away from 'negative' affordances) [14]. To that end, the *spatial module* provides for each affordance instance a tuple (a, i, d) denoting the afforded interaction a, an interaction i that brings it closer, and the distance d estimating its proximity (i.e. the minimum number of interactions to enact to reach it). An interaction adds to its valence an utility value depending on affordance instances that it can lead to (Eq. 1):

$$u_i \;=\; \nu_i \;+\; \beta \times \sum_{(a_k,j_k,d_k)\in M_t/j_k=i} \nu_{a_k} \times f(d_k) \quad (1)$$

where v_{a_k} is the valence of the interaction a_k, (a_k, j_k, d_k) an instance of the affordance of a_k stored in memory M_t which can be brought closer by enacting the interaction i (i.e. $i = j_k$). f is a strictly decreasing and positive function characterizing the importance of an affordance according to its distance in the agent's decisions-making process, and β is a coefficient for delayed satisfaction [14]. We can note that a change in valences (e.g. the value of 'eat' decreases when satisfied) will instantly modify the attractiveness of surrounding affordances.

In the case of a simple reactive behaviour, the operating mechanism selects the interaction with the highest utility from among those considered enactable in the current context. More complex decision mechanisms accounting for the agent's ability to modify its environment were presented in a previous work [26].

The possibility of enacting an interaction i according to the agent's current feedback E_t is provided by the *Entity module* (Sec. III-C) as the confidence in the possibility or impossibility of enacting i. When this confidence is low (e.g. initial learning or changes in the environment's properties), a curiosity mechanism is triggered, to select and test an interaction to reinforce the learned structures. As the agent comes to characterize its environment, confidences rise and curiosity gradually gives way to exploitation (although curiosity can be used in the case of changes in the environment).

C. Entity module

This module determines which interactions can be enacted given a feedback E_t. Such sets, which characterize the properties of an affordance from a sensorimotor point of view, are called *signatures*. Formally, a signature is defined as a function S_i that gives the agent's *confidence* in the prediction of the outcome, success or failure, of i's enaction as a function of the feedback E_t (absolute confidence in success if $S_i(E_t) = 1$ and absolute confidence in failure if $S_i(E_t) = -1$). The parameters of a signature S_i are adjusted after each intention of i according to the outcome experienced, to improve the predictions.

The signature S_i identifies one or more sets of interactions $\{j_1, ..., j_n\}$ such that if $\{j_1, ..., j_n\} \subset E_t$, then i can be successfully enacted. These sets thus characterize the affordance of i from the agent's point of view. The signature implementation must define a pseudo-reverse function $\hat{S}_i$ providing sets $\{j_1, ...j_n\}$ characterizing the presence of the affordance of i for $\hat{S}_i(1)$ or its absence for $\hat{S}_i(-1)$. Signatures integrating static or predictable affordances were implemented as formal neurons [14] and lists of interactions (LUT) [26].

Affordances offered by mobile entities are susceptible to move during the enaction of an interaction. Thus even if the affordance is present, the interaction may still fail. Conversely, the mobile entity may afford the interaction from multiple positions k depending on its movements.

