



HAL
open science

The intrinsic motivation of reinforcement and imitation learning for sequential tasks

Sao Mai Nguyen

► **To cite this version:**

Sao Mai Nguyen. The intrinsic motivation of reinforcement and imitation learning for sequential tasks. Machine Learning [cs.LG]. Sorbonne Université, 2024. <tel-04853270v2>

HAL Id: tel-04853270

<https://hal.science/tel-04853270v2>

Submitted on 9 Jan 2025

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

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



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

The intrinsic motivation of reinforcement and imitation learning for sequential tasks

Body movement analysis, activity of daily living analysis and intrinsically motivated robot for hierarchical reinforcement learning

A Dissertation Presented

by

Nguyen Sao Mai

Computer Science, Data and AI

in partial fulfillment of the requirements

for the degree of

Habilitation à Diriger des Recherches

in the subject of Engineering Science

with the Sorbonne Université

Paris, France

27/09/2024

Thesis Rapporteurs

| | | |
|-------------------------|--|---------------------------|
| Prof. LOLA CAÑAMERO | ETIS Lab CY Cergy Paris Université | Director |
| Prof. MOHAMED CHETOUANI | Institut des Systèmes Intelligents et de Robotique | Adjunct Director |
| Dr. SYLVAIN CALINON | Idiap Research Institute, Switzerland | Senior Research Scientist |

Jury President :

| | | |
|--------------------|------------------------|-----------|
| Prof. RICHARD DURO | Universidade da Coruña | Professor |
|--------------------|------------------------|-----------|

Jury Member :

| | | |
|--------------|---|-----------|
| Pr. JUN TANI | Okinawa Institute of Science and Technology | Professor |
|--------------|---|-----------|

© 27/09/2024 - *Nguyen Sao Mai*
ALL RIGHTS RESERVED.

The intrinsic motivation of reinforcement and imitation learning for sequential tasks

My work in the field of developmental cognitive robotics aims to devise a new domain bridging between reinforcement learning and imitation learning, with a **model of the intrinsic motivation for learning agents to learn with guidance from tutors multiple tasks, including sequential tasks**. My main contribution has been to propose a common formulation of intrinsic motivation based on empirical progress for a learning agent to choose automatically its learning curriculum by actively choosing its learning strategy for simple or sequential tasks: which task to learn, between autonomous exploration or imitation learning, between low-level actions or task decomposition, between several tutors. The originality is to design a learner that benefits not only passively from data provided by tutors, but to actively choose when to request tutoring and what and whom to ask. The learner is thus more robust to the quality of the tutoring and learns faster with fewer demonstrations.

With my previous advisors, colleagues and my collaborators, we developed the framework of **socially guided intrinsic motivation** with machine learning algorithms to learn multiple tasks by taking advantage of the generalisability properties of human demonstrations in a passive manner or in an active manner through requests of demonstrations from the best tutor for simple and composing subtasks. The latter relies on a representation of subtask composition proposed for a construction process, which should be refined by representations used for observational processes of analysing human movements and activities of daily living. With the outlook of a language-like communication with the tutor, we investigated the emergence of a symbolic representation of the continuous sensorimotor space and of tasks using intrinsic motivation. These works proposed within the reinforcement learning framework, a **reward function for interacting with tutors** for multi-task learning. The formulation of this reward unifies interactive learning, multi-task learning and hierarchical learning through the same empirical competence progress measure as the intrinsic motivation to choose the different aspects its learning strategy.

The socially guided intrinsic motivation framework could be used for intelligent tutoring systems to be applied in socially assistive robotics for a robot coach for physical rehabilitation of low-back pain patients analysing whole body movements, or for a robot coach for ASD children in an imitation game.

The project I wish to develop in cognitive developmental learning pertains both to the reinforcement learning framework of machine learning, to control learning in the perception-action loop of cognitive robotics and the theory of intrinsic motivation from developmental psychology in cognitive science. The objective is to propose new reinforcement learning algorithms for robots to learn by actively interacting with their environment, enabling robots to master control tasks of growing complexity, while extending the current theory of intrinsic motivation towards a unified model of motivations for interacting with tutors. The project relies on two axes : a model of interaction with tutors relying on an emerging representation of tasks by the robot, and a model of motivation of human learners through the use case of a robot coach for physical rehabilitation. This work will enrich the formulation of the motivation for social guidance with various motivational drives of learners to interact with tutors, taking the perspective of communication and emotions. This will also highlight the importance of emerging internal representations for understanding social guidance, thus emphasising the interdependence between autonomous learning and social learning.

Keywords : Interactive Reinforcement Learning, Active Imitation Learning, Social Guidance, Intrinsic Motivation, Reinforcement Learning, Hierarchical Learning, Human Movement Analysis, Activities of Daily Living, Activity Recognition, Socially Assistive Robotics, Smart Home

La motivation intrinsèque d'apprendre par renforcement et imitation des tâches séquentielles

Ma recherche dans le domaine de la robotique cognitive développementale cherche à concevoir un nouveau domaine qui allie apprentissage par renforcement et apprentissage par imitation, avec un modèle de motivation intrinsèque pour que des agents apprennent plusieurs tâches, dont des tâches séquentielles, avec la guidance de tuteurs. Ma principale contribution a été de proposer une formulation commune de la motivation intrinsèque, basée sur le progrès empirique, pour qu'un agent choisisse automatiquement son programme d'apprentissage en choisissant activement sa stratégie d'apprentissage pour des tâches simples ou séquentielles : la tâche à apprendre; le tuteur à imiter; stratégie entre exploration autonome ou apprentissage par imitation, entre des actions bas niveau ou une décomposition de tâches. L'originalité est de concevoir un agent d'apprentissage qui bénéficie non seulement passivement des données provisionnées par des tuteurs, mais aussi en choisissant activement quand demander de l'aide et quand et à qui demander. L'apprenant est ainsi plus robuste à la qualité du tutorat et apprend plus vite avec moins de démonstrations.

Avec mes encadrants, collègues et collaborateurs, nous avons développé le cadre de la **motivation intrinsèque guidée socialement** des algorithmes d'apprentissage multi-tâches, en profitant des propriétés de généralisation des démonstrations par des humains observées de manière passive ou de manière active par le biais de requêtes au meilleur tuteur de démonstrations pour des tâches simples ou pour composer des tâches complexes. Cet apprenant exploite des représentations de composition de sous-tâches que nous proposons pour une démarche constructiviste, et qui devrait être précisées par des représentations utilisées pour des démarches observationnelles comme l'analyse de mouvements humains et d'activités de la vie quotidienne. Dans la perspective d'une communication par un pseudo-langage avec le tuteur, nous avons étudié l'émergence de représentations symboliques de l'espace continu sensorimoteur et des tâches, grâce à la motivation intrinsèque. Nos travaux ont proposé dans le cadre de l'apprentissage par renforcement, une **fonction de récompense pour interagir avec des tuteurs** pour l'apprentissage multi-tâches. La formulation de cette récompense unifiant l'apprentissage interactif, l'apprentissage multi-tâches et l'apprentissage hiérarchique, au travers de la même mesure empirique de progrès en compétence, comme motivation intrinsèque pour choisir les différents aspects de sa stratégie d'apprentissage.

Ce cadre de la motivation intrinsèque guidée socialement pourrait être utilisée pour les systèmes de tutorat intelligent, pour être appliquée en robotique d'assistance sociale, pour un robot coach pour la rééducation physique pour des patients de lombalgie qui analyse les mouvements corporels, ou pour un robot coach pour des enfants ayant un TSA dans des jeux d'imitation.

Le projet que j'aimerais développer en apprentissage cognitif développemental se rapporte à la fois au cadre de l'apprentissage par renforcement en apprentissage automatique, pour l'apprentissage du contrôle dans une boucle action-perception de la robotique cognitive, et de la théorie de la motivation intrinsèque qui provient de la psychologie développementale in science cognitive. L'objectif est de proposer de nouveaux algorithmes d'apprentissage par renforcement pour permettre aux robots d'apprendre en interagissant de manière active dans leur environnement, en permettant aux robots de maîtriser des tâches de contrôle de complexité croissante, tout en développant la théorie actuelle de la motivation intrinsèque pour un modèle unifié des motivations à interagir avec des tuteurs. Le projet s'appuie sur deux axes de recherche : un modèle d'interaction avec des tuteurs reposant sur une représentation émergente des tâches; et un modèle de la motivation des apprenants humains à travers le cas d'un robot coach pour la rééducation

physique. Ce travail va enrichir la formulation de la motivation pour le guidage social avec différentes impulsions motivationnelles des apprenants pour interagir avec des tuteurs, en adoptant le point de vue de la communication et des émotions. Ceci va aussi mettre en lumière l'importance de représentations internes émergentes pour la compréhension du guidage social, ainsi souligner l'interdépendance entre apprentissage automatique et apprentissage social.

