A new paradigm to study social and physical affordances as model-based reinforcement learning

Augustin Chartouny, Keivan Amini, Mehdi Khamassi, Benoît Girard

Cognitive Robotics 4 (2024) 142–155

KeAi
CHINESE ROOTS
GLOBAL IMPACT

Contents lists available at ScienceDirect

Cognitive Robotics

journal homepage: http://www.keaipublishing.com/en/journals/cognitive-robotics/

Research Paper

A new paradigm to study social and physical affordances as model-based reinforcement learning

Augustin Chartouny*, Keivan Amini, Mehdi Khamassi, Benoît Girard

Institut des systèmes intelligents et de robotique, UMR 7222, Sorbonne Université - CNRS, Paris, France

ARTICLE INFO

Keywords:
Social affordances
Model-based reinforcement learning
Navigation
Human-Robot Interaction

ABSTRACT

Social affordances, although key in human-robot interaction processes, have received little attention in robotics. Hence, it remains unclear whether the prevailing mechanisms to exploit and learn affordances in the absence of human interaction can be extended to affordances in social contexts. This study provides a review of the concept of affordance in psychology and robotics and proposes a new view on social affordances in robotics and their differences from physical affordances. We moreover show how the model-based reinforcement learning theory provides a useful framework to study and compare social and physical affordances. To further study their differences, we present a new benchmark task mixing navigation and social interaction, in which a robot has to make a human follow and reach different goal positions in a row. This new task is solved in simulation using a modular architecture and reinforcement learning.

1. Introduction

The concept of affordances introduced by Gibson [1] has been extensively studied in robotics [2,3]. However, most studies on affordances in robotics, defined as action possibilities for robots, focus on object manipulation without social interaction or humans in the close environment [4]. In this article, we study the concept of social affordances, which are action possibilities in social contexts. Social affordances are central to learning in robots: efficient social interaction permits better integration within society and enables the robot to ask for human help when facing a task beyond its ability [5].

The contribution of this article is threefold. First, we refine the concept of social affordances in robotics and suggest studying them in the context of model-based reinforcement learning. Second, we introduce a new task to study physical and social affordances concurrently in a new benchmark combining navigation and human-robot interaction. Finally, we propose a first solution to this new task with the reinforcement learning framework. In the discussion, we highlight the new open questions and possibilities, related to affordances, theory of mind, and open-ended learning, that our task will permit to address in future work.

2. Affordances

2.1. What is an affordance?

The psychologist J.J. Gibson introduced the concept of affordances to refer to what the environment affords to an organism: *"The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill"*[1]. According to this view, animals perceive their environment in terms of action possibilities.

* Corresponding author.
E-mail address: augustin.chartouny@gmail.com (A. Chartouny).

https://doi.org/10.1016/j.cogr.2024.08.001
Received 29 December 2023; Received in revised form 8 August 2024; Accepted 8 August 2024
Available online 14 August 2024

This ecological view of perception clashes with the inference view, in which animals use internal prior information and neural representations to perceive and predict sensory stimuli. According to perceptual inference, a chair is perceived as an ensemble of features (e.g. color, shape, size) that are assembled to create or identify the chair's representation. In ecological perception, these features are perceived through what they afford, i.e. a place to sit, climb, or stand for a human or a surface to land for flies. Moreover, ecological psychology assumes direct perception: affordances are perceived before semantic categorization since low-level features are sufficient to understand most of the possible actions (e.g. flat surfaces afford sit-ability, before categorizing whether the object at hand is a chair, a table, or anything else) [6].

The scope of affordances has been studied through various lenses, from psychology to robotics. In the next sections, we wish to review some work from these two fields.

2.2. Theoretical considerations about affordances

Since affordances are action possibilities that the environment offers to the animal, they could be properties of the environment, of the agent, or emerge from the environment-agent relation. Historical views sometimes describe affordances as dispositions or latent properties of the environment, which would be discovered by the agent through interaction. For instance, Turvey interprets affordances as properties of the environment that are complemented by "effectivities", the propensity of the animals to effect an action. Even if the two dispositions are not defined without each other: *"an affordance is not defined (i.e., is nonexistent) without a complementing animal property"*, affordances are the environmental relata of the agent-environment system only [7].

However, later views on affordances mainly define them as relations emerging from the interaction of the agent and the environment. A fundamental view is the one of Sahin and colleagues who define affordances as acquired relations between two equivalence classes: an effect equivalence class and an (environment entity, agent behavior) equivalence class [2]. With this formalism, a *behavior* (an action) applied to an *entity* of the environment (e.g. an object) should produce an *effect* (a change in the environment or the state of the agent). In addition, they propose that affordances, as relations, can be perceived from three different perspectives. The environment perspective encompasses all the possible affordances applicable to a particular entity of the environment. The agent perspective encompasses all the possibilities of action of an agent interacting with an environment. Finally, the observer perspective is the perception of an external agent witnessing an affordance, without taking part in it. As opposed to historic views of affordances in psychology which sometimes emphasized the role of the environment perspective, they propose that the agent perspective is the most interesting when considering learning processes. The observer view, external to the environment-agent relation provides insights on imitation learning and permits to learn affordances from demonstration.

Indeed, even if some affordances may be innate such as reflex grasp-ability for human infants, affordances are mostly acquired by an organism when interacting with its environment over its lifetime. The most famous example of a learned affordance is the one of the *visual cliff*, introduced by Gibson and Walk [8]. In this experiment, animals cross a flat transparent surface that affords locomotion until halfway, where there seems to be a drop-off of more than one meter. Animals generally stop moving forward when they see the drop-off. Infants that have limited locomotive experience fail to avoid the drop-off in cliff-related tasks, which shows that self-experienced locomotion is necessary to learn and perceive the associated affordance [9]. In addition, infants with limited walking experience fail to avoid the cliff while walking, although they can perceive the cliff while crawling [9]. This shows that what infants learn about the environment while crawling fails to generalize to walking. Hence, infant perception of the environment is altered by how they explore it. After becoming experienced walkers, infants can perceive the cliff and find methods to cross it [10].

The capacity to generalize to new situations, entities, and actions would bootstrap the learning of affordances in humans and robots. In the formalism of Şahin and colleagues, generalization can come either from the entity (environment), the behavior (action), or the affordance. Learning an affordance is then the capacity to learn the invariant properties that specify it [1]. Knowing one affordance, e.g. the cross-ability of a road by walking from one side to another, one could try to further learn this affordance either by trying to cross a slightly flooded road walking (entity equivalence), to cross a road running (behavior equivalence) or to cross a river swimming (affordance equivalence) [2]. This exploration process to further specify and perceive affordances is linked to goal generation: self-generated goals help perceive and learn affordances.

Indeed, affordance perception may change depending on the agent's goal. For example, if an adult wants to replace a broken light bulb, they would perceive a chair as climbable rather than sittable. The intent modifies the perceived affordances. In addition, the chair located right below the light bulb affords climb-ability more certainly than the other chairs: the goal highlighted the climb-ability affordance of the chair. Goal-related affordances match economical views on ecological perception, where only relevant information is picked up from the environment as perceiving all the affordances of all the close items would be costly [1].

Self-generating goals help to learn affordances. When someone sees an affordance from the observer perspective, they might try to reproduce the behavior on the same entity to see if it produces the same effect or try to reproduce the same effect with different behaviors. An infant who sees their sibling cry and get attention from their parents might try to cry as well to get attention. Likewise, an infant who sees their parents climbing the stairs might want to start climbing them. However, learning affordances through intents and self-generated goals is not limited to mimicking human actions. In Montesano et al. [11], robots learning from a human demonstrator try to reproduce the effect of the human action by choosing an action from their model of affordances rather than by copying the human's action, which sometimes produces very different behaviors. This goal-directed method seems central to learning the consequences of the agent's actions on their environment, which are the affordances of the agent-environment system.