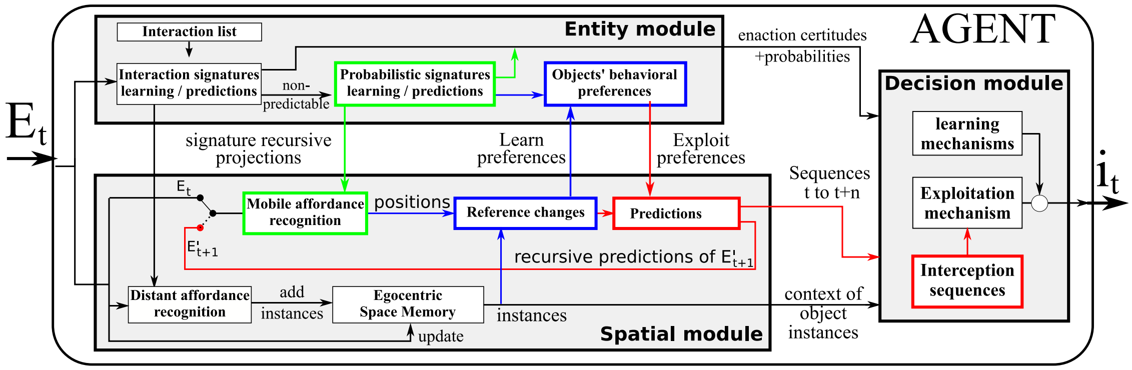

Fig. 2. Architecture for inference and prediction of mobile entities in a Radical Interactionism model (mechanisms of the static affordance architecture [14] are in black). In red, the prediction and decision mechanisms presented in this paper, finalize the complete architecture, enabling efficient behaviours for interacting with mobile entities. Green and blue mechanisms were studied respectively in [22] and [23]. The *decision module* (Section III-B) exploits the environment model generated by the two other modules to generate behaviours satisfying the interactional motivation principles, or to test and reinforce data structures. The *entity module* (Section III-C) contains the initial set of sensorimotor schemes (interactions) and learns to predict the results of interactions by constructing data structures, called signatures, that characterize the affordances of interactions. The *spatial module* (Section III-D) exploits signatures to detect distant instances of affordances. Detected static affordances are stored in the Egocentric Space Memory that maintains a context of affordances while the agent moves. The module performs reference changes on the affordance context to infer behavioural preferences of mobile affordances. These preferences are then exploited to predict future positions of mobile affordances, and define data structures called interception sequences, that can be directly exploited by the decision mechanism.

We introduced the *partial signature* $C_{i,k}$ to characterize the presence of the affordance of i ($C_{i,k}(E_t) = 1$) or its absence ($C_{i,k}(E_t) = -1$) at one of the positions k. A signature learning process must discover the possible positions of the mobile entity affording i, and construct the signature S_i as a set of partial signatures $C_{i,k}$. The notion of *enaction probability*, independent of the enaction confidence, quantifies for each partial signature $C_{i,k}$, the probability $p_{i,k}$ that the detection of an affordance of i at position k will actually lead to the success of i, i.e. the probability that the mobile entity moves in the direction allowing the enaction of i. Our implementation, described in [22], uses multiple neurons combined by a Winner-Takes-All rule to integrate multiple contexts, with $S_i(E_t) = max_k(C_{i,k}(E_t))$, and estimate the probability of each position k.

D. Space module

This module exploits interaction signatures to identify remote affordances (i.e. requiring prior movement to reach them), and generate a model able to store, locate and track the position of instances of affordances, according to the movement of the agent and possible occlusion (i.e. object persistence [11]).

1) Detecting instances of affordance: This mechanism leverages a property of interaction signatures: a signature S_i defines sets of interactions $\{j_1, ..., j_l, ..., j_n\} \in \hat{S}_i(1)$ characterizing the affordance of i. However, these interactions j_l may have their own signatures. Since some interactions result in transformations in space (i.e. movement of the agent), it is possible to 'project' the required affordance's position into space, in agent-centered reference frame, by means of an interaction j, defining an element designated by S_i^j which, if the agent enacts j, will afford i. This process can be applied recursively through an interaction sequence $\sigma = \langle i_1, ..., i_m \rangle$, thus giving a projected signature S_i^σ. Several sequences may lead to the same projected signatures, in which case only the shortest is retained. When a high confidence (i.e. greater than a threshold) $S_i^\sigma(E_t)$ is established, we consider that an affordance *instance* is discovered at a position characterized by σ. A sequence σ characterises a *position* in space independently of its accessibility (i.e. σ is not necessarily enactable).