Mots clés : Apprentissage par renforcement interactif, Apprentissage par imitation actif, Guidage social, Motivation Intrinsèque, Apprentissage par renforcement, Apprentissage hiérarchique, Analyse de mouvement humain, Activités de la vie quotidienne, Reconnaissance d'activité

Contents

| | | |
|----------|--|-----------|
| 1 | APPROACH AND STATE OF THE ART | 2 |
| 1.1 | Developmental Robotics | 2 |
| 1.2 | Reinforcement Learning and Imitation Learning | 4 |
| 1.3 | Framework | 6 |
| 1.3.1 | Formalising the learning problem | 6 |
| 1.3.2 | Formalising the imitation process | 7 |
| 1.4 | Active Imitation Learning | 7 |
| 1.4.1 | Optimising What, How, When and Who to imitate | 8 |
| 1.4.2 | The Reward Hypothesis of Social Interaction | 11 |
| 2 | INTRINSIC MOTIVATION TO LEARN COMPOSITIONAL TASKS | 13 |
| 2.1 | Multi-Arm Bandit with Intrinsic Motivation | 14 |
| 2.1.1 | Formalisation | 15 |
| 2.1.2 | Algorithmic Architecture | 15 |
| 2.1.3 | Learn Task Hierarchy from a Static Set of Tasks | 17 |
| 2.1.4 | Symbolic Representations of Tasks | 18 |
| 2.1.5 | Learn Task Hierarchy from a Dynamic Set of Tasks : Adaptive learning of an affordance hierarchy through exploration | 18 |
| 2.2 | STAR : Goal-Conditioned Hierarchical Learning | 19 |
| 2.3 | Representations used for the Analysis of Human Activity of Daily Living | 22 |
| 3 | INTRINSIC MOTIVATION TO LEARN WITH TEACHERS | 24 |
| 3.1 | Learning a Parametrised Task | 24 |
| 3.2 | Influence of demonstrations | 27 |
| 3.3 | Robustness to the quality of demonstrations | 27 |
| 3.4 | Compositional Tasks with Active Imitation Learning | 28 |
| 3.5 | Socially Guided Intrinsic Motivation | 30 |
| 4 | APPLICATION TO SOCIALLY ASSISTIVE ROBOTICS | 33 |
| 4.1 | A Robot Coach Proposing an Automatic Curriculum | 33 |
| 4.2 | Robot Coach for Physical Rehabilitation | 34 |
| 4.2.1 | Humanoid Socially Assistive Robot in a Clinical Trial | 34 |

| | | |
|----------|--|-----------|
| 4.2.2 | Four Challenges for a Dataset | 35 |
| 4.2.3 | Personalisation of the coaching to different morphologies | 35 |
| 4.3 | Imitation Game for ASD children | 36 |
| 5 | THE MOTIVATION TO INTERACT ACTIVELY WITH TUTORS | 38 |
| 5.1 | Research problem | 39 |
| 5.2 | Representations of tasks to enable social learning | 40 |
| 5.2.1 | Challenges | 40 |
| 5.2.2 | Learning a symbolic representation | 40 |
| 5.3 | Different factors of the motivational drive of learning agents to interact with tutors | 41 |
| 5.4 | Conclusion | 42 |
| 6 | SELECTED ARTICLES | 43 |
| 7 | PUBLICATION LIST OF NGUYEN SAO MAI | 44 |
| 7.1 | Book chapter | 44 |
| 7.2 | Thesis | 44 |
| 7.3 | Editorials | 45 |
| 7.4 | Peer Reviewed Journals | 45 |
| 7.5 | Conferences with Proceedings | 46 |
| 7.6 | International Conferences with Peer Review | 50 |

1

Approach and State of the Art

My research has been investigating in developmental cognitive robotics [Cangelosi and Schlesinger, 2015; Asada et al., 2009; Lungarella et al., 2003], the motivations of learning agents to interact with teachers, with the outlook of a fully autonomous robot that must continuously learn new tasks and how to perform more complex tasks relying on simpler ones, so as to adapt to the human daily environment and user's needs that change constantly. The paradigm of a permanent learning process is called *lifelong learning* or *continual learning* [Ring, 1994].

More specifically, learning in an unbounded environment a non-finite number of tasks is referred to as *open-ended learning* [Doncieux et al., 2018]. Open-ended learning is characterised by multi-task learning, where the set of tasks is not known in advance, including interrelated tasks where easy tasks can be composed into more complex tasks. Thus state, action and task spaces are a priori unbounded in time, space and complexity, thus can be of infinite dimensionality. The **challenges of open-ended learning are that (1) the state and action spaces are continuous and high-dimensional, even of infinite dimensionality, and (2) the mappings to learn can be stochastic and redundant.**

In this chapter, I will describe the developmental learning approach, the challenge and outlook taken by our works, and the relations of our works in active imitation learning with the state of the art.

1.1 DEVELOPMENTAL ROBOTICS

Uncovering the mechanism babies acquire new skills has been proposed as a key to enable artificial intelligence agents, since Alan Turing suggested :

Instead of trying to produce a program to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education, one would obtain the adult brain [...] Our hope is that there is so little mechanism in the child brain that something like it can be easily programmed. The amount of work in the education we can assume, as a first approximation, to be much the same as for the human child.

(Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, LIX(236):433–46).

Taking Turing's suggestion, *developmental robotics*, first introduced in [Asada et al., 2001], proposes an "interdisciplinary approach to the autonomous design of behavioral and cognitive capabilities in artificial agents (robots) that takes direct inspiration from the developmental principles and mechanisms observed in the natural cognitive systems of children", as phrased in [Cangelosi and Schlesinger, 2015]. The main goal is to uncover the principles regulating the acquisition process of an increasingly complex set of sensorimotor and cognitive skills, through real-time interaction between its brain, its body and its environment [Cangelosi and Asada, 2022, Chapter 3]. This online, open-ended, cumulative learning process is generally viewed as a progressive evolution of abilities and skills, akin to the developmental stages of infants as described by Piaget [1952], which principle derived into the machine learning fields of *curriculum learning* [Bengio et al., 2009].

Part of the highly interdisciplinary field of cognitive science which includes developmental psychology, neuroscience, robotics, linguistics and computer science, developmental robotics is centered around two key concepts:

- **enactivism**, introduced by Varela et al. [1991], hypothesises that cognition is based on situated, embodied agents. It is a theoretical approach to understanding the mind which emphasises the way that organisms and human mind organise themselves by interacting with the environment. Enactivism thus uses the notion of *embodiment* [Brooks, 1991; Pfeifer and Scheier, 1999] and action for cognition, which hypothesises that the mind is largely determined by the form of the organism's body and the actions it can do. The embodied cognition is grounded on self-experience. The way we conceptualise and reason depends on "the kinds of bodies we have, the kinds of environments we inhabit, and the symbolic systems we inherit, which are themselves grounded in our embodiment" [Johnson, 1987]. The mind builds from the personal history of each agent of his sensorimotor perception, and is not mere manipulation or operations of abstractions. This translates Piaget's theory emphasis on the sensorimotor bases of mental development and on the environmental approach balancing the biological approach. Drawing a parallel with machine learning, this learning paradigm is the closest to reinforcement learning, where the agent learns a policy by interacting with its environment to learn online from reward signals while discovering its environment.
- the **social learning theory** proposes that new behaviors can be acquired by direct instructions or observing and imitating others [Bandura, 1971]. The learning benefits from the observation of behavior, but also the observation of rewards and punishments. This learning paradigm relies on data induced by a teacher. Vygotsky's sociocultural theory [Vygotsky, 1978] describes the social environment's role as important as the physical environment on the scaffolding of the child's cognitive system owing to the role of adults and peers in guiding the development of the child. Educational psychology has outlined the importance of social learning to bootstrap autonomous learning, and put

forward the idea of *social scaffolding* : children explore and learn on their own, but in the presence of a teacher, they can take advantage of the observations of social cues and the teacher’s behavior to accomplish more. A teacher often guides a learner by providing timely feedback or instructions, luring them to perform desired behaviours, and controlling the environment to simplify the analysis of the state, for instance by highlighting the appropriate cues to make them salient. [Vygotsky \[1978\]](#) defines the concept of *zone of proximal development* as the space between what a learner is capable of doing autonomously and what the learner cannot do even with support :

[The zone of proximal development] is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers.

[Vygotsky \[1934\]](#)

Backed by findings on mirror neurons in the brain [[Rizzolatti et al., 2008](#)] that suggests a natural predisposition for imitation, computational models of imitation have been developed such as in [[Oztop et al., 2006](#)]. More broadly, social learning has contributed to human-robot imitation studies such as imitation, gesture, vocalization, joint attention, turn taking, etc. In machine learning, this paradigm translates in its simplest form as supervised learning, with a wealth of algorithms named *imitation learning* [[Nehaniv and Dautenhahn, 2007](#); [Schaal, 1999](#)], programming by demonstration [[Calinon, 2009](#); [Billard et al., 2007](#)] or learning from demonstration [[Argall et al., 2009](#)].

Our works falls within the developmental robotics approach, to propose machine learning algorithms, with a focus on motor control and action learning.

1.2 REINFORCEMENT LEARNING AND IMITATION LEARNING

From the machine learning perspective too, these two key concepts seem to shape two families of action learning algorithms. They can be seen as pertaining to the trial-and-error paradigm, where the learning agent collects data by exploring its physical environment autonomously and observing the outcomes of its actions, or to the imitation learning, where the learning relies on data provided by a tutor.

While imitation learning has been investigated to learn a policy for an Markov Decision Process given expert demonstrations, for a long time in robotics [[Schaal et al., 2003](#); [Billard and Siegward, 2004](#)], socially guided learning has been strongly relying on the involvement of the human user or the demonstration dataset. However, the more dependent on the human, the more challenging learning from interactions with a human is, due to limitations of the provided demonstration dataset because of lack of human patience, attention, memory, or the sparsity of teaching datasets, the absence of teaching for some subspaces, ambiguous and suboptimal human input, correspondence problems, etc, as highlighted in [[Nehaniv and Dautenhahn, 2007](#)]. This is one of the reasons why in most approaches to robot learning of motor skills, either in supervised learning or inverse reinforcement learning, only a few movements or motor policies were learnt in any single studies. More recently, behavioural cloning has been implemented using deep neural networks, as reviewed in [[Mandlekar et al., 2021](#)], but show the same limitations and a reliance on a higher amount of demonstration data. In particular, imitation learning has been performing poorly for

multi-task learning, including hierarchical tasks. A limitation is that as the number of tasks increases, the number of data needed and the required quality of data from tutors increase, thus increasing the cost of human tutoring. Robot learning algorithms need to decrease the cost of human guidance, by both leveraging demonstrations on simple tasks to learn more complex tasks, and by reducing the cost for human tutors of the interactions. **Increasing the learner’s autonomy from provided demonstrations could address these limitations.**

On the other side, the trial-and-error strategy has been implemented in computational models as reinforcement learning algorithms [Sutton and Barto, 1998] to infer the optimal behaviour, which is also referred to as the *policy*. To improve the efficiency of exploration, methods reusing some of the concepts elaborated in the statistical active learning framework [Fedorov, 1972; Cohn et al., 1996; Roy and McCallum, 2001] have recently been developed in the fields of developmental learning using a heuristics for its active choice [Oudeyer and Kaplan, 2009; Schmidhuber, 2010; Gottlieb et al., 2013]. This heuristics is inspired by *intrinsic motivation* in psychology. Intrinsic motivation was described in [White, 1959] : “While the purpose is not known to animal or child, an intrinsic need to deal with the environment seems to exist and satisfaction (the feeling of efficacy) is derived from it.” Intrinsic motivations triggers spontaneous exploration and curiosity in humans.