Whether an affordance i.e. an association of a (behavior, entity) and an effect are reinforced or not depends on factors of uncertainty (an action can sometimes fail to produce an expected effect) and of non-stationarity (the environment and the agent change over time). Affordances rely on the agent's perception of features and are thus prone to uncertainty. For example, a broken chair might

seem to afford sit-ability, although it has a chance of collapsing if someone sits on it. As an agent learns new skills, its affordance map changes accordingly. For instance, growing infants develop new motor skills which bring new possible actions and thus new affordances [10]. Affordances of objects change as well: a cup affords fill-ability only when not already full.

2.3. Distinctions between social and physical affordances

One central question in the recent literature is to which extent affordances in social contexts require specific mechanisms in comparison to non-social situations. Building on one of de Carvalho's definitions [12], for whom social affordances can be *"possibilities for interaction that other persons or animals afford"*, we define social affordances as action possibilities offered by the presence of other agents within the close environment. In social affordances, the environmental relata of the affordance is either another agent who can reciprocate or an object whose affordance is available thanks to the presence of another agent; as opposed to physical affordances in which there is no social interaction. For example, the outcome of a social affordance could be to make a human wave their hand or to make a human grab an item that is out of reach for the agent. However, social affordances may, in general, be more volatile and uncertain than in non-social domains, because other agents' behavior changes depending on context, mood, and inter-individual differences. For example, waving at someone on the street might trigger various effects. The person could wave back, stop, stare at the person waving, or come to them and others. Even with a well-defined design, taking into consideration social and cultural rules or whether the two people know each other, the produced effect remains uncertain. As an illustration, in a human-robot interaction task, Uyanik and colleagues expected that the voice behavior *bye* from the robot would make the participants leave the scene, but they mostly kept their position and waved at the robot [5].

One remaining question is whether social affordances and physical affordances are fundamentally different or whether they can be perceived and learned using the same tools. Perception of physical affordances mainly relies on visual cues while social affordances may not. For instance, there is a good chance of knowing whether an item affords grasp-ability simply by looking at the size of the gripper and of the object, or a good chance of knowing whether it affords fill-ability based on its shape. Although other non-visual information about the environment can help further perceive affordances (e.g. weight, texture), vision is the main tool to study physical affordances in psychology and robotics [4]. Similarly, for social affordances, it is sometimes possible to infer what affordances are feasible by the other agent using visual cues: most humans can wave their hands, reach a tool based on the length of their arm, grasp an item depending on the size of their hands and the item, climb the stairs depending on their leg lengths and the height of a step [13]. However, even if it is sometimes possible to infer whether some affordances are feasible, it is hard to predict what behavior from the agent will trigger the desired action from the human, possibly leading to the desired effect (e.g. induce the other agent to do a grasping tentative on a red cube). The most efficient and common way to solve this problem is to ask the human to perform an action explicitly using language [5], which however assumes common learned representations for the two social agents (e.g. knowing what a red cube is, and knowing what grasping means).

Contextual information or basic models of interaction are mandatory to perceive social affordances, instead of visual cues only. For example, factors such as human attention and engagement are highly variable and a given action directed towards the same human twice may not produce the same effects. A human who's been asked for a handshake a few times in a row may refuse the interaction, whereas a cube is always graspable as it does not reciprocate.

Finally, humans have different ranges of skills and do not afford the same complex tasks. For instance, two people with different jobs are likely to have different fields of knowledge, hence producing different effects when asked to do a task related to their job. Social affordances may thus differ from physical affordances and higher uncertainties coming from human particularities could induce specific ways of learning them. Since humans have different skills and personalities, we propose that humans and robots may need to build two distinct types of social affordance models: human-general affordances, which would predict the average effect of an action in a social context (e.g. humans generally wave back); and human-specific affordances, with specific ways of behaving depending on the human partner. Human-general, human-specific, or physical affordances could then be learned in parallel or successively, depending on the task and the experience of the agent.

Based on this brief literature review and theoretical considerations, from now on, we will consider affordances as relations between an agent and its environment, producing an effect either on the agent or on the environment. Affordances are learned by the agent e.g. using self-generated goals, are subject to uncertainty and non-stationarity, and can be physical or social.

2.4. Affordances in robotics

Robots need to understand how to act on their environment to operate efficiently. On the one hand, they need to interact with inanimate objects, to solve tasks of manipulation or navigation. On the other hand, they need to interact efficiently with humans to integrate into society. We propose to study these two aspects of robot autonomy through the scope of physical and social affordances.

Affordances in robotics mostly focus on physical affordances, which are the action possibilities associated with classes of objects. There were several reviews in the field [3,4,6,14]. Jamone et al. review work on affordances in psychology, neuroscience, and robotics [6]. Zech et al. provide a systematic review of affordances in robotics and classify the articles using a taxonomy that they defined [4]. Ardón and colleagues provide a review focusing on affordance learning and classify the articles depending on whether the affordances are learned, partially learned, or known beforehand [14]. These three reviews show that affordances in robotics mainly focus on object manipulation, such as grasping affordances [15]. Jamone et al. highlight the fact that affordances can help improve the quality of human-robot interaction [6]. Nevertheless, none of the reviews addresses the question of learning and using social affordances in robotics.

Uyanik and colleagues [5] propose to extend the approach of Sahin et al. [2] to social affordances. In their task, an iCub robot learns to interact with humans and physical objects using the same affordance system. First, it samples (effect, behavior, and entity) triplets thanks to a guided interaction provided by the experimenters. Then, it learns to predict the effect of each behavior on a given entity using a support vector machine. Given the object and a desired effect, the support vector machine chooses the behavior that works with the highest probability. The robot can then use these probabilities to plan towards a given goal with forward chaining and breadth-first tree search. This experiment shows that social affordances can be studied with the same formalization as physical affordances. Nevertheless, the affordances are only partially learned in this experiment, through offline guided learning and the robot does not have behavioral flexibility to relearn its mapping online during social interaction. Shu et al. [16] focus on learning social affordances. Using human demonstration RGB-D videos, their algorithm learns semantically meaningful social affordances and implement them on a Baxter robot. However, the robot learns stereotypical social behaviors from videos and cannot reason in terms of the effects of its action, extend its knowledge online out of the given dataset, or chain social affordances for good social interaction. Montesano et al. take a developmental approach with the robot knowing basic sensory-motor coordination skills and acquiring skills of increasing difficulty as it interacts online with the environment [11]. They use Bayesian networks, probabilistic graphical models, to encode affordances. However, the interaction with a human is limited to observing a human-object affordance and trying to reproduce the effect of the affordance in the robot-object system. In a recent paper, Munguia et al. use affordances in a human-robot interaction task to bootstrap the learning of model-free reinforcement learning algorithms [17]. They show that a variant of Q-learning that creates initial Q-tables using affordances and semantics information outperforms a deep reinforcement learning agent. In addition, tabular reinforcement learning agents can quickly update and relearn the desired behavior online, a property that seems essential in social contexts.

This short review on affordances in robotics shows that social affordances have been little studied so far, although mandatory for efficient human-robot interactions. Affordances are often learned offline, with little to no adaptation online which reduces the possible applications in human-robot contexts that are highly variable. Moreover, the reinforcement learning theoretical framework [18] seems to provide an interesting formalism and method to learn social affordances on-the-fly, to adapt to different human behaviors, even in non-stationary and uncertain contexts.

2.5. Affordances and reinforcement learning

Reinforcement learning is a theoretical tool to explain autonomous learning and behaviors of animals, through trial and error [18]. In reinforcement learning, agents try to optimize their behavior in an environment with explicit rewards, by maximizing the sum of rewards they get discounted over time. The reinforcement learning framework is often divided between model-free (MF) and model-based (MB) methods. Model-based methods create and store a model of the environment and infer a policy on it. Model-free methods do not build a model of the environment: the agent's policy is only based on progressively learned (state, action)-value associations.

In model-based RL, the effects of the agent's actions on the environment are stored in a model, which draws a direct parallel with formal definitions of affordances [2]. These acquired relations between agent, environment, and effect are comparable to acquired relations between behavior, entity, and effect. To reinforce the link between model-based reinforcement learning and affordances, Khetarpal et al. [19] introduced the notion of intents, which are desired consequences on the environment or the agent in model-based reinforcement learning. Learning an affordance in this context means finding the subset of the state-action space that generates an intent. This method helps generalize reinforcement learning to equivalence classes depending on the agent's intent (or goal). With their formalism, effects can be on the environment (e.g. changing the position of an item) or on the agent (e.g. moving in a direction).