The detection of remote instances of mobile affordances adds the complexity of multiple partial signatures $C_{j,k}$, projected independently with each new interaction added to σ. The probabilities of the partial signatures are multiplied consecutively with each new projection. The projection of a signature S_i would thus lead to the generation of numerous projected partial signatures $C_{i,k}^\sigma$ with low probabilities. Our solution [23] is to keep only $C_{i,k}^\sigma$ with the highest enaction probability for each initial partial signature $C_{i,k}$, at each new projection. Thus, to each $C_{i,k}$ of a signature S_i and sequence σ corresponds a single projected partial signature $C_{i,k}^\sigma$. This set of partial projected signature $C_{i,k}^\sigma$ constitutes the projected signature S_i^σ, and designates the positions from where a mobile affordance of i may afford i after enacting σ. Consequently, the same instance of affordance of i will be detected by multiple projected signatures $S_i^{\sigma_k}$ of S_i, each σ_k sequence corresponding to a future position of the affordance instance at time $t+1$.

The Egocentric Space Memory (ESM) sub-module manages a representation of local space with the affordances discovered. Affordance instances are recorded as tuples (a, i, d) used by the decision module, which indicate the afforded interaction a, and its position derived from σ as the elements i and d. i represents the "direction" of the affordance (first interaction of σ), d corresponds to the distance (length of σ). The set of affordances forms the agent's affordance context, denoted M_t. The ESM learns to predict the changes in the position of affordances in the egocentric space when the agent performs an interaction, thus implicitly generating a geometric structure of space based on interactions [14].

2) Inferring the motivations of mobile entities: This module uses the hypothesis that, like the RI agent, all agents are driven by a decision mechanism that reacts to surrounding affordances (eq. 1) according to behavioural preferences that are unknown *a priori*. These preferences can be estimated according to the movement of other agents relative to the affordances (i.e. variations in distance).

The RI agent cannot know the interactions available to another agent, nor their valences. It can however observe changes in the other agent's position and estimate the attractiveness of the other agent to its own affordance.

Formally, the RI agent determines a set of values ν_i^w indicating the valence that the affordance of interaction i (of the RI agent) has for the other agent, itself identified by the RI agent as the affordance of interaction w. Thus, the RI agent requires interactions not only with other agents, but also with any objects likely to affect their behaviour. For instance, a predator must be able to interact with a prey (i.e. eat it), but also have a specific way to interact with the prey's food source allowing it to discriminate this element, and therefore to identify it as an attractor of prey.

The interactionist representation of space does not allow measuring the Euclidian distance between two distant affordances (Sec. III-A). Inferring the behavioural preferences of a target agent requires changing the frame of reference to the point of view of the target to estimate the distances to its surrounding affordances. A trained ESM already possesses all the operations required to implement this process [14]. The change of frame of reference can be achieved by simulating the enaction of a σ sequence characterising the position of the target agent. This simulation in the ESM produces an affordance context M_t^σ locating affordances from this point of view as tuples (a_n, i_n, d_n).

To account for the multiple σ_k sequences representing the instance m of a mobile affordance (i.e. another agent), the distances between the target agent and the surrounding affordance instances are defined as the average of the distances for the different σ_k positions. The average simulation $\overline{M}_t^w$ provides the average affordance context, giving the nature (i.e. a_k) and average distance $\overline{d}_n$ of the affordances from this point of view, as couples $(a_n, \overline{d}_n)$. A utility value can be defined according to the current context $\overline{M}_t^w$ and estimated preferences ν_i^w as Eq. (2):

$$u_t^w = \sum_{(a_n,\overline{d}_n)\in\overline{M}_t^w} \nu_{a_n}^w \ \times \ f(\overline{d}_n) \tag{2}$$

where f is the same decreasing function as in eq. 1, giving a greater importance to the closest affordances.

The variations in distances over consecutive steps reveal which affordances the target agent tends to move towards, or to flee from. Assuming the other agent behaviour is mainly reactive, it will try to maximise the utility of its actions. The variation $u_t^w - u_{t-1}^w$ between two time steps should be positive, otherwise the estimated valences ν_a^w must be corrected for each type of affordance present. A preliminary work implementing this principle [23] shows that the behavioral preferences can be estimated accurately.