“Intrinsic motivation is defined as the doing of an activity for its inherent satisfaction rather than for some separable consequence. When intrinsically motivated, a person is moved to act for the fun or challenge entailed rather than because of external products, pressures or reward” [Ryan and Deci, 2000]

Intrinsic motivation is to be contrasted with extrinsic motivation, which is “a construct that pertains whenever an activity is done in order to attain some separable outcome. Extrinsic motivation thus contrasts with intrinsic motivation, which refers to doing an activity simply for the enjoyment of the activity itself, rather than its instrumental value” [Ryan and Deci, 2000].

Still, the current reinforcement learning algorithms fail to transpose to robots because: (1) they require the robot to execute a large number of real actions to sample the continuous high-dimensional space of states and actions, hence requiring a lot of time and are not scalable for robots; (2) they cannot leverage the learned tasks to compose more complex tasks for open-ended learning, as the more sequential the considered tasks, the higher dimensionality the state-action spaces are. Therefore, **complementary developmental mechanisms need to constrain the growth of the size and complexity of the exploration areas and structure the environment and the learning curriculum**, by guiding them toward learnable subspaces and away from unlearnable subspaces.

While reinforcement learning and socially guided learning have so far often been studied separately in developmental robotics and robot learning literature, we believe their integration has high potential. Their combination could push the respective limits of each family of exploration mechanisms. I argue that **social guidance, leveraging knowledge and skills of others, can be key for bootstrapping the intrinsically motivated learning of such models.**

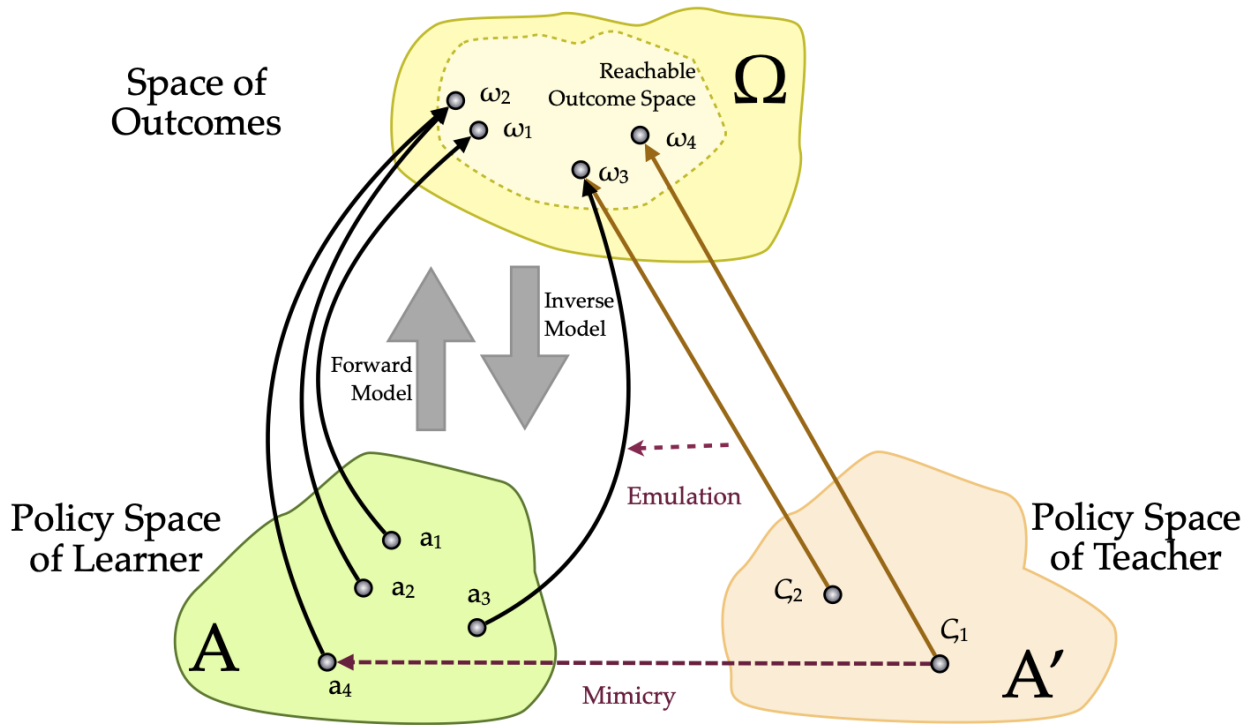


Figure 1.3.1: The goal is to learn a mapping from the policy space A of the learner and the outcome space Ω . The mapping from the policy space to the outcome space is the forward model, used for predicting the effects of an action. The mapping from the outcome space to the policy space is the inverse model, used to choose the appropriate action to reach the goal. Emulation and mimicry are two types of imitation learning. In emulation, the learner tries to reproduce the outcome ω_3 demonstrated by the teacher without trying to reproduce the teacher's movement ζ_2 , but uses its own movement a_3 . In mimicry, the learner reproduces the demonstrated movement ζ_1 with policy parameters a_4 , without trying to reproduce the demonstrated outcome ω_4 .

1.3 FRAMEWORK

Let us outline a framework for learning open-ended problems from reinforcement learning and imitation learning.

1.3.1 FORMALISING THE LEARNING PROBLEM

Let us consider a robot interacting with a non-rewarding environment by performing sequences of motions of unbounded length in order to induce changes in its surroundings.

Each of these motions is named a *primitive action*, described by a parametrised function with p parameters: $a \in \mathcal{A} \subset \mathbb{R}^p$. We call the continuous multidimensional space \mathcal{A} the *primitive action space*. Our robot can perform sequences of primitive actions.

The actions performed by the robot have consequences on its environment, which we call outcomes $\omega \in \Omega$,

where Ω is a subspace of the state space S defining the control tasks to learn. Once the robot knows how to cause an outcome ω , we say the outcome is then *controllable*. The set of controllable outcomes is $\Omega_{cont} \subset \Omega$ and this set changes as the robot learns new tasks. For convenience, we define the *controllable space* $\mathcal{C} = \mathcal{A} \cup \Omega_{cont}$, regrouping both primitive actions \mathcal{A} and observables that may be controlled, Ω_{cont} .

The robot learns tasks/models T , each mapping a controllable $c \in \mathcal{C}_T \subset \mathcal{C}$ to an outcome $\omega \in \Omega_T \subset \Omega$ within a given context $s \in S_T \subset S$. More formally, a task is a set of: a **forward model** $M_T : (S_T, \mathcal{C}_T) \rightarrow \Omega_T$ and an **inverse model** $L_T : (S_T, \Omega_T) \rightarrow \mathcal{C}_T$. The forward model is used to predict the observable consequence ω from a given context s of a controllable c , which is either a primitive action or a goal state that can be induced by a compound action. Conversely, the inverse model is used to estimate a controllable \tilde{c} to be performed in a given context s to induce a goal observable state $\tilde{\omega}$: if \tilde{c} is a primitive action the robot executes the action, otherwise it sets \tilde{c} as a goal state and infers the necessary actions using other inverse models.

These models are trained on the data recorded by the robot along its exploration in its dataset \mathcal{D} . We define a strategy σ any process for exploration. For instance, we consider autonomous exploration and imitation learning strategies. Formally, we define it as a data collection heuristics, based on the current data, that outputs a set of triplets of initial state, action and outcome: $\sigma : \mathcal{D} \mapsto \{(s, a, \omega)\}$.

1.3.2 FORMALISING THE IMITATION PROCESS

Imitation learning can be separated into two broad categories of social learning [Call and Carpenter, 2002], as represented in fig. 1.3.1 : the mimicry strategy where the learner copies the observed trajectory ζ_1 with policy parameters a_4 , and emulation strategy where the learner attempts to replicate goals $\omega \in \Omega$.

Several teachers generate trajectories according to their policies π_k , which are a priori unknown to the learner. A priori, the action space of the teacher is not equal to the action space of the learner. We suppose that π_k depends on the current state, the goal, but can also be influenced by the query of the learner. In the next sections, we discuss the query process from a learner to a tutor.

1.4 ACTIVE IMITATION LEARNING

While reinforcement learning and imitation learning have been traditionally opposed – the former being typically a data collection algorithm through exploration of the environment, and the latter typically using a preset dataset of labelled data – we show that both worlds are on the contrary complementary and highlight the merits of merging the two fields. While an increasing stream of machine learning works propose to combine the two paradigms [Price and Boutilier, 1999; Ng and Russell, 2000], and has been named *interactive reinforcement learning* Thomaz et al. [2005]; Arzate Cruz and Igarashi [2020]. This stream of work includes reinforcement learning from human feedback [Thomaz and Breazeal, 2008; Knox et al., 2013; Christiano et al., 2017] including for large language models [Ouyang et al., 2022]. In most interactive reinforcement learning works, the agent in these works undergoes the interaction passively. We show how a learning agent could optimise its interaction with tutors.

Research in developmental psychology indicates that social interaction is not only for social pleasure but can serve for learning and exploration of the environment. While most theories of infant social learning focus on how infants learn whatever and whenever the adults decide to teach them, recent findings suggest

that social learning is not a passive process but that infants play an active role in collecting information and adapting their learning strategy according to their interests. Indeed, infants show a preference to learn from reliable people [Poulin-Dubois et al., 2011], and their attention towards adults is influenced by the adult’s role in bringing new information [Begus et al., 2016]. Infants show curiosity and active contribution to social transmission of knowledge [Begus and Southgate, 2018]. Thus, we propose to frame the interaction with human teachers as a reinforcement learning problem. Reinforcement learning is therefore not only about an agent interacting with a physical environment but also with a social environment. Could the reinforcement framework be extended to all aspects of the interaction with teachers?