Graves et al. [20] describe affordances as value predictions in reinforcement learning. In their view, however, only predicting reward is too task-specific and restrictive for learning affordances. Hence, they propose to learn affordances using general value functions, which extend the prediction of rewards to the prediction of any scalar. They claim that perceiving an affordance is similar to learning the initiation set of an option, temporally abstract actions introduced by Sutton et al. [21], tied to its success. Their view of affordances in reinforcement learning is close to the notion of intents introduced by Khetarpal et al. [19]. However, Graves and colleagues [20] promote the use of general value functions as a means to solve the reinforcement learning problem, as opposed to Khetarpal and colleagues [19] who use partial models. In a follow-up paper, Khetarpal and colleagues extend their results on intents for actions to intent for options [22], which narrows the gap between these two views of affordances in reinforcement learning.

Recent work demonstrates that reinforcement learning can be a useful tool to model and learn affordances. First, learned transitions of the transition model are an interesting proxy for affordances in model-based reinforcement learning. Second, intents further formalize affordances in model-based reinforcement learning [19]. Third, options allow chaining affordances together so that complex and efficient behaviors are learned [22]. Fourth, reinforcement learning's flexibility permits to adapt to social and non-social contexts online [17]. One limitation to studying affordances in reinforcement learning may come from the task-specific reward function [20]. Thus, goal-conditioned reinforcement learning and reward-free exploration approaches could be other interesting means of learning affordances in reinforcement learning [23].

2.6. Summary

With this short review, we highlighted that affordances in robotics were mostly learned using offline supervised learning in non-social contexts. Social affordances, defined as action possibilities offered by the presence of other social agents within the close environment are key factors for efficient human-robot interaction. We argue that they should be thoroughly studied to improve

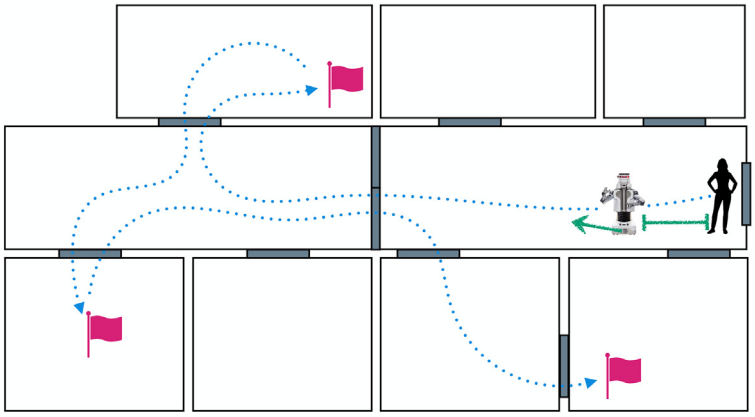

Fig. 1. The "visit the lab" task. The robot, here represented with a PR2 picture, has to lead a human to several goal rooms in a row, represented by the red flags. The robot needs to learn to enter the human visual field, grab the human's attention, and lead them to the goal rooms, without losing their attention. All the figures in this article are available under a CC-BY 4.0 licence (https://doi.org/10.6084/m9.figshare.24910998).

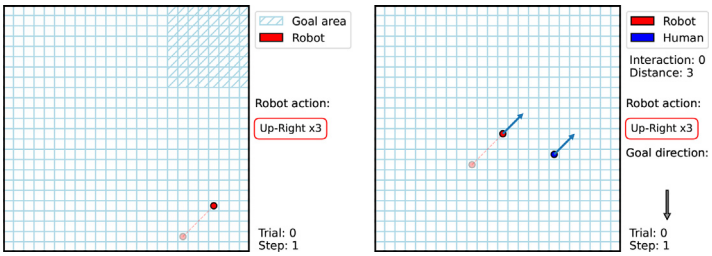

Fig. 2. Graphical representations of the navigation and the social task. (Left) Simulated navigation task. The robot's goal is to reach the hatched area (which represents one of the nine rooms) and to stay there, here in the top-right corner. (Right) Simulated social task. The goal of the robot is to reach the maximal interaction for which the human follows it. Then, it should try to induce a human movement in the indicated goal direction, here southwards. The robot is in red and the human is in blue. The robot's last action is represented with a dashed line and its previous position by a dimmed robot.

the autonomy of robots. Moreover, the reinforcement learning framework can build internal models of affordances online. More specifically, we propose that model-based reinforcement learning combined with goal-conditioning for reward-free exploration is a great framework for learning physical and social affordances under non-stationarity and uncertainty.

3. Presentation of the task

We propose a new task to study physical and social affordances in a reinforcement learning context. In this task, a robot and a human interact in a simulated environment composed of a corridor and several rooms. The goal of the robot is to lead the human to different goal rooms in a row. This task mixes navigation as the robot and the human move in the environment, and social interaction as the robot and the human can communicate using social actions. This task represents a hypothetical scenario in which a robot offers a laboratory visit to a human who visits it for the first time.

3.1. Interest of the task

The requirements for the robot are twofold: it needs to understand the consequences of its movement in the environment and to understand how to interact with the human. Mixing social interaction and navigation permits to study physical and social affordances concurrently. For instance, if the door of a room closes, the effect of some navigation actions will be modified, and the robot may thus need to relearn its action policy to go to the goal room. Similarly, when the robot faces a new unknown human, it may need to adapt to their singularities. We propose to use model-based reinforcement learning to learn both models of navigation and social interaction, as a proxy for affordances.

3.2. Modelling the environment

The environment we simulated consists of a tabular grid world, with different goal areas. We use a square environment of 24×24 cells divided in 9 goal rooms of size 8×8, shown in Fig. 2. This version of the task does not include walls to simplify the vision

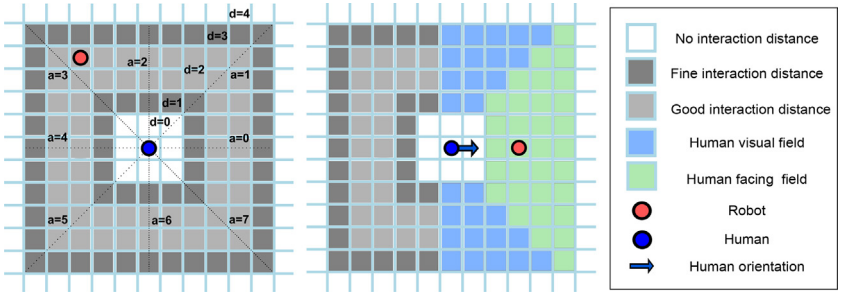

Fig. 3. Model of social interaction. Appropriate distances for social interaction are represented in gray. The human is in blue and the robot is in red. (Left) The robot gets ordered categorical distance from 0 to 4, with 0 being too close and 4 being too far, triggering a loss of attention from the human. Distances of 1, 2, and 3 are appropriate for interaction. However, the robot can learn that it is getting too close or too far from the human if the distances are 1 or 3 respectively (dark gray). The robot gets the angle of its relative position to the human with discrete values from 0 to 7. The dashed lines represent these 8 possible values of angles. The robot gets the angle value of the closest dashed line to its position. Distances are indicated with the letter *d* and angles with the letter *a*. In this figure, the robot is at a distance of 2 with an angle of 3. (Right) The robot needs to execute actions in the human's visual field (in blue) or their facing field (in green) to get their attention. Being in the facing field grasps the human's attention more certainly than when being in the visual field only.