E. Tracking and predicting the movement of mobile entities

The inferred model of motivation can be used to predict the most likely future position of the target agent. As it is expected to move toward affordances affording interactions with the greatest valences, we can make the hypothesis that its next position will be the position, among the ones that it can reach at step $t+1$, that has the greatest utility value. Thus, this position can be predicted by computing the utility value of each detected position σ_k of the mobile affordance w (Eq. 3), and selecting the $\sigma_{k,max}$ sequence producing the highest utility value as the next position of the target agent.

$$u_{\sigma_k}^w = \sum_{(a_n,i_n,d_n)\in M_t^{\sigma_k}} \nu_{a_n}^w \ \times \ f(d_n) \tag{3}$$

To predict further steps, this process can be repeated by considering the predicted position $\sigma_{k,max}$ as the current position of the target agent. To that end, this predicted position must be integrated in a simulated feedback E_t' that the RI agent would have experienced if the target agent were in this position, i.e. the set of interactions $\{j_1, ..., j_n\}$ that would be experienced as a consequence of the presence of the target agent at $\sigma_{k,max}$. The context designated by the partial signature $C_{w,k}^{\sigma_{k,max}}$ associated with $\sigma_{k,max}$, with the greatest *enaction probability*, is used.

The simulated feedback E_t' is used to repeat the prediction process recursively and derive a probable trajectory for the target agent. The recursion may be halted when the predicted position may be reached by the RI agent in a number of steps equal to the number prediction steps (i.e. an *interception point* is found), when the ESM does not allow the next step to be simulated (e.g. an affordance is no longer located with sufficient precision, see [14]), or by a technical limit, such as a maximum length of sequences.

When an interception trajectory is thus defined, its sequence is considered as the actual position of the moving entity. However, such a "position" can only be maintained while the agent strictly perform the interception sequence, the position being lost if the agent must change its trajectory, for example, to avoid an obstacle. Thus, we propose to maintain a set of interception sequences of different lengths to allow the agent to keep track of the position even if it should deviate from the shortest path, offering the possibility of taking into account other elements of the environment.

The interception sequences cannot be managed by the ESM: as the 'position' of the mobile entity depends on the interception sequence (especially its length), it cannot be updated like a static affordance. To track a mobile entity without resorting to the ESM, the sequential nature of such a 'position' is exploited. It is possible to update the position to the next instant by removing the first interaction from the sequences which begin with the enacted interaction, and discarding the others as no longer valid. An interception sequence $\sigma = \langle i, \sigma' \rangle$ is updated to σ' when the agent enacts i. Thus, the mobile entity is accounted for in the decision process, even if it or its predicted interception point leave the perceptual field of the agent.

The interception sequences are provided to the *decision module* (Sec. III-B) as (a, i, d) tuples, using the first element and the length of a σ interception sequence as the i and d parameters, respectively.

IV. Demonstration on a working system

To illustrate our proposal, we have implemented our architecture in a prey-predator context, where a predator must infer prey behaviour (i.e. attraction to its food source) to predict its movement and intercept it efficiently.

The predator is endowed with the following interactions (valences in parentheses): move forward ▷ (2), collide with an obstacle ▶ (-5), eat ▶ (50), turn right ⌓ (-3) and left ⌒ (-3) by 90°. It can also experience interactions associated with its visual system as an outcome of other interactions. These interactions occur in a field of view of 180° in front of the agent. As mentioned in Section III-D.2, the predator must have an interaction with the prey's food source to infer attraction behaviour, even if this interaction has no purpose in itself. This is modelled by a particular sensation of contact (different that of empty space), when moving over a prey's food source, defining the *slide* interaction ▶ (0). The visual system provides a set of additional outcomes produced when enacting a primary interaction (except *bump* that does not produce movement). This visual system discretizes space in front of the agent as a regular grid of 15x9 positions and perceive 3 colors, for a total of 15x9x3x5=2025 uninterpreted 'visual' interactions.