In the next section, we review the literature describing on how a learning agent devises its learning strategy between the two learning paradigms and aspects of its interaction with its tutors. We will refer the approaches combining autonomous exploration and social guidance in an active way as *active imitation learning*, a term defined in [Shon et al., 2007], as an ‘imitation learning paradigm covering cases where an observer can influence the frequency and the value of demonstrations that it is shown’. This corresponds to a more general definition than the *ask for help* approach where ‘The learner is completely responsible for its interaction with the trainer, asking the trainer to provide actions. The trainer will always give an action when asked’ [Clouse, 1996]. Later on, Judah et al. [2014] defined used active imitation learning, in a more restricted case where ‘the learner asks queries about specific states, which the expert labels with the correct actions. The goal is to learn a policy that is nearly as good as the expert’s policy using as few queries as possible’.

1.4.1 OPTIMISING WHAT, HOW, WHEN AND WHO TO IMITATE

In order to define the social interaction that we wish to consider, let us characterise the different possibilities of information flow as reviewed in [Argall et al., 2009; Billard et al., 2007; Schaal et al., 2003; Lopes et al., 2009] with respect to: what, how, when and who to imitate. This categorisation has been introduced in [Dautenhahn and Nehaniv, 2002; Breazeal and Scassellati, 2002]. In this section, we examine the related works studying the closed-loop phenomenon of a learner making queries to a tutor that influence what data are added to its training set. Compared to the review by Zhang et al. [2019] which focuses on the deep reinforcement learning and analyses the state of the art with respect to the types of learning algorithm, we propose to analyse in this section the existing works under the human-robot interaction perspective. As opposed to the more general review in [Ravichandar et al., 2020], we will focus on active choices by the learner on its interaction with teachers, and contrast our works to the state of the art.

Please note that the current review does not aim at friendly human-robot interaction, the social rules for a comfortable and natural interaction or goal understanding and intentionality.

In this section, we focus on the information from the tutor to the learning agent and its efficiency to convey content by human-robot interaction. Thus we examine what information is given by the tutor.

- What to imitate and query? Let us examine the target of the information given by the teacher, or mathematically speaking, the space on which he operates. This can be either the policy, context or outcome spaces, or combinations of them. Merging autonomous exploration with socially guided exploration can optimise the number of data needed for multi-task learning by reusing knowledge across tasks by interpolating for parameterised tasks or using hindsight replay [Nguyen and Oudeyer,

| Algorithm | Environ. Tasks | Imitation | | | Query | | | |
|---------------------------------------|--|---------------------------------------|--------------------------|----------------------|---------------------|---|--------------------------------------|--|
| | | What | How | When | Who | What | When | Who |
| AfH [Clouse, 1996] | Single | Policy | Low-level | Learner's initiative | Unhelpful | Difference lowest Q values | NA | NA |
| [Shon et al., 2007] | Single | Discrete States | Low-level | Learner's Initiative | Unhelpful, several | NA | Value of demonstration | Policy regret |
| CBA [Chernova and Veloso, 2009] | Single | Discrete Policy | Corrective demonstration | Learner's initiative | Helpful, one | Current state | Confidence on state | NA |
| Dagger [Ross et al., 2010] | Single | Policy | Low-level | After each episode | Helpful | Action for the visited states | NA | NA |
| FPS [Argall et al., 2011] | Continuous Tasks of different complexities | Trajectory segment | Advice, Reward | Always | One, helpful | Feedback on visited trajectories | NA | NA |
| CLIC [Fournier et al., 2019] | Multiple discrete | discrete Task, policy | Low-level | Fixed frequency | Unhelpful | Task with highest learning progress | NA | NA |
| [Kulak and Calimon, 2023] | Single continuous | Continuous policy | NA | Learner's initiative | Helpful, one | Goal based on epistemic uncertainty | expected epistemic entropy reduction | NA |
| RPI [Liu et al., 2024] | Single | Continuous policy | Low-Level | Learner's initiative | Several | Action for random state in trajectory | Highest Value | Highest Value |
| SGIM-ACTS [Nguyen and Oudeyer, 2012d] | Multiple continuous | Continuous policy, Continuous outcome | Low-level | Learner's initiative | Unhelpful, multiple | Action for the outcome with highest competence progress | Competence progress | Tutor enabling highest competence progress |
| SGIM-PB [Duminy et al., 2021] | Continuous tasks of different complexities | Policy, Outcome, task decomp | Low-level | Learner's initiative | helpful, multiple | Action for the outcome with highest competence progress | Competence progress | Tutor enabling highest competence progress |

Table 1.4.1: Comparison of the active imitation learning algorithms: aspects of the imitation and of the interaction the learner actively decides on.

2012d], and by querying and combining different types of demonstrations such as policy demonstration, goal demonstration and task decomposition [Nguyen and Oudeyer, 2014; Duminy et al., 2019]. The learner can query demonstrations for all states visited in the previous episode [Ross et al., 2010], or for a state with a low difference between the highest and the lowest Q values Clouse [1996]. In [Fournier et al., 2019] and in our works [Nguyen et al., 2011a; Duminy et al., 2018b], the learner queries a demonstration for the task with highest competence progress.

- How to imitate ? Whichever the target, the information can be communicated from the teacher to the learner through low level (as in our algorithm SGIM-ACTS [Nguyen and Oudeyer, 2014]) or high level demonstrations, through advice, reward or labelling. A language protocol often enables instructions of policies [Argall et al., 2011]. Language become a cognitive tool to support efficient acquisition of open-ended repertoires of skills as described in [Colas et al., 2019].
- When to imitate ? The timing of the interaction varies with respect to its timing within an episode, and with respect to its general activity during the whole learning process. Demonstration data can be used at the beginning of the learning process to initialise the model, at a fixed or a random frequency, at the teacher’s initiative or at the learner’s initiative. When at the learner’s initiative, it optimises the interaction with human supervisors, so as to minimise the cost of human guidance in terms of number of demonstrations. The learner query teachers’ guidance based on criteria such as uncertainty Chernova and Veloso [2009] and in particular epistemic uncertainty Kulak and Calinon [2023]. Our works used a criteria competence progress to select the best strategy between autonomous exploration and the imitation strategies [Nguyen and Oudeyer, 2012d,c; Duminy et al., 2019],
- Who: Being able to request help to different experts is also an efficient way to address the problem of the reliability of the teacher. Imitation learning studies often rely on the quality of the demonstrations, whereas in reality a teacher can be performant for some outcomes but not for others. Demonstrations can be ambiguous, unsuccessful or suboptimal in certain areas. Like students who learn from different teachers who are experts in the different topics of a curriculum, a robot learner should be able to determine its best teacher for the different outcomes it wants to achieve. To our knowledge, Shon et al. [2007] were the first to propose a framework to enable the learning agent to decide who to imitate, using a side payment. Liu et al. [2024] proposed a criteria by comparing the value functions of each teacher’s policy. Our works [Nguyen and Oudeyer, 2012d; Duminy et al., 2019, 2021] make the most of several teachers by choosing the expert for each query based on the competence progress each teacher has enabled. Learners also minimise the cost of human guidance by considering their availability through a cost parameter to choose whom to ask guidance to. These methods are robust to imperfect teachers Shon et al. [2007]; Fournier et al. [2019]; Nguyen et al. [2013]; Nguyen and Oudeyer [2012a]; Duminy et al. [2016].

A synthesis of how the state-of-the-art tackles these aspects of active imitation learning is presented in Table 1.4.1.

While learning single tasks can be addressed by naive methods, continuous task spaces require more efficient reinforcement learning algorithms or models. Further, state-of-the-art methods are the most profitable for multi-task learning where tasks of different nature constitute composite continuous task spaces. In particular, hierarchical/sequential task learning can combine transfer knowledge from simple tasks to complex tasks after

identifying through learning the relationships between tasks [Duminy et al., 2021], thus entailing automatic curriculum learning that self-schedules the learning process with simple tasks before the agent starts to master more complex/sequential tasks.

1.4.2 THE REWARD HYPOTHESIS OF SOCIAL INTERACTION

Yet, no algorithms have yet proposed a comprehensive framework to optimise all types of social interaction. Most works have focused on addressing what and how to interact. Only a few have addressed when and who to imitate. A few works such as [Nicolescu and Mataric, 2003; Thomaz, 2006] have combined several types of messages such as policy demonstrations, highlight on the context elements, corrective and reward feedback, but they rely on the teacher to optimise the interaction. The learning agent does not actively choose the aspects of this interaction.

In active imitation learning, intrinsic motivation based algorithms such as CLIC [Fournier et al., 2019] or SGIM [Nguyen et al., 2021] have proposed a common criteria to optimize what, how, when and who to imitate: the learning progress. These optimisations rely on empirical estimations of the impact of additional data on the learning progress for each source of information, and on the active selection of the information to be queried. The algorithms also optimise the number of data needed for multi-task learning by reusing knowledge across tasks and by querying and combining different types of information, after identifying through learning the relationships between tasks. The learning progress measure can be used to merge autonomous exploration with socially guided exploration, select the best among several teachers, and be robust to imperfect teachers. In particular, choosing aspects of the interaction with the teachers, such as who to imitate and when to imitate can be optimised with, as criteria, the learning progress a teacher can bring to the learner, weighted by the cost of querying a demonstration to the teacher. This criteria can thus be viewed as a reward to interact with teachers.

While reinforcement learning and social interaction seem to live in different frameworks, our review shows that they can be combined. However a common framework is still needed. In reinforcement learning, Sutton and Barto [2018] posited the reward hypothesis as :

“all of what we mean by goals and purposes can be well thought of as maximization of the expected value of the cumulative sum of a received scalar signal (called reward).” [Sutton and Barto, 2018]

However, this reward hypothesis has only been tested in autonomous learning settings, and has not been extended to settings where the learning agent interacts with teachers, who may be perfect or imperfect. How can a learning agent learn to choose right interaction actions with teachers ? **Which reward function should shape the learning agent’s interactive behaviour ?**

While Human-Computer Interaction and Affective computing research postulate that the drive for interaction with others comes from emotions, thus emotional signals can act as an external reward signals, we take **the outlook of interaction as an active learning process to gain information, and the others as sources of information**, which probability distribution the learner has to model through sampling. Thus, we strive to **model the intrinsic motivation of interacting with teachers**.