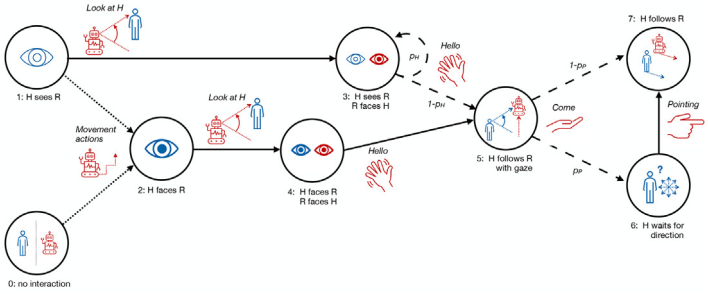

Fig. 4. The eighth interaction levels and the associated transition graph. Interaction levels are represented by circles and the actions that allow transitions to other levels by arrows. Dotted arrows indicate that multiple movement actions will be necessary to reach interaction level 2 from levels 0 or 1; in our first version of the control architecture presented in Fig. 5, this is handled by a dedicated option. Dashed arrows indicate uncertain actions, whose success depends on p_H or p_P (see text for details). From level 0 to level 4 included, only the relative positions and orientations of the human and the robot change the interaction level.

processes for social interaction presented in section 3.3: the robot and the human see each other when looking in each other's direction. At each time step, the agent can move in one among 8 possible directions, at one among three possible speeds (1, 2, or 3 tiles away from its current position), or stay in its current cell. In total, there are $24 \times 24 \times 9 = 5184$ states and $8 \times 3 + 1 = 25$ actions, hence $25 \times 5184 = 129600$ (state, action) couples. The robot's goal is to learn sets of actions to go to each goal room. The information provided to the robot is composed of (1) its current position and (2) the current goal room to reach.

3.3. Modelling social interaction

The robot has to get human attention to engage them in the navigation process. First, it needs to go to their visual field (blue or green zones on Fig. 3, right) at an appropriate distance (gray zones on Fig. 3, left), to face them, do the 'hello' action so that the human follows the robot with their gaze (using rotation movements), and to ask to follow with the 'come' action. Some humans also need the robot to point in a direction to start following the robot. These three actions are expected to be implemented by gestures in real robots. The robot is rewarded when it reaches the maximum level of interaction and makes the human move in the input goal direction. The inputs for the social module are the goal direction between 0 and 7, the interaction level between 0 and 7, the distance between 0 and 4, and the relative angle between the human position and the robot position mapped from the continuous range $[0, 2\pi]$ to the discrete range from 0 to 7. Distances and angles are represented in Fig. 3.

The robot and the human have 8 levels of interaction that are provided to the reinforcement learning system (Fig. 4). If the robot is in the visual field of the human but not in the facing field of the human (interaction level 3), the 'hello' action fails with the failing rate p_H. In the facing field (interaction level 4), this action succeeds with a 100% success rate. If the 'hello' action succeeds, the robot can then try to make the human follow its movements with the 'come' action. The 'come' action either makes the human follow the robot with the probability $1 - p_P$ or makes the human wait for a pointing gesture towards the desired direction with the probability p_P. From level 5 and higher, the human follows the robot with their gaze and rotates to keep it in their facing field. At level 7, the human follows the robot. From level 5 and higher, the human can lose their attention with the rate p_A. If the human loses their attention, the interaction level becomes the level of the relative positions and orientations (from 0 to 4).

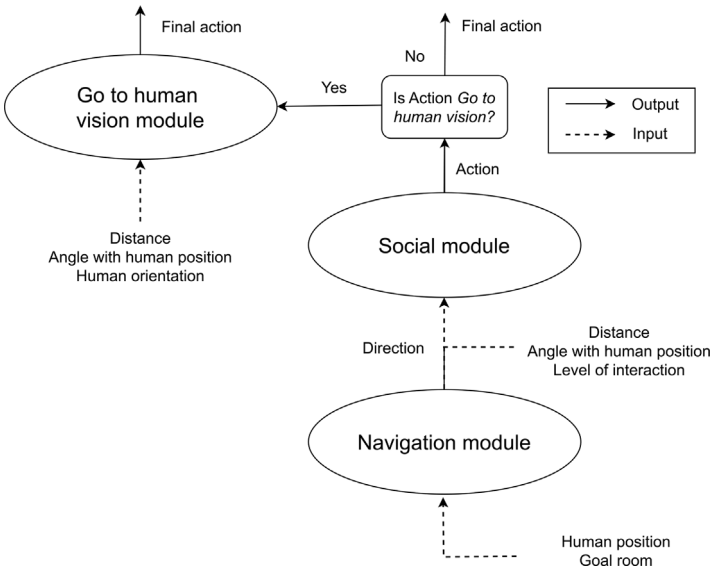

Fig. 5. Modular architecture to guide the robot behavior. First, the navigation module finds the direction towards which the human should move to reach the goal room. Second, the social module decides which action the robot should take depending on this direction and the level of interaction. If this action is not the 'go to human vision' option, then it is the final robot action. Otherwise, the option module provides the final action.

Table 1
Parameters of three humans used for simulating social behaviors. The three human behaviors are ranked in order of social interaction difficulty, from a constant speed to a close to random speed. Different human behaviors are studied in Fig. 8.

Property	Symbol	Human 1	Human 2	Human 3
Human speeds	p_S	[1,0,0]	[0,0.5,0.5]	[0.33,0.33,0.34]
Failing hello	p_H	0	0.05	0.5
Random movement	p_M	0.1	0.05	0.2
Random eye movement	p_O	0.1	0.15	0.3
Losing Attention	p_A	0	0.05	0.1
Pointing need	p_P	0	0	0.5

In the social module, the robot has access to the 25 navigation actions and to 5 social actions: going to the human visual field, looking at the human, saying 'hello' to the human, asking the human to come, and pointing in the desired movement direction. The input information is composed of the relative distance between the robot and the human, the relative position angle, the desired movement direction, and the level of interaction. In total, there are $5 \times 8 \times 8 \times 8 = 2560$ states and $25 + 5 = 30$ actions, thus $2560 \times 30 = 76800$ (state, action) couples.

3.4. Modelling human behaviors

To account for human variability in social interaction, we modeled different human behaviors. Humans can move at three different speeds, represented by a probability distribution over speeds p_S. When they move, their speed is sampled at each step from this distribution. For example, the vector [0, 0.3, 0.7] means that at every time step, the human has 30% chances of moving at speed 2 and 70% chances of moving at speed 3. When the robot is in the human visual field and not in the human-facing field, the 'hello' action fails with the probability p_H which depends on the human. Humans randomly move when not engaged with the probability p_M and randomly change their visual field orientation with the probability p_O. When engaged, i.e. when their interaction level is equal to 5 or higher, they can lose their attention with the probability p_A, returning to the level between 0 and 4 accurately representing the situation. When the robot executes the 'come' action, humans start following with the probability $1 - p_P$ or wait for a pointing action with the probability p_P. Parameters of simulated humans are presented in Table 1.

4. Solving the task

There are several approaches to solve the new task presented in section 3. Here, we present one method to solve the task with reinforcement learning agents and a predefined modular architecture for the robot's decision-making.

4.1. Architecture of the robot's decision-making process

To demonstrate that reinforcement learning can solve the "visit the lab" task, we designed a first hierarchical architecture (Fig. 5). The navigation module provides the direction towards which the human should move based on the human position and the goal room. For each direction, the agent averages the Q-values of the three actions and chooses the direction with the best average value. If the navigation module finds more than one best direction and if the preferred direction the agent chose so far is not part of the best directions, the new preferred direction is chosen randomly amongst the best directions. The social module takes this direction as an entry, with the human-robot distance, the angle between their two positions and the level of social interaction. If the action chosen by the social module is not 'go to human vision', then it is the final action. Otherwise, the 'go to human vision' module provides the final action depending on the relative distance, the relative angle, and the human orientation.

For learning the social interaction module, going to the human vision action teleports the robot to the human visual field. In the "visit the lab" task, this action is the 'go to human vision' option which is learned by an independent module (Fig. 5). The robot has access to the 25 navigation actions to learn this option and takes its decision based on the 5 relative distances with the human, the 8 angles with the human position, and the 8 possible human orientations. In total, there are $8 \times 8 \times 5 = 320$ states and 25 actions, thus $320 * 25 = 8000$ (state, action) couples.