This experiment only assesses the prediction module, and uses signatures acquired in previous experiments. For more details on how to generate signatures, see [23]. To exclude possible interference with the evaluation of the prediction module, the ESM (Sec.III-D) is substituted for a static program updating affordance contexts M_t from detected instances of affordances. Since the prediction module relies on discrete abstractions of the environment generated by other modules, the environment chosen for the experiment is also discrete to simplify the interpretation of the results. The localisation of affordances in a continuous environment is covered in [14]. A previous work [23] demonstrated the ability of the agent to infer valences for prey with an accuracy close to the ground truth. For this demonstration, the valences of *bump* (-0.2) and *slide* (2) for the preys are provided to the agent.

Fig. 3 shows and compares the generation of behaviour, with and without the interception mechanism. After enacting an initial interaction, the ESM registers three static affordances, corresponding to the interaction *bump* (wall blocks) on the left and right, and the interaction *slide* (food source for the prey, i.e. the seaweed) on the top right. The content of the ESM is shown in the top block at step 0: tuples of the form (a, i, d). When the prediction mechanism is disabled (Fig. 3a), the agent locates the mobile affordance of eat (fish) at time t using the most likely sequences leading to its current position (bottom block at step 0). The agent attempts to reach

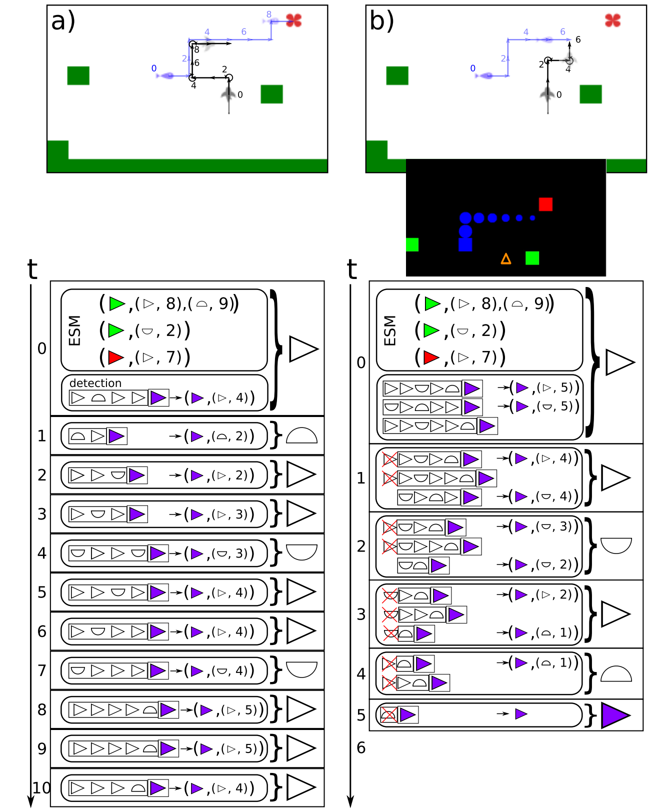

Fig. 3. Generation of an behaviour to capture a prey. a) The prediction mechanism is disabled. Top: the agent (grey shark) in its environment, and the trajectory over the next ten decision cycles. Bottom: the timeline displays at each step one of the most probable positions (among multiple sequences), and the proposed interaction. At step 0, the timeline details the position of detected static affordances, encoded ad triplets *(a,i,d)*: two affordances of *bump* and one affordance of *slide*. As the fish moves at same speed than the agent, the latter cannot reach the fish until it stops. b) The prediction mechanism is enabled. The middle insert shows the predicted trajectory of the prey at step 0: future positions are represented with a sequence of blue circles of decreasing diameters. The timeline displays at each step a sample of interception sequences and updates, and the interaction proposed by the mechanism. By enacting one of the interception sequence, the agent is able to intercept the prey before it reaches the algae. The set of interception sequences allowing to keep track the future position of the prey even when leaving the agent's perceptual field (steps 3 and 4).

the current position of the prey, and as this affordance is mobile, its position is recalculated at each step, as shown by the fluctuation in the interaction sequences from steps 1 to 10. Since the fish moves at the same speed as the agent, the latter will pursue its prey and will only be able to enact *eat* when the fish reaches the seaweed and stops.