Recently, some works have proposed a universal reward function for reinforcement learning of social interaction : the learning progress. In [Fournier et al., 2019; Nguyen et al., 2021], the same formulation of the

intrinsic motivation is valid both for autonomous exploration and for social guidance, may the demonstrations requested to tutors be low-level policies, goals or decomposition into subgoals. Thus, a common framework is possible by **using intrinsic motivation as the common criteria to choose each learning strategy**. Our works shows that by considering that **each choice of teacher and type of demonstration is a learning strategy, the common criteria to choose a learning strategy σ and task ω can be formulated as $im(\sigma, \omega) = \kappa(\sigma) * progress(\sigma, \omega)$** where *progress* is the empirical progress measured through the last episodes with strategy σ and goal ω , and $\kappa(\sigma)$ is the cost of the strategy, representing the availability of teachers, their willingness to interact with the robot ...

At the convergence of active imitation learning and open-ended learning in non-rewarding environments, we describe in the next chapter our multi-task learning algorithms for multiple parametrised tasks and for hierarchical tasks, using intrinsic motivation. In chapter 3, we detail our active imitation learning algorithms. In chapter 4, we describe application projects in socially assistive robotics and assistive technologies. Lastly, chapter 5 sketches future research directions.

2

Intrinsic Motivation to Learn Compositional Tasks

In this chapter, we describe, for open-ended learning in environments without external rewards, our works to learn with active imitation learning without external rewards from the environment or the teachers, and to solve long-horizon tasks such as compositional tasks, also referred to as sequential tasks. While most works on multi-task learning focus on parametrised tasks or tasks of the same level of complexity, my team examined how to learn *compositional tasks*, ie tasks can be considered as composed of simpler tasks.

Indeed, humans, and other animals, cannot only learn simple behaviors and generalise them across various states, but they can also combine simple behaviors to create more complex behaviors. This idea of decomposition and composition of actions into sequences of reusable primitives has been formalised as the motor schemata theory by [Arbib \[1981\]](#). The theory considers that a set of action primitives might be memorised, to be retrieved and combined by the higher level to generate desired actions. The ability to compositionally combine behaviors is thought to be central to generalized intelligence in humans and a necessary component for artificial intelligent systems. The ability to learn a variety of compositional, long-horizon skills while generalizing to novel concepts remains an open challenge. Long-horizon tasks demand sophisticated exploration strategies and structured reasoning, while generalization requires suitable representations.

Applying *Reinforcement Learning* (RL) [[Sutton and Barto, 1998](#)] to the complex environments commonly found in robotics is often challenging, due to the high-dimensional and continuous state space and sparse rewards for temporally extended tasks, especially for *open-ended learning* [[Doncieux et al., 2018](#)].

To solve open-ended learning, several successive representational redescription processes through

| Algorithm | Strategies σ | Tasks set | Composite Actions |
|-----------|---|------------------------------|-------------------|
| IM-PB | Action exploration, Outcome space exploration, Task decomposition exploration | Static set | Subgoals sequence |
| CHIME | Action exploration, outcome space exploration | Dynamic set (emerging tasks) | Planning |
| SGIM-PB | Action exploration, outcome space exploration, Task decomposition exploration; active imitation | Static set | Subgoals sequence |

Table 2.1.1: Differences between the 3 implementations of SGIM-SAHT to learn compositional tasks

unsupervised acquisition of a hierarchy of adapted representations is suggested in [Doncieux et al., 2018] to be a key component. Open-ended learning should be solved by an unsupervised acquisition of a hierarchy of adapted representations. To tackle learning multiple tasks of complexities unknown a-priori, the robot needs to exploit the knowledge from simple tasks to compose tasks of growing difficulty. A hierarchical description of actions has been proposed in neuroscience (eg [Grafton and de C. Hamilton, 2007]) and in behavioural psychology (eg. [Eckstein and Collins, 2021]). This idea has been transcribed in machine learning : Hierarchical Reinforcement Learning (HRL) [Barto and Mahadevan, 2003] is a recent approach for learning to solve long and complex tasks by decomposing them into simpler subtasks. HRL could be regarded as an extension of the standard Reinforcement Learning (RL) setting as it features high-level agents selecting subtasks to perform and low-level agents learning actions or policies to achieve them. Hierarchical Reinforcement Learning tackles the environment’s complexity by introducing a hierarchical structure of agents that work at different levels of temporal and behavioral abstractions : lower-level agents learn to solve easier tasks and higher-level agents learn to solve more complex tasks, using the low-level agents’ policies. However, the reinforcement learning algorithms still need millions of learning samples to learn and are not scalable for robot learning.

In this section, we hypothesize that one critical element for solving such problems is the notion of compositionality. Hierarchical RL (HRL) has designed this compositionality as an ability to learn concepts and sub-skills that can be composed to solve longer tasks. However, acquiring effective yet general abstractions for hierarchical RL is remarkably challenging. In this chapter, I only examine autonomous exploration paradigms, and not the imitation paradigm.

2.1 MULTI-ARM BANDIT WITH INTRINSIC MOTIVATION

My team examines how an agent can learn both the policies for simple and sequential tasks and **the relationship between tasks for the purpose of transferring knowledge from simpler tasks into sequential tasks** and proposes an architectural architecture, Socially Guided Intrinsic Motivation for Sequence of Actions through Hierarchical Tasks (SGIM-SAHT), common to the implementations of algorithms IM-PB, CHIME, SGIM-PB. Table 2.1.1 outlines the differences between them.

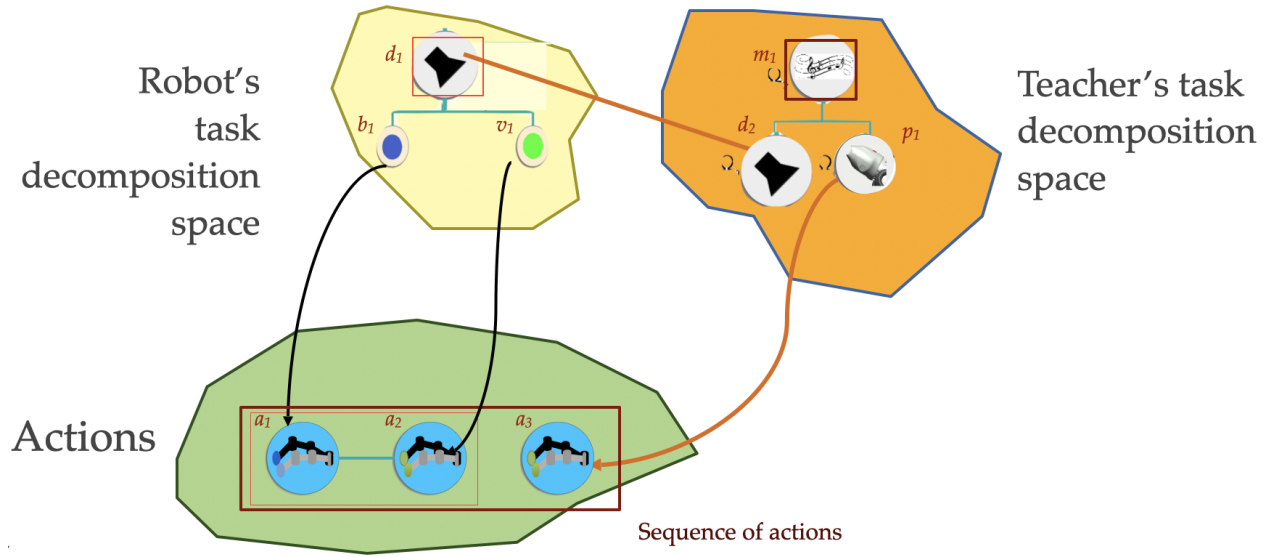


Figure 2.1.1: Our goal-oriented representation of sequential tasks decomposes a task into goals of subtasks. Each subtask can then recursively be decomposed into subtasks, to be in fine executed by a policy : m_1 is decomposed into (d_2, p_1) . d_2 is the decomposed into (b_2, v_2) Thus a sequential task is in fine realised by a sequence of policies : m_1 is realised by the sequence (a_1, a_2, a_3) . When learning with a teacher, the learner can ask for a demonstration of low-level policy (a_3 realises p_1) or task decomposition (m_1 is decomposed into subtasks (d_2, p_1)).

Our study relies on representations of task decomposition as illustrated in fig. 2.1.1.

2.1.1 FORMALISATION

We extend the formalisation proposed in Section 1.3.1 for sequential tasks.

Our robot can perform sequences of primitive actions. Let a *compound action* be a sequence of any length $n \in \mathbb{N}$ primitive actions, and be described by $n * p$ parameters : $a = [a_1, \dots, a_n] \in \mathcal{A}^n$. Thus the action space exploitable by the robot is a continuous space of infinite dimensionality $\mathcal{A}^{\mathbb{N}} \subset \mathbb{R}^{\mathbb{N}}$.

Let us note \mathcal{H} the hierarchy of the models used by our robot. \mathcal{H} is formally defined as a directed graph where each node is a task T and its successors are the components of \mathcal{C}_T . As our robot learns this hierarchy, \mathcal{H} varies along time.

2.1.2 ALGORITHMIC ARCHITECTURE

Socially Guided Intrinsic Motivation for Sequence of Actions through Hierarchical Tasks (SGIM-SAHT), detailed in Algo. 2.1.1 and outlined in Fig. 2.1.2, learns by episodes in which a task T to work on, a goal outcome $\omega_g \in \Omega_T$ and a strategy σ have been selected to optimize progress and according to an interest map (see below). The selected strategy σ applied to the chosen goal outcome ω_g chooses a sequence of controllables $l_c = [c_1, \dots, c_m]$ as a candidate to reach the ω_g (Alg.2.1.1, 1.3).