4.2. Reinforcement learning algorithms

To solve the task, we use and compare three classical reinforcement learning algorithms such as Q-learning ϵ-greedy (model-free ϵ-greedy), ϵ-greedy value iteration (model-based ϵ-greedy) and Rmax (model-based greedy in the face of uncertainty) [18,24]. Each of the three modules can be learned separately by any learning agent. The learning parameters of the agents can be fine-tuned depending on the module to learn. One iteration of the value iteration inference process is done on all of the (state, action) couples for the model-based agents every 1000 step. This process saves computational costs due to large tables of rewards and transitions. The parameters used for our simulations are $\gamma = 0.9$ for all agents, $\epsilon = 0.05$ for ϵ-greedy agents, $\alpha = 0.5$ for the model-free agent, $R\max = 1$ for Rmax and $m = 1$ for Rmax in the navigation task, $m = 5$ for Rmax in the two other tasks.

Model-based agents learn an internal model of the environment using the frequency of observed events. For each action they take in a given state, they store a reward model and a transition model. Let $\hat{R}$ be the reward function, $\hat{T}$ be the transition function, for all state s, action a and arrival state s',

$$\hat{T}(s,a,s') = \frac{n(s,a,s')}{n(s,a)}, \qquad \hat{R}(s,a) = \sum_{k=1}^{n(s,a)} \frac{r_k(s,a)}{n(s,a)},$$

where $n(s,a,s')$ is the number of times the agent arrived in state s' after taking action a in state s, $n(s,a)$ is the number of times the agent took action a in state s, and $r_k(s,a)$ is the reward obtained the k-th time the agent took the action a in state s.

When each module is learned, it can be assembled with the others to assess the performance of the robot in the global task using the architecture presented in Fig. 5. Beyond this proof of concept, we propose to study the interest of a model-based/model-free combination and show the capacity to adapt to different human behaviors shown in Table 1, provided by reinforcement learning. As the architecture makes an explicit distinction between social interaction and navigation, the robot can re-explore one module specifically. For example, if the robot encounters a new human in the same environment, it can re-learn appropriate social behaviors online, without changing its navigation module.

5. Results

We evaluate each agent over 10 seeds for each plot. The plots show the mean and the standard deviation for each condition. For the three modules, agents are evaluated on sequential tasks of 20 steps, corresponding to 20 action choices for the robot. In the global task (when putting the modules together), trials last 100 steps. In the navigation task, the agent is rewarded when it reaches the goal room and stays in it. In the social task, the agent gets rewarded when the human is following it and moves in a randomly selected direction. In the 'go to human vision' task, the agent is rewarded when it is in the visual field of the human. For the "visit the lab" task, with the three modules together, the reward is the number of goal rooms the human reached, with the goal room changing randomly every time the previous goal is reached.

In the social task, we train the reinforcement learning agents in an environment in which the human cannot reach external walls to prevent side effects (e.g. the robot cannot make the human move in the desired direction because of a wall). In practice, we train the agents in a squared environment of size 120×120, with humans starting in the middle of the environment. Thus, the human cannot reach the walls in 20 steps even with a speed of 3. To add noise in the simulation and to fit better with what the robot will need to do in the "visit the lab" task, the human desired direction of movement changes randomly at every time step with a probability of 20%.

5.1. Learning the three modules

In the navigation task (Fig. 6, top-left), ϵ-greedy agents converge faster than Rmax but at a slightly lower value. This is partly due to the ϵ rate of exploration that ϵ-greedy agents keep even when they have learned the task, as opposed to Rmax. In the 'go to human' task (Fig. 6, top-right), the performance of all three agents is rather similar and they all converge quickly. In the social task (Fig. 6, bottom-left), the model-based version of ϵ-greedy converges faster than Rmax, while the model-free agent does not converge

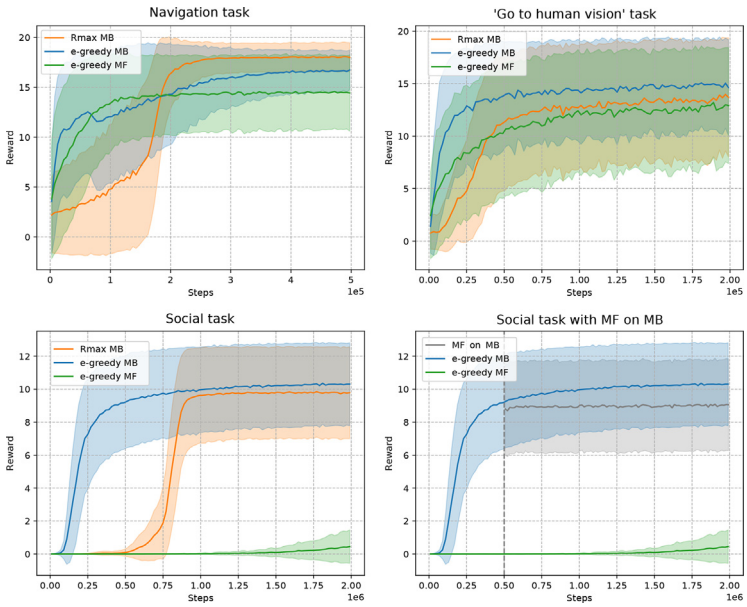

Fig. 6. Learning of the three modules by reinforcement learning agents. (Top and Bottom-Left) Learning of the three modules by the three agents described in Section 4.2. In the three tasks, model-based agents perform as well or better than the model-free agent. (Top-Left) Navigation task with the agent learning to go to the 9 goal rooms. (Bottom-Left) Social Task with Human 1 as described in Table 1. (Top-Right) Go to the human visual field task. (Bottom-Right) Performance on the social task of three ϵ-greedy agents, one model-based (in blue) and two model-free (in green and gray). The gray model-free agent learns from observing the rewards of the model-based agent until step 500,000, represented by a dotted vertical line. From this step, it takes decisions alone based on its learned Q-values. The bootstrapped model-free agent outperforms the model-free agent alone.

within the studied timescale. In all three tasks, model-based agents perform as well or better than the model-free agent and can thus be used for learning all three modules.

5.2. Model-based / model-free cooperation

As shown in Fig. 6, bottom-left panel, the model-free ϵ-greedy agent did not learn to solve the task after 2 million time steps of training in the social task. This poor performance is due to the poor sample efficiency of model-free agents, which need many experiences to converge. Although sample efficient, model-based agents have a huge computational cost due to inference processes, especially in large environments: as presented in Section 4.2, we decided that the model-based agents would perform one step of value iteration only once every 1000 steps. Even if they are not sample efficient, model-free agents have a computational cost advantage over model-based agents, as they do not have to infer the Q-values based on their model of the environment. Cooperation of model-based and model-free agents precisely enables keeping a performance level close to the one of the model-based agent, while reducing the global computational cost [25]. With the coordination criterion we proposed in previous works [26], the model-free agent first learns by observing decisions made by the model-based one, and gradually takes control of the agent, thereby reducing computational costs.

As we plan to use this strategy in future developments, we make here a preliminary test. We check whether the performance of the model-free agent improves faster when observing the model-based agent at the beginning of learning.

To do so, the model-free agent is passively learning from the model-based agent experience for 500,000 time steps. Then, the model-free agent splits up from the model-based one and makes decisions based on the Q-table learned from observation. The bottom-right panel of Fig. 6 shows that this bootstrapped model-free agent outperforms the model-free agent alone. This confirms that the dynamic coordination of both agents has the potential to improve performance while diminishing computing resources.

However, what the model-free agent observes depends on what the model-based agent does. For example, the exploration method of the model-based agent modifies what the model-free agent learns. In Fig. 7, we show the pattern of Q-values learned in the navigation task by a model-free agent either alone, either observing an ϵ-greedy model-based agent or observing a Rmax agent. The Q-values learned by the model-free agent from the Rmax behavior are very different than for ϵ-greedy agents. For Rmax, we observe a jump every three steps which comes from a systematic exploration method followed by a greedy exploitation. Hence, exploration methods of the model-free and the model-based agents need to be studied to further understand their dynamic cooperation.