With the prediction mechanism active (Fig. 3b), a set of interception sequences is generated (bottom block at step 0). The agent moves towards the calculated interception point by following one of the possible sequences. As the position of the interception point does not change, we can see that the interaction sequences evolve from one stage to the next by unstacking the enacted interaction. The prediction mechanism maintains several possible interception sequences, which allows the agent to favour one or the other according to the presence, or even the discovery of new affordances. In the present example, the interception sequence keeping the agent distant from the obstacle is preferred, due to the negative valence of *bump* afforded by the obstacle.

At step 2, we notice that the agent turns towards the interception point instead of its current position, which demonstrates the ability to anticipate the future position of the affordance. During this manoeuvre, the prey leaves the agent's perceptual field, and although the interception point could not be recalculated if unforeseen changes occurred, updating the sequence acts as a memory that maintains the affordance in the decision process. Finally, at step 5, the agent finds itself in a configuration where *eat* can be enacted, and fosters this interaction, due to its high valence, at step 6.

V. Conclusion

This paper introduces an architecture based on interactionist principles, with which an agent builds an emergent model of its environment and generates behaviours enabling agent-to-agent interactions in multiagent contexts. To do so, it integrates mobile entities into its model, infers their behavioral preferences, predict their future positions and generate behaviours that maximise the chances of interacting with them. This mechanism requires no *a priori* knowledge of the environment and its entities, and is capable of inferring properties by interacting through sensorimotor patterns.

This architecture presents two limitations inherent to learning exclusively from sensorimotor possibilities. Firstly, the interactions available to the agent must encompass the affordances relevant to the moving entities to detect them. Secondly, predictions are biased by the agent's experience in the environment and by the operation of its own decision system. We plan to study how these biases can affect the emergence of collaborative behaviour in populations composed of agents with different sensorimotor capabilities.

Our next step is to study a high-level decision mechanism exploiting these predictions to generate behaviours of greater complexity, taking into account the use of this mechanisms by other agents. We aim at the emergence of collaborative behaviour in multi-agent contexts, such as the influence of the predator's presence on the prey's behaviour, or the emergence of encircling tactics when hunting large prey. Such applications require the integration of the agent itself in the prediction of the behaviour of other agents, a first step in developing intersubjectivity between agents. We will also study the inference of a hidden affordance through other agents' behaviour.

As mentioned previously, the demonstration shown in this article uses a simple and intuitive environment, in order to make the structures generated easy to interpret and evaluate. However, the mechanisms of the architecture are not strictly limited to Euclidean or even physical spaces. Their general function is the discovery of regularities in transformations produced by interactions in spatially homogeneous and isotropic environments, generating abstract notions from interaction possibilities. Further avenues of research could investigate scenarios involving different topological (non-topographical) spaces, or even abstract spaces, such as dynamic social organizations and communication.

References

[1] P.-Y. Oudeyer, "Developmental robotics," in *Encyclopedia of the Sciences of Learning*. Springer, 2012, pp. 969–972.

[2] A. Ricci, ""go to the children": Rethinking intelligent agent design and programming in a developmental learning perspective," in *In Proc. 21st Int. Conf. AAMAS*, 2022, p. 1809–1813.

[3] A. K. Shakya, G. Pillai, and S. Chakrabarty, "Reinforcement learning algorithms: A brief survey," *Expert Systems with Applications*, vol. 231, p. 120495, 2023.

[4] Q. Li, T. Lin, Q. Yu, H. Du, J. Li, and X. Fu, "Review of deep reinforcement learning and its application in modern renewable power system control," *Energies*, vol. 16, no. 10, 2023.