Algorithm 2.1.1 SGIM-SAHT

Require: the different strategies $\sigma_1, \dots, \sigma_n$
Require: the initial model hierarchy \mathcal{H}
Ensure: partition of outcome spaces $R \leftarrow \bigsqcup_i \{\Omega_i\}$
Ensure: episodic memory $Memory \leftarrow \emptyset$

 1: **loop**

 2: $\sigma, T, \omega_g \leftarrow$ Select Strategy, Task & Goal Outcome(R, \mathcal{H})

 3: $l_c \leftarrow$ Apply Strategy(σ, ω_g)

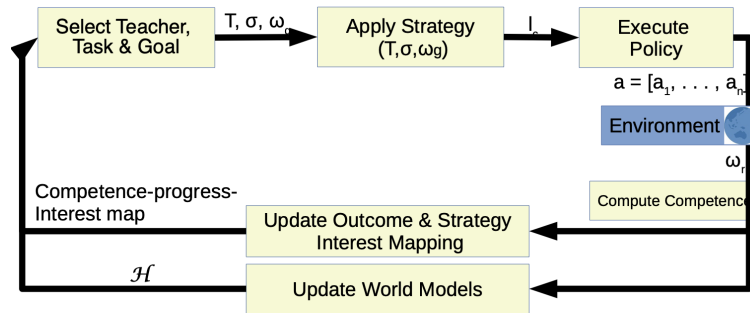
 4: $\mathcal{D} \leftarrow (\omega_r, a, l_c) \leftarrow$ Execute Sequence(l_c)

 5: $(comp(\omega_g), comp(\omega_r)) \leftarrow$ Compute Competence(ω_g, ω_r)

 6: Update M_T, L_T, \mathcal{H} with $(\mathcal{D}, comp(\omega_g), comp(\omega_r))$

 7: $R_i \leftarrow$ Update Outcome and Strategy Interest Map(R, \mathcal{D}, ω_g)

 8: **end loop**


Figure 2.1.2: The algorithmic architecture of SGIM-SAHT

SGIM-SAHT chooses an adequate task T_i , i.e. when the input space of L_{T_i} includes (s, ω_g) . Then it applies L_{T_i} to find the action $a^i = L_{T_i}(c_i)$. This inference process may be recursive until the output is an action, using the hierarchy between tasks. SGIM-SAHT thus infers from l_c a compound action $a = [a_1, \dots, a_n] \in \mathcal{A}^N$, to be executed by the robot. The trajectory of the episode with the primitive actions and controllables sequence and goal and reached outcomes ω_g, ω_r are recorded in the memory (Alg.2.1.1, 1.4). Then, it computes the learner’s competence for the goal outcome. In the RL framework, this competence can be seen as the reward for the goal outcome. In our multi-task learning setting, we use as competence a reward function common to all goals based on the Euclidean distance between the goal outcome ω_g and the reached outcome ω_r (Alg.2.1.1, 1.5). The memory and competence are used to update the models M_T and L_T , the set of tasks, and the hierarchy of models \mathcal{H} (Alg.2.1.1, 1.6). These mechanisms differ in our 3 algorithms. Besides, the competence is used to obtain an interest map that associates to each strategy and region of outcome space partition an interest measure to guide the exploration. The interest measure is computed as the progress or derivative of the competence for the goal outcomes (see [Nguyen and Oudeyer, 2012d]).

When the number of outcomes added to a region R_i exceeds a fixed limit, the region is split into two regions with a clustering boundary that separates outcomes with low from those with high interest (details in [Nguyen and Oudeyer, 2012d]). This mechanism of progress-based intrinsic motivation leads to **automatic curriculum**, automatically organising the training phase from simple tasks, for which progress can be made earlier, to complex tasks, for which progress are made later. This curriculum is all the more structured as complex tasks are compositions of simpler tasks in the hierarchical learning setting.

We detail IM-PB and CHIME and compare in the next subsections. The description SGIM-PB will be

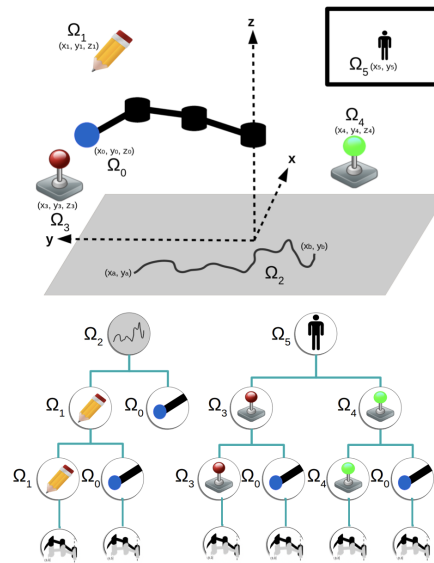


Figure 2.1.3: Top: Experimental setup: a robotic arm, can interact with the different objects in its environment (a pen and two joysticks). Both joysticks enable to control a video-game character (represented in top-right corner). A grey floor limits its motions and can be drawn upon using the pen (a possible drawing is represented). Bottom: Task decomposition as designed in the experimental setup.

detailed in the next chapter.

2.1.3 LEARN TASK HIERARCHY FROM A STATIC SET OF TASKS

A first stream of works proposed a representation of task hierarchy as a sequence of parametrised subgoals, in a teleological approach [Csibra, 2003] which considers actions as goal-oriented. As illustrated in Fig. 2.1.4, a task is specified with its parameters and can be decomposed as sequence of two subgoals, which are specified by variables and their parameters. This decomposition can be recursive, providing generalisability to several levels of hierarchy. Learning a task decomposition thus means identifying both subtasks and their parameters.

Let us consider that the set of tasks is given. The robot needs to choose which are easier to learn first, which are more difficult, and which tasks can be reused as sub-goals for more difficult models: we use knowledge transfer in the context of automatic curriculum learning. To learn the hierarchy \mathcal{H} between tasks, we proposed in [Duminy et al., 2018c] an implementation, IM-PB (Intrinsically Motivated Procedure Babbling) based on sequencing and interpolation of subgoals. Learning to draw with a pen and control a game character with joysticks, as illustrated in Fig. 2.1.3, IM-PB uses intrinsic motivation to learn sequential tasks by composing more simple tasks by learning autonomously from a known set of possible decompositions. IM-PB involves an exploration of the possible task decompositions, driven by intrinsic motivation.

The results in [Duminy et al., 2018c] report that IM-PB can learn the task decomposition as designed, owing to the task hierarchy representation. IM-PB can then use the learned relationship between tasks to

transfer knowledge from simple tasks to solve higher level tasks.

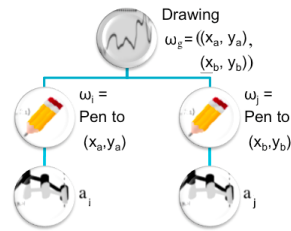


Figure 2.1.4: Representation of task hierarchy as a sequence of parametrised subgoals. To make a drawing ω_g between points (x_a, y_a) and (x_b, y_b) , a robot can recruit subtasks consisting in (ω_i) moving the pen to (x_a, y_a) , then (ω_j) moving the pen to (x_b, y_b) . These subtasks will be completed respectively with actions a_i and a_j . To complete this drawing, the learning agent can use the sequence of actions (a_i, a_j) . $\omega_i, \omega_j, a_i, a_j$ are inferred by regression.

2.1.4 SYMBOLIC REPRESENTATIONS OF TASKS

As I tackle the learning in continuous open-ended environments, the exploration of the continuous high-dimensional sensorimotor space faces the curse of dimensionality. In IM-PB, the curse of dimensionality problem increases with the addition of the space of all possible task decompositions to the exploration. Our experimental setup features a task space that is a composite and continuous space with tasks up to 4 dimensions and the action space is of dimension 14. Our results show that SGIM-PB can handle such complexity to learn hierarchical tasks. However, scaling this algorithm to higher dimensional sensorimotor spaces can prove difficult. Thus, in order to improve the exploration of the task decomposition space, **as a means to reduce the size of the task space, I examined whether one can derive a symbolic (ie. discrete) representation of tasks or task decomposition.**

2.1.5 LEARN TASK HIERARCHY FROM A DYNAMIC SET OF TASKS : ADAPTIVE LEARNING OF AN AFFORDANCE HIERARCHY THROUGH EXPLORATION

In the first line of work, I removed an assumption made in the previous studies with SGIM : that the specification of all tasks is known from the beginning by the robot. Thus, the set of all possible task decompositions are known, and mappings only need their parameters to be tuned. In [Manoury et al., 2019b], on the contrary, the learning agent only discovers through its exploration which observables of the environment are controllable and creates a new task to learn. Thus, the robot learns which objects it can control. More concretely, it creates a new neural network to learn the mapping for this task, with the observable as output. To allow quick learning with few data, I considered how to use small size neural networks, the algorithm CHIME selects the appropriate inputs through an evolutionary selection. The criteria for the stochastic selection is the estimated prediction accuracy. Once the input and output of the neural network chosen, the weights are learned by supervised learning with the data collected during intrinsically motivated exploration. Moreover, a task A is added to the set of possible subtasks only when the robot reaches a mastery level for task A above a threshold. Thus, the space of all possible task decompositions

grows progressively as the robot masters tasks. For instance, the manipulation of an object of the environment needs to reach a good level of mastery, before this object can be used as a tool to solve more complex tasks.

Thus, **CHIME implements the notion of tool use and affordance learning [Gibson, 1979] with a mechanism of emergence of tasks and a mechanism for task composition into hierarchically more complex tasks based on the same formulation of intrinsic motivation.** While it uses deep neural networks, its originality lies in the proposition of a network which architecture evolves throughout its learning curriculum, enabling it to learn simple tasks quickly. The nested models emerge online, in parallel to the learning by the robot of its automatic curriculum. The results in [Manoury et al., 2019b] show that the robot, owing to the intrinsic motivation heuristic, learns first the easy tasks, before the tasks higher in hierarchy: the reward in controlling its self-position increases first, in controlling the position of a movable object after, and at last pushing an object with another.