5.3. Human variability

Depending on the human the agent is interacting with, the robot is more or less likely to succeed in bringing the human in the desired direction. For example, a very inattentive human will not engage well in the interaction and it will be harder for the robot to

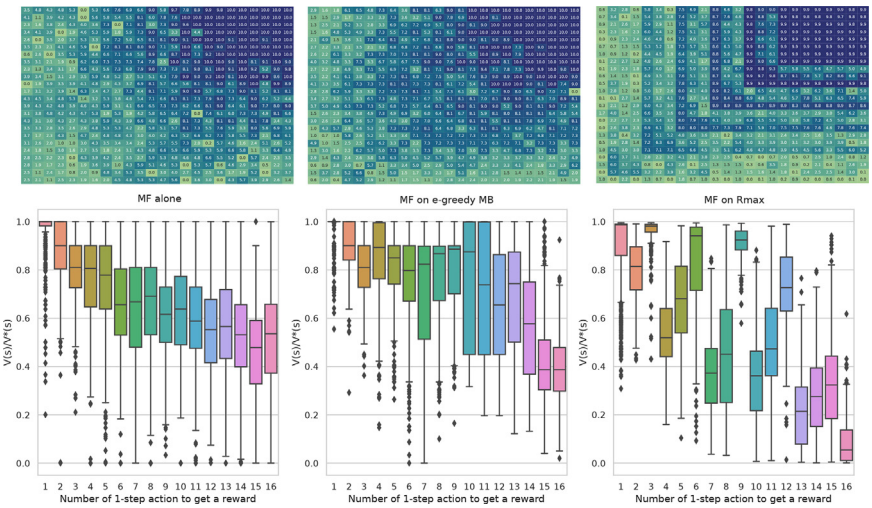

Fig. 7. Q-values learned by a model-free agent on the navigation task. The model-free agent either learns the values exploring alone (Left), observing an ϵ-greedy model-based (Middle), or observing Rmax in the navigation task (Right). The map learned by the model-free agent depends on the exploration method of the model-based agent. Agents are all studied after 400,000 steps where all model-based agents have converged as shown on the top-left plot of Fig. 6, with the parameters indicated section 4.2. (Top) Examples of 2D heat-maps of the learned Q-values for all three models, when the reward area is the top-right corner. (Bottom) Barplots showing the learned Q-values over the best theoretical Q-values depending on the Manhattan distance to the goal room averaged for the 9 goal rooms for the three agents and on 10 trials for each agent. Because of the systematic initial exploration of Rmax, followed by a greedy policy, the Q-values the model-free agent learned are high every three steps.

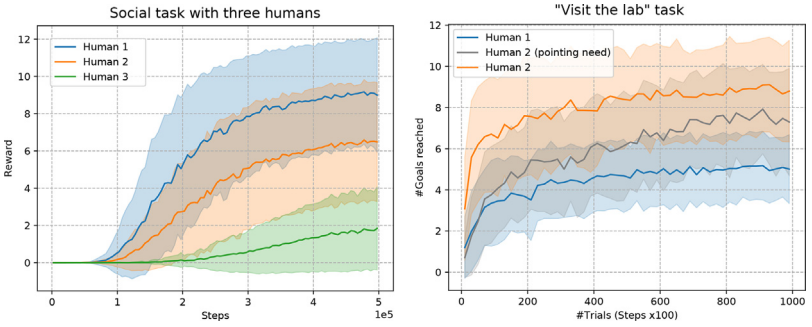

Fig. 8. Learning and adaptability to different human behaviors. (Left) Learning of the social module for the three humans presented in Table 1, with the same model-based ϵ-greedy agent. (Right) Number of areas reached in the "visit the lab" task depending on the human the robot is interacting with. The robot has learned the task with Human 2 on the social task and tries to generalize its human-specific model to other social behaviors. Human 1 is slower than Human 2 so the agent needs to learn to lower its speed so as not to lose human attention. To interact with human 2 with a pointing need, the agent needs to learn how to use the pointing social action which it didn't need so far.

bring them to goal rooms. To model these different behaviors, we propose to use three models of humans shown in Table 1. The first human moves at a constant (slow) speed and is very attentive. The second human moves rather fast and can infrequently lose their attention. The third human is very inattentive, moves at a random speed between the three possible speeds, and randomly moves when not engaged more frequently than the two other agents.

For the three humans, a model-based ϵ-greedy agent is trained to learn the social task. If going too fast, the robot can easily exceed the maximal interaction distance and lose human attention. The left plot of Fig. 8 shows that the robot gets more rewards interacting with Human 1. This matches our intuition that social behavior is not hard to learn with an attentive human who moves at a constant speed. The second human's interaction pattern is more predictable than the third human's one. Thus, the robot manages to get more rewards with Human 2 than with Human 3 on average. The results obtained in this social task show that with our simulated humans, we can produce a wide range of behaviors from easy to hard to learn.

However, one central question in our view of social affordances is whether one behavior learned while interacting with one human can be generalized to different classes of humans, and whether the robot should adapt by building human-specific models when the behavior is too extreme. In Fig. 8-right, we show the performance of the robot on the global task (bringing the human to several goal rooms in a row). Each module has been learned using the agent which converged to the highest reward for each task: Rmax for the navigation task and ϵ-greedy model-based for the social and the 'go to human vision' task. The social agent updates its Q-values to adapt its behavior online to the different humans it faces (whereas the Q-values for the navigation and the 'go to human vision' tasks are not modified). The Q-values and the model of the environment are saved for one agent after 500,000 steps for the navigation

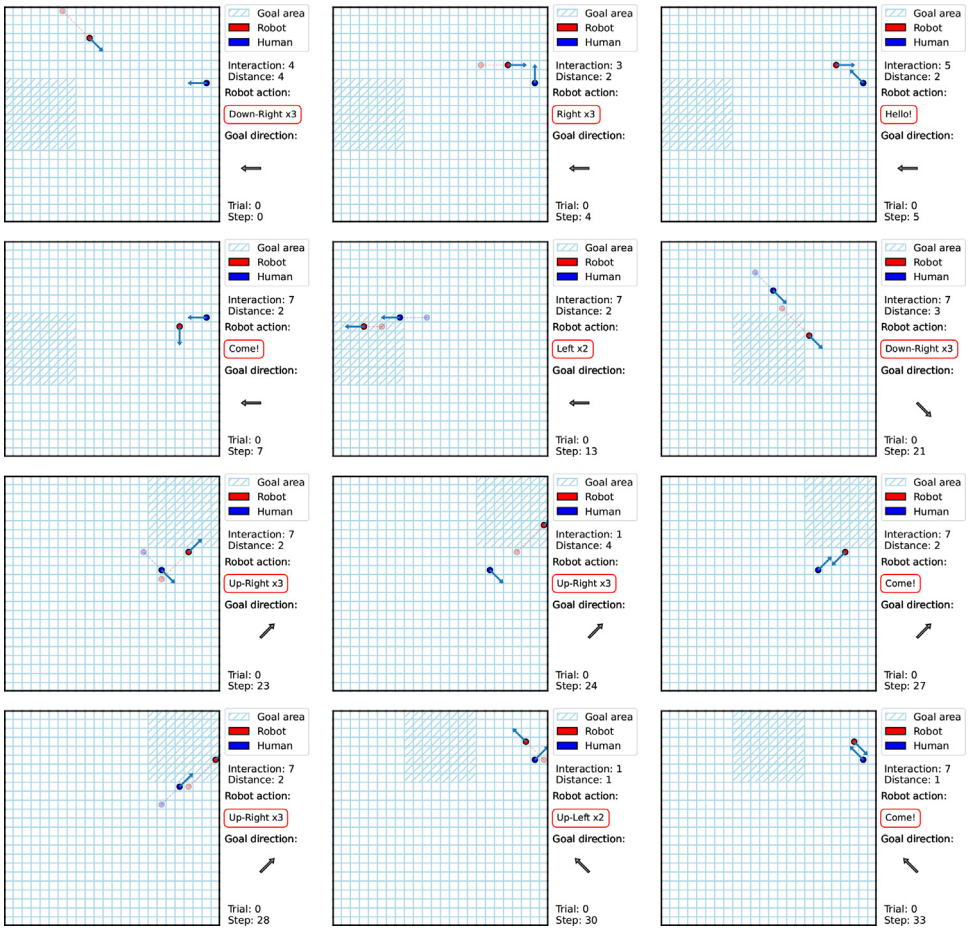

Fig. 9. Example of an interaction between the robot and Human 2 after learning the three modules. Robot actions, steps, and the desired directions to reach the goal room taken from the navigation module are indicated for each picture. At the start of the experiment, the robot tries to go to the human visual field. It reaches it at a proper distance at step 4. Then, it gets the human attention with the hello and the come action (step 5 and step 7). At step 13, it reaches the first goal room with the human, receiving a reward. At step 22, it reaches the second goal room. Unfortunately, the human loses its attention at step 24, right before the human reaches the goal room, due to a non-zero probability for human 2 to lose its attention, even if at the right distance of interaction ($p_A = 5\%$). The robot manages to get the human attention back at step 27 and the human reaches the goal room at step 29. However, the human goes too close to the human and the interaction level drops to 1. The robot has to get the human attention back, which it manages to do at step 33.