[5] C. Zhu, M. Dastani, and S. Wang, "A survey of multi-agent deep reinforcement learning with communication," *AAMAS*, vol. 38, no. 1, Jan. 2024.

[6] P.-Y. Oudeyer and F. Kaplan, "What is intrinsic motivation? A typology of computational approaches," *Frontiers in Neurorobotics*, vol. 1, no. 6, 2007.

[7] O. L. Georgeon and D. Aha, "The Radical Interactionism Conceptual Commitment," *Journal of Artificial General Intelligence*, vol. 4, no. 2, pp. 31–36, 2013.

[8] R. Bajcsy, Y. Aloimonos, and J. Tsotsos, "Revisiting active perception," *Autonomous Robots*, vol. 42, 02 2018.

[9] K. R. Thórisson, *A New Constructivist AI: From Manual Methods to Self-Constructive Systems*. Atlantis Press, 2012, pp. 145–171.

[10] T. Froese and T. Ziemke, "Enactive artificial intelligence: Investigating the systemic organization of life and mind," *Artificial Intelligence*, vol. 173, no. 3, pp. 466–500, 2009.

[11] J. Piaget, *The construction of reality in the child.* New York: Basic Books, 1954.

[12] 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*, vol. 2, no. 4, pp. 304–324, 2010.

[13] O. Georgeon, J. Marshall, and S. Gay, "Interactional Motivation in Artificial Systems: Between Extrinsic and Intrinsic Motivation," in *In IEEE ICDL-Epirob*, 2012, pp. 1–2.

[14] S. Gay, A. Mille, O. Georgeon, and A. Dutech, "Autonomous construction and exploitation of a spatial memory by a self-motivated agent," *Cognitive Systems Research*, vol. 41, 08 2016.

[15] J. J. Gibson, "The theory of affordances. perceiving, acting, and knowing: toward an ecological psychology," *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, pp. 67–82, 1977.

[16] E. Sahin, M. Cakmak, M. Dogar, E. Ugur, and G. Üçoluk, "To afford or not to afford: A new formalization of affordances toward affordance-based robot control," *Adaptive Behavior*, vol. 15, pp. 447–472, 2007.

[17] P. Ardón, È. Pairet, K. S. Lohan, S. Ramamoorthy, and R. P. A. Petrick, "Affordances in robotic tasks - A survey," *CoRR*, 2020.

[18] 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*, vol. 80, 2015.

[19] K. Qian, X. Jing, Y. Duan, B. Zhou, F. Fang, J. Xia, and X. Ma, "Grasp pose detection with affordance-based task constraint learning in single-view point clouds," *J. of Intelligent and Robotic Systems*, vol. 100, 2020.

[20] C. Pohl, K. Hitzler, R. Grimm, A. Zea, U. D. Hanebeck, and T. Asfour, "Affordance-based grasping and manipulation in real world applications," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2020, pp. 9569–9576.

[21] Y. Kubera, P. Mathieu, and S. Picault, "Ioda: an interaction-oriented approach for multi-agent based simulations," *Autonomous Agents and Multi-Agent Systems*, vol. 23, pp. 303–343, 2011.

[22] S. Gay, J.-P. Jamont, and O. L. Georgeon, "Identifying and localizing dynamic affordances to improve interactions with other agents," in *IEEE ICDL*, 2022.

[23] S. Gay, F. Suro, O. Georgeon, and J.-P. Jamont, "An affordance-based intersubjectivity mechanism to infer the behaviour of other agents," in *IEEE ICDL*, 2023, pp. 403–408.

[24] T. A. Stoffregen, "Affordances as properties of the animal-environment system," *Ecological Psychology*, vol. 15, no. 2, pp. 115–134, 2003.

[25] A. Chemero, "An outline of a theory of affordances," *Ecological Psychology*, vol. 15, no. 2, pp. 181–195, 2003.

[26] S. L. Gay, A. Mille, and A. Cordier, "Autonomous affordance construction without planning for environment-agnostic agents," in *Joint IEEE ICDL-EpiRob*, 2016, pp. 111–116.