Like SGIM-PB, CHIME implements a temporal abstraction representation of tasks : a task is not defined by its duration but by its goal.

The experimental setup used in this chapter is presented in Figure 2.1.5.

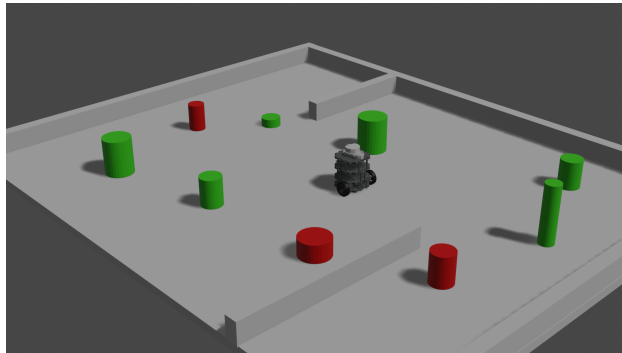


Figure 2.1.5: Experimental setup used: at the center is a mobile robot, the green objects represent movable entities, at the opposite of the red ones. The room is closed.

2.2 STAR : GOAL-CONDITIONED HIERARCHICAL LEARNING

Yet, CHIME still shows limitations for scaling to higher dimensional spaces because each task defines a continuous goal space. This is why in a second line of work, we build upon the framework of hierarchical reinforcement learning (HRL)[Dayan and Hinton, 1992; Sutton et al., 1999], which offers a potential solution for learning long-horizon tasks by training a hierarchy of policies. However, the abstraction used by policies is critical for good performance. Hard-coded abstractions often lack modeling flexibility and are task-specific. Moreover, for long-horizon tasks, it is crucial to engage in hierarchical reasoning across spatial and temporal scales. This entails planning abstract subgoal sequences, thus we aimed for an abstract representation of space. We examined how to discretise online during exploration the continuous goal space using formal methods, in collaboration with Sergio Mover, from LIX, Ecole Polytechnique, expert in formal methods. Recent studies show that representations that preserve temporally abstract environment dynamics are successful in solving difficult problems and provide theoretical guarantees for optimality. In [Zadem et al.,

2024], we propose a novel three-layer HRL algorithm, named Spatio-Temporal Abstraction via Reachability (STAR), that uses, at different levels of the hierarchy, both a spatial and a temporal goal abstraction. The goal space abstraction through reachability analysis results in a discretisation of the goal space, and thus a reduction of the exploration space. Thus, **STAR exploits an emergent discrete representation of the goal space, to be used more easily for task composition into sequential tasks.**

STAR uses as framework, a goal-conditioned Markov Decision Process $(\mathcal{S}, \mathcal{A}, P, r_{ext})$, where $\mathcal{S} \subseteq \mathbb{R}^n$ is a continuous state space, \mathcal{A} is an action space, $P(s_{t+1}|s_t, a_t)$ is a probabilistic transition function, and $r_{ext} : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ is a parameterised reward function, defined as the negative distance to the task goal $g^* \in \mathcal{S}$, i.e $r_{ext}(s, g^*) = -\|g^* - s\|_2$. The *multi-task reinforcement learning problem* consists in learning a goal conditioned policy π to sample at each time step t an action $a \sim \pi(s_t | g^*)$, so as to maximize the expected cumulative reward. The spatial goal abstraction is modeled by a *set-based abstraction* defined by a function $\mathcal{N} : \mathcal{S} \rightarrow 2^{\mathcal{S}}$ that maps concrete states to sets of states (i.e., $\forall s \in \mathcal{S}, \mathcal{N}(s) \subseteq \mathcal{S}$). We write $\mathcal{G}_{\mathcal{N}}$ to refer to the range of the abstraction \mathcal{N} , which is intuitively the abstract goal space. We further drop the subscript (i.e., write \mathcal{G}) when \mathcal{N} is clear from the context and denote elements of \mathcal{G} with the upper case letter G .

STAR learns, at the same time, a spatial goal abstraction \mathcal{N} and policies at multiple time scales. The STAR algorithm, shown in Figure 2.2.1, has two main components: a 3-levels Feudal HRL algorithm (enclosed in the **red dashed lines**); and an abstraction refinement component (shown in the **blue solid lines**). STAR runs the Feudal HRL algorithm and the abstraction refinement in a feedback loop, refining the abstraction \mathcal{N} at the end of every learning episode.

Our work presents the following contributions:

1. A novel Feudal HRL algorithm, STAR, to learn online

(Fig. 2.2.1) with the three RL agents :

- (a) *Commander*: the highest-level agent learns the policy $\pi_{\text{Comm}} : \mathcal{S} \times \mathcal{S} \rightarrow \mathcal{G}$ that is a goal-conditioned on g^* and samples an abstract **goal** $G \in \mathcal{G}$ that should help to reach the task goal g^* from the current agent’s state ($G_{t+k} \sim \pi_{\text{Comm}}(s_t, g^*)$).
- (b) *Tutor*: the mid-level agent is conditioned by the *Commander* goal G . It learns the policy $\pi_{\text{Tut}} : \mathcal{S} \times \mathcal{G} \rightarrow \mathcal{S}$ and picks **subgoals** in the state space ($g_{t+l} \sim \pi_{\text{Tut}}(s_t, G_{t+k})$).
- (c) *Controller*: the low-level policy $\pi_{\text{Cont}} : \mathcal{S} \times \mathcal{S} \rightarrow \mathcal{A}$ is goal-conditioned by the *Tutor*’s subgoal g and samples actions to reach given goal ($a \sim \pi_{\text{Cont}}(s_t, g_{t+l})$).

2. An emergent symbolic representation of the environment, based on reachability analysis. Each symbol representing a region of the state space, we estimate the k-step reachability of the forward model from each region. We refine the representation by splitting the regions online. We provide a theoretical motivation for using reachability-aware goal representations, showing convergence to a reachability-aware abstraction after applying a finite number of refinements, and showing a guarantee of a suboptimality bound for the converged policy.

Compared to the optimal policy, the policy learned on the abstraction has a near optimal value : $|V_{\pi^*}(s_i) - V_{\pi_{\mathcal{N}}}(s_i)| \leq U(\epsilon, i, \gamma)$, where U is a function of γ the discount factor, ϵ a bound on reward values, and i the time step index.

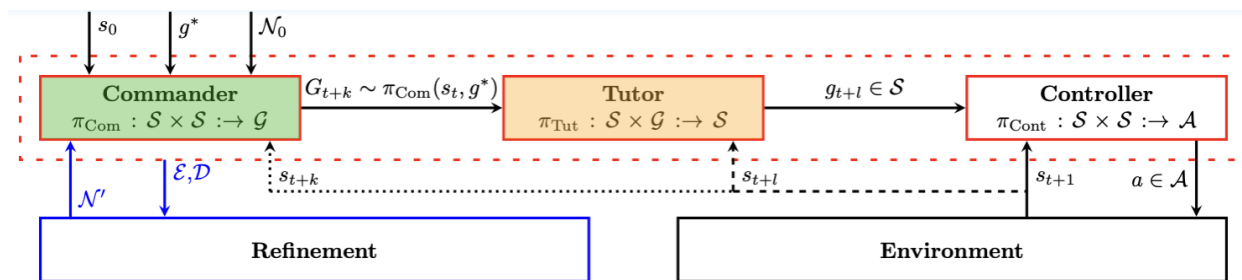


Figure 2.2.1: Architecture of STAR. The algorithm’s inputs are the initial state s_0 , the task goal g^* , and an initial abstraction \mathcal{N}_0 . STAR runs in a feedback loop a Feudal HRL algorithm (dashed red block) and an abstraction refinement (blue box). The solid red blocks show the HRL agents (Commander, Tutor, Controller). The agents run at different timescales ($k > l > 1$), shown with the solid, dashed, and dotted lines carrying the feedback from the environment to the agents. The Refinement uses as inputs the past episodes (\mathcal{D}) and a the list of abstract goals (\mathcal{E}) visited during the last episode, and outputs an abstraction.

- Empirical results showing that STAR successfully combines both temporal and spatial abstraction for more efficient learning, and that the reachability-aware abstraction scales to tasks with more complex dynamics in Ant environments (Fig. 2.2.2).

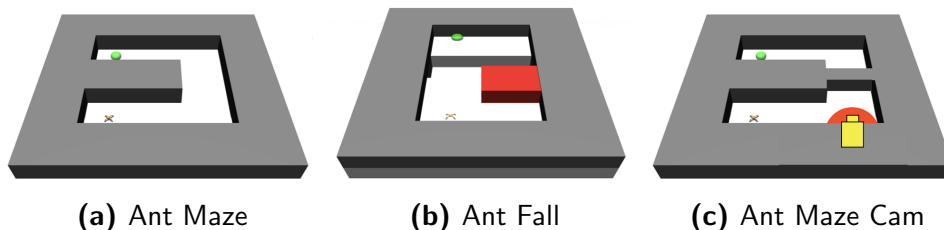


Figure 2.2.2: Ant environments. **Ant Maze:** in this task, the ant must navigate a ‘D’-shaped maze to reach the exit positioned at the top left. **Ant Fall:** the environment is composed of two raised platforms separated by a chasm. The ant starts on one of the platforms and must safely cross to the exit without falling. A movable block can be pushed into the chasm to serve as a bridge. Besides the precise maneuvers required by the ant, falling into the chasm is a very likely yet irreversible mistake. **Ant Maze Cam:** this is a more challenging version of Ant Maze. The upper half of the maze is fully blocked by an additional obstacle that can only be opened when the ant looks at the camera (in yellow) when on the red spot. The exit remains unchanged.