module and the social module, and 200,000 steps for the 'go to human vision' module. The social agent used in this global task is a model-based ϵ-greedy similar to the one used so far but which performs one step of value iteration once every 100 steps instead of once every 1000 steps when learning the three modules. We increased the budget for planning for our model-based reinforcement learning agent to speed up the learning on the "visit the lab" task.

The social module has been learned by interacting with Human 2 presented in Table 1. We test the adaptability of the robot when facing Human 2, Human 1, and an alternative version of Human 2 with a pointing need (with the same parameters but $p_P = 1$ instead of $p_P = 0$). For the three humans, there is a rise at the start of the interaction when putting all modules together. Indeed, the social module was trained with a 'go to human vision' action which teleported the robot in front of the human and never failed. However, now that the agent uses the 'go to human vision' option, reaching the visual field of the human is not immediate. Thus, the agent needs to adjust the value of this option, which results in a learning phase. Apart from this initial rise, the number of goals reached when interacting with Human 2 only increased slightly over trials, which shows that the social interaction was already learned. When facing Human 2 with a pointing need, the robot needs to learn to point in the desired direction to engage the human's interest. The performance of the agent when facing Human 2 with a pointing need increases from 1 (3 after the initial rise) to more than 7 areas reached per trial. The robot also manages to adapt to the behavior of Human 1 going from 1.5 (3 after the initial rise) to 5 goal areas reached per trial. However, the robot seems to adapt less to the behavior of Human 1 than it does to Human 2 with a pointing need. This comes from the fact that the behavior of Human 2 with a pointing need is close to that of Human 2. On the opposite, the robot needs to learn a new speed when facing Human 1. In addition, Human 1 moves slower than Human 2, which also limits the number of goal areas Human 1 can reach in a given number of time steps. To illustrate the "visit the lab" task, we show one example of interaction with Human 2 in Fig. 9.

6. Conclusions

In this work, we proposed to study the combined learning of social affordances (to interact with other social agents) and physical affordances (to interact with the physical world, such as when navigating or manipulating objects) within the model-based reinforcement learning framework. We first reviewed the affordance literature in psychology and robotics and highlighted that little attention had been paid to social affordances, and then that most previous papers used supervised learning to learn them, rather than RL. We then proposed a new task that prompts robots to learn and combine social and physical affordances: the ”visit the lab” task. The robot needs to engage the human in following it towards a series of goal rooms. It further needs to learn to react to different levels of distractability in different humans. It finally needs to combine these social affordances with learned navigation behavior to reach the goal.

We presented a series of simulations of the ”visit the lab” task, comparing the ability of different model-based and one model-free reinforcement learning algorithms. First, we showed that the model-based agents were able to learn the three modules, while the model-free agent didn’t manage to learn the social task. However, it was possible to bootstrap the performance of the model-free agent with the passive observation of a model-based agent at the start of the experiment. Then we demonstrated that our task permits studying different human behaviors, with different success rates depending on the human parameters. Finally, we proved that it was possible to solve the task using the architecture proposed in Fig. 5. In this architecture, the affordances are the transition models of the model-based agents, either social in the social module or physical in the navigation module. Furthermore, we showed that reinforcement learning permits to adapt to new behaviors observed online.

7. Limitations and future work

More elaborated reinforcement learning agents could further speed up the adaptability to new social behaviors online. A first approach to deal with changes is to build the transition and reward model of the model-based agents only with the last occurrences for each (state, action) couple [25]. Although this approach adapts well to changing environments with a low computational cost, the number of passages the agents consider limits the precision of the models. In addition, the agents forget previous information as they adapt to the new behaviors online, without the capacity to re-use previous knowledge. With this approach, the robot would have to relearn the model of a human even if it interacted with them before. Another solution is to use multi-model reinforcement learning agents to store different models and select one depending on the human the agents are interacting with. The detection of a change could occur when the current model is not accurate enough at predicting the recent observed human behavior (see [27] for a review on context detection). After interacting with enough humans, the robot might have a human-general model, which explains well many human behaviors, and human-specific models when facing contrasting behaviors.

We showed that reinforcement learning agents manage to solve the task with a pre-defined architecture. However, robots could be more autonomous. In our task, robots learn the different pre-defined modules in parallel without exploring when and how learning each module becomes relevant. For example, if the human starts following the robot before the robot knows where to go, the robot could quickly lose the human’s interest. Hence, the robot may want to learn how to navigate to the goal room before learning how to engage the human interest. Choosing which module to explore preferentially could be further assessed through curriculum learning. The robot could focus on learning modules of increasing difficulty using a measure of performance on the sub-modules, such as learning progress [28]. Thanks to curriculum learning, the robot would also naturally explore affordances that have recently changed and need to be learned again. Finally, the robot could create its own architecture combining curriculum learning and goal generation [29,30], without relying on a pre-defined architecture. Goal-conditioned reinforcement learning, which focuses on learning how to achieve generated goals would permit the robot to explore its environment and extract interesting features for solving the task. All these different possible extensions go beyond the scope of the present article, but the proposed task provides a benchmark to compare the efficiency and robustness of these methods.

Generating goals and learning how to solve them also extends the vision of affordances we have used in our simulations. The affordances we presented in our simulations are the consequences of an action on the environment in terms of ’transitions’, learned and handled with the transition function of the model-based reinforcement learning framework. However, as highlighted in the review on affordances, this characterization of affordances through model-based reinforcement learning can be extended to further fit Gibson’s view [1]. Reinforcement learning agents that generate intents and learn them could group actions with a similar effect, hence narrowing the gap with Sahin et al.’s equivalence classes [2]. In addition, learning how to solve long-term goals permits understanding the effect of a sequence of actions on the environment [21,22]. Robots could chain actions to achieve a desired effect they self-generated. We argue that model-based reinforcement learning offers a relevant framework to analyze learned internal models, to merge transitions with similar effects, to concatenate actions into sequences using the ’options’ framework [21], and to compare different merging and concatenation methods in the present paradigm.

One concept we have not debated in this article although central to social interaction is theory of mind. In our architecture, the robot learns the social task using an explicit variable about the level of interaction. However, this hand-coded parameter could be removed if the robot infers the hidden interaction state of the human depending on its behavior. For example, parameters such as human gaze, and displacements, are key for the robot to perceive the human’s internal (cognitive) state. In addition, the robot could use theory of mind to model the capabilities of each human. For instance, the robot could measure the different movement speeds of a human and adapt its own speed accordingly, further specifying the learning of human-specific models. Thus, theory of mind could help the robot understand the hidden variables in social interaction. Furthermore, it could help differentiate social and physical affordances.

One limit of model-based reinforcement learning agents is the planning computational cost in large (state, action) spaces. One possible solution proposed in this article consists of progressively automatizing action control from model-based to model-free, to speed up decision-making. Indeed, we showed that tasks could be learned by a model-free agent passively learning from observing a model-based agent. Recent literature has shown that combining model-based and model-free experts with a meta-controller arbitration reduces computational costs in robotic contexts [25,26]. Another possibility to reduce the computational cost of model-based reinforcement learning is to use hierarchical reinforcement learning. For example, chaining actions together with the option framework reduces the number of elementary operations and the computational cost of planning [21].

In this article, we presented a new paradigm to study social and physical affordances in simulations. Our human model can simulate different movement speeds and reactions to the robot's actions. However, simulations cannot capture all the complexity of human behaviors and human-robot interaction. Future work will adapt the "visit the lab" task to robotic contexts, with a mobile robot equipped with an arm, such as a PR2 or a Tiago robot. Extending our simulated task to real-world experiments necessitates a work of integration so that the robot can detect its position and the human position and movements. Gathering human data from social interaction in the "visit the lab" scenario will help further specify the simulated human models and improve the proposed paradigm.

Copyright

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Augustin Chartouny: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Methodology, Investigation, Conceptualization. **Keivan Amini:** Visualization, Software, Methodology, Investigation. **Mehdi Khamassi:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Benoît Girard:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Acknowledgments

This research was funded in whole, or in part, by the French Agence Nationale de la Recherche (ANR) (ANR-21-CE33-0019-01 ELSA project).

References

[1] J.J. Gibson, The Ecological Approach to Visual Perception, Houghton, Mifflin and Company, 1979.
[2] E. Şahin, M. Cakmak, M.R. Doğar, E. Uğur, G. Üçoluk, To afford or not to afford: a new formalization of affordances toward affordance-based robot control, Adapt. Behav. 15 (4) (2007) 447–472.
[3] E. Renaudo, P. Zech, R. Chatila, M. Khamassi, Computational models of affordance for robotics, Front. Neurorobotic. 16 (2022) 1045355.
[4] P. Zech, S. Haller, S.R. Lakani, B. Ridge, E. Ugur, J. Piater, Computational models of affordance in robotics: a taxonomy and systematic classification, Adapt. Behav. 25 (5) (2017) 235–271.
[5] K.F. Uyanik, Y. Calskan, A.K. Bozcuoglu, O. Yuruten, S. Kalkan, E. Sahin, Learning social affordances and using them for planning, in: Proceedings of the Annual Meeting of the Cognitive Science Society, volume 35, 2013, pp. 3604–3609.
[6] L. Jamone, E. Ugur, A. Cangelosi, L. Fadiga, A. Bernardino, J. Piater, J. Santos-Victor, Affordances in psychology, neuroscience, and robotics: a survey, IEEE Trans. Cognit. Dev. Syst. 10 (1) (2016) 4–25.
[7] M.T. Turvey, Affordances and prospective control: an outline of the ontology, Ecol. Psychol. 4 (3) (1992) 173–187.
[8] E.J. Gibson, R.D. Walk, The" visual cliff", Sci. Am. 202 (4) (1960) 64–71.
[9] K.S. Kretch, K.E. Adolph, Cliff or step? posture-specific learning at the edge of a drop-off, Child Dev. 84 (1) (2012) 226–240.
[10] K.E. Adolph, S.R. Robinson, Motor development, Handbook Child Psychol. Dev. Sci. (2015) 1–45.
[11] L. Montesano, M.C. Lopes, A. Bernardino, J. Santos-Victor, Learning object affordances: from sensory–motor coordination to imitation, IEEE Trans. Robot. 24 (2008) 15–26.
[12] E.M. de Carvalho, Social affordance, in: Encyclopedia of Animal Cognition and Behavior, Springer International Publishing, 2020, pp. 1–4.
[13] W.H. Warren, Perceiving affordances: visual guidance of stair climbing, J. Exp. Psychol.: Human Percept. Perform. 10 (5) (1984) 683.
[14] P. Ardón, È. Pairet, K.S. Lohan, S. Ramamoorthy, R. Petrick, Affordances in robotic tasks–a survey, arXiv preprint arXiv:2004.07400 (2020).
[15] P. Ardón, E. Pairet, R.P. Petrick, S. Ramamoorthy, K.S. Lohan, Learning grasp affordance reasoning through semantic relations, IEEE Robot. Automat. Lett. 4 (4) (2019) 4571–4578.
[16] T. Shu, X. Gao, M.S. Ryoo, S.-C. Zhu, Learning social affordance grammar from videos: transferring human interactions to human-robot interactions, in: 2017 IEEE international conference on robotics and automation (ICRA), IEEE, 2017, pp. 1669–1676.
[17] F. Munguia-Galeano, S. Veeramani, J.D. Hernández, Q. Wen, Z. Ji, Affordance-based human–robot interaction with reinforcement learning, IEEE Access 11 (2023) 31282–31292.
[18] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT press, 2018.
[19] K. Khetarpal, Z. Ahmed, G. Comanici, D. Abel, D. Precup, What can i do here? A theory of affordances in reinforcement learning, in: International Conference on Machine Learning, PMLR, 2020, pp. 5243–5253.
[20] D. Graves, J. Günther, J. Luo, Affordance as general value function: a computational model, Adapt. Behav. 30 (4) (2021) 307327.
[21] R.S. Sutton, D. Precup, S. Singh, Between mdps and semi-mdps: a framework for temporal abstraction in reinforcement learning, Artif. Intell. 112 (1-2) (1999) 181–211.

[22] K. Khetarpal, Z. Ahmed, G. Comanici, D. Precup, Temporally abstract partial models, Adv. Neural Inf. Process. Syst. 34 (2021) 1979–1991.
[23] M. Liu, M. Zhu, W. Zhang, Goal-conditioned reinforcement learning: problems and solutions, arXiv preprint arXiv:2201.08299 (2022).
[24] R.I. Brafman, M. Tennenholtz, R-max-a general polynomial time algorithm for near-optimal reinforcement learning, J. Mach. Learn. Res. 3 (Oct) (2002) 213–231.
[25] E. Massi, J. Barthélemy, J. Mailly, R. Dromnelle, J. Canitrot, E. Poniatowski, B. Girard, M. Khamassi, Model-based and model-free replay mechanisms for reinforcement learning in neurorobotics, Front. Neurorobot. 16 (2022) 864380.
[26] R. Dromnelle, E. Renaudo, M. Chetouani, P. Maragos, R. Chatila, B. Girard, M. Khamassi, Reducing computational cost during robot navigation and human–robot interaction with a human-inspired reinforcement learning architecture, Int. J. Soc. Robot. 15 (8) (2023) 1297–1323.
[27] K. Khetarpal, M. Riemer, I. Rish, D. Precup, Towards continual reinforcement learning: a review and perspectives, J. Artif. Intell. Res. 75 (2022) 1401–1476.
[28] S. Forestier, P.-Y. Oudeyer, Modular active curiosity-driven discovery of tool use, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2016, pp. 3965–3972.
[29] A. Manoury, S.M. Nguyen, C. Buche, Hierarchical affordance discovery using intrinsic motivation, in: Proceedings of the 7th international conference on human-agent interaction, 2019, pp. 186–193.
[30] S. Forestier, R. Portelas, Y. Mollard, P.-Y. Oudeyer, Intrinsically motivated goal exploration processes with automatic curriculum learning, J. Mach. Learn. Res. 23 (1) (2022) 6818–6858.

Augustin Chartouny is a researcher working at the Institute of Intelligent Systems and Robotics on the campus of Sorbonne Université, Paris, France. His main topics of research include decision-making and reinforcement learning in robots and humans, and questions of exploration in the face of uncertainty.

Keivan Amini is a researcher working at the Institute of Intelligent Systems and Robotics on the campus of Sorbonne Université, Paris, France. His main topics of research include reinforcement learning in robots and swarm robotics.

Mehdi Khamassi is a research director with CNRS, working at the Institute of Intelligent Systems and Robotics on the campus of Sorbonne Université, Paris, France. His main topics of research include decision-making and reinforcement learning in robots and humans, the role of social and non-social rewards in learning, and ethical questions raised by machine autonomous decision-making.

Benoît Girard is a research director with CNRS, working at the Institute of Intelligent Systems and Robotics on the campus of Sorbonne Université, Paris, France. His main topics of research include decision-making and reinforcement learning in robots and animals, he contributes to both artificial intelligence and computational neuroscience.