Our evaluation of the Ant environments show superior success rate of STAR compared to the state of the art. To analyse how the learned abstraction decomposes the environment to allow for learning successful policies, we examine the abstraction learned by STAR’s *Commander* agent at different timesteps during learning when solving the Ant Maze. Fig. 2.2.3 provides some insight on how STAR gradually refines the

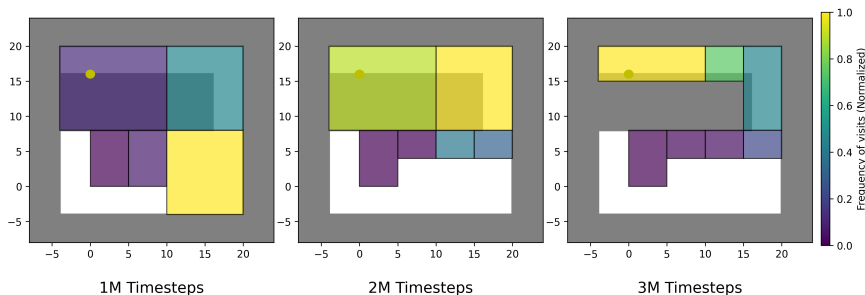


Figure 2.2.3: Frequency of goals visited by the *Commander* when evaluating a policy learned after 1M, 2M, and 3M timesteps (averaged over 5 different evaluations with 500 maximum timesteps). The subdivision of the mazes represent (abstract) goals. The color gradient represents the frequency of visits of each goal. Grey areas correspond to the obstacles of the environment in Ant Maze.

goal abstraction to identify successful trajectories in the environment. We can see that, progressively, the ant explores trajectories leading to the goal of the task. The most visited areas move from regions before the obstacle wall to the corner past the wall, then finally the region of the goals. Additionally, the frequency of visiting goals in the difficult areas of the maze (e.g., the tight corners) is higher, and these goals are eventually refined later in the training, fitting with the configuration of the obstacles. In particular, STAR learns a more precise abstraction in bottleneck areas where only a few subset of states manage to reach the next goal.

2.3 REPRESENTATIONS USED FOR THE ANALYSIS OF HUMAN ACTIVITY OF DAILY LIVING

In parallel to the previous investigations on representations of sequential tasks for learning agents to build them up, we also examined from the observational point of view, representations used in analysis methods. As the most complex motor activities in terms of compositionality and hierarchical decomposition into subtasks, we inspected our daily tasks : *activities of daily living* (ADL) include sleeping, eating, cooking, bathing ... as listed in datasets such as CASAS [Cook et al., 2012] or the ADL definitions provided by Katz [1983]. Indeed, an activity of daily living such as cooking and cleaning can vary greatly from one day to the other depending on the context and goal of the inhabitant. Moreover, ADL are sequential actions of variable length combining several primitive actions with multilevel temporal dependencies. Activities can also be deeply interconnected, underscoring the significance of temporal context in human activity recognition. Each activity can be viewed as a combination of unit actions that are selected and organised for the completion of a temporally distant goal. The activities of daily living share common properties as the sequential actions we considered earlier : the activities can be described as sequences of subtasks as multilevel time dependencies, the subtasks are recruited to complete a goal, and the temporal length and complexity degree of each task is not a-priori bounded. How complex actions such as activities of daily living can be described by hierarchical models ?

Our investigations first examined the recognition of activities based on video data as part of the project AMUSAAL. We proposed a hierarchical LSTM to classify the human poses in [Devanne et al., 2019], and showed that a **hierarchical representation is better** suited for activities of daily living.

As part of a collaboration with specialists in smart homes, Christophe Lohr from IMT Atlantique and the company Delta Dore, we then examined the problem of recognition of activities based IoT network data with sensors motion, door open/close, temperature... Due to specificities of the privacy-respecting design even-triggered sensors, the time series to analyse does not have the Markov property. Understanding one sensor activation often requires contextualization from the history and from other sensors, as shown in [Bouchabou et al., 2021b] where Word2Vec encoding can capture contextualized semantics, and in [Bouchabou et al., 2023c] where GPT transformer decoder can capture longer contexts. Compounding these issues, ambient sensor data manifest as noisy, multi-variate, and irregular time series with long-term dependencies, hardly modelled by traditional machine learning and deep learning methods. Examining multi-level temporal dependency, our work integrates (1) attention mechanisms for discerning the importance of sensor signals across a whole sequence, (2) pre-trained generative transformer embeddings capturing sensor inter-relations [Bouchabou et al., 2023c], (3) a hierarchical model emphasising activity succession for long-horizon dependency, and (4) a temporal encoding model to harness the timing of events. Despite semantic contextualisation, long-term dependency sequence models, and temporal encoding, a **hierarchical model of the time series leads to better recognition of the sequential activities**.

Indeed, for human activity recognition, hierarchical models have been proposed as ontologies of context-aware activities for recognition of activities of daily living in smart homes [Hong et al., 2009], hierarchical hidden Markov models [Asghari et al., 2019] for recognition of activities of daily living in smart homes, or hierarchical LSTM with two hidden layers for activity recognition from wearable sensors [Wang and Liu, 2020]. The hypothesis that sequential actions need to be represented by hierarchical models has been proposed in reinforcement learning : to solve sequential tasks, hierarchical reinforcement learning [Barto and Mahadevan, 2003; Barto et al., 2013] enabled tackling sequential tasks by decomposing into subtasks. These machine learning models confirm the neuroscience description that our nervous system selects and organizes actions in a hierarchical model [Grafton and de C. Hamilton, 2007], as well as the behavioural psychology studies showing that humans use hierarchical representations of action sequences of efficient planning and flexibility [Eckstein and Collins, 2021]. Our study provides another computation model supporting this hypothesis, but has the particularity of analysing activities of daily living which are very complex and variable tasks.

In relation to our robot learning algorithms STAR, CHIME and IM-PB, these works seem to **support our hypothesis that robots building up sequential tasks should rely on a hierarchical representation of tasks**.

3

Intrinsic Motivation to Learn with Teachers

While in the previous chapter, the learning agent was exploring its environment autonomously, in this chapter, I integrate social guidance in the learning process. I used theories of imitation learning and intrinsic motivation to bridge the paradigms of reinforcement learning and supervised learning, and devise the framework of socially guided intrinsic motivation. I proposed algorithms in the framework of reinforcement learning to enable learning agents to **learn multiple parametrised tasks by devising their own learning strategy: they choose actively what do learn, when to learn and thus their own curriculum; and also what, when and whom to imitate.**

From the point of view of cognitive science, our work can be seen as a first outline for model of the intrinsic motivation to interact with teachers, and thus a first step towards a **model of motivations for social interaction.**

3.1 LEARNING A PARAMETRISED TASK

My PhD research has taken a stance opposite to the current research in intrinsic motivation. While computational models of intrinsic motivation have examined how agents learn by interacting autonomously with its environment to control it without external reward by implementing an intrinsic reward based on its empirical progress, my PhD research has proposed a formulation of the intrinsic motivation of interacting with teachers: **what can drive a learning robot to actively interact with teachers?** Our proposition is inspired by developmental psychology studies that describe how children can learn by trial and error during their exploration with the environment, and also by requesting their parents' help in an active manner.

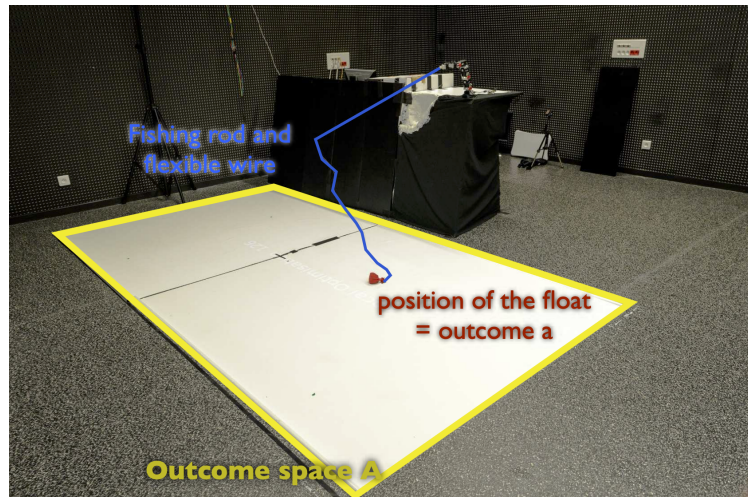


Figure 3.1.1: The robot observes the outcome a , the final position of the ball after its movement. It learns which movement can reach different positions on the floor. The camera is placed above the white surface, and can only see the white surface which materialises the outcome space A .

My first step was to study how control learning in continuous sensorimotor spaces for parametrised tasks with a continuous goal space, can benefit from using both paradigms.

With a robot arm learning policies for manipulating a fishing rod to place the float to any position (see fig 3.1.1), I showed that action and outcome demonstrations by humans lead to better generalisation to a field of tasks, in [Nguyen and Oudeyer \[2014\]](#). In [Nguyen and Oudeyer \[2012a\]](#), I showed that human demonstrations given by kinesthetics display different distribution characteristics in the trajectory profile. Comparing the demonstrations recorded from a human and movements provided by a robot that has finished learning, the most noticeable difference is that the joint value trajectories of the human demonstrator are all monotonous, and seem to have the same shape, only scaled to match different final values. Indeed, the comparison indicates that human demonstrations are not randomly generated but are well structured and regular, while the demonstrations from expert robots come from different probability distributions. Therefore, the human demonstrator shows a bias through his demonstrations to the robot, and orients the exploration towards different subspaces of the policy space. Because **human demonstrations have the same shape, they belong to a smaller, denser and more structured subset of trajectories from which it is easier for the learner to generalise, and build upon further knowledge**. Thus, a robot can exploit this bias to learn to generalise to more tasks by using data from both autonomous goal-directed exploration and from a few demonstrations. Devising the framework of Socially Guided Intrinsic Motivation, I proposed several implementations, which are particular cases of SGIM-SAHT (Algo. 2.1.1). Fig. 3.1.2 outlines the difference between the algorithms, the experimental setups they were tested on, and the main results; Fig. 3.1.3 illustrates different strategies used by each algorithm on the timeline of the learning process.

First, the algorithm SGIM-D, for learning online parametrised tasks in a passive manner, alternates at a fixed frequency between autonomous exploration and imitation learning [[Nguyen and Oudeyer, 2014](#)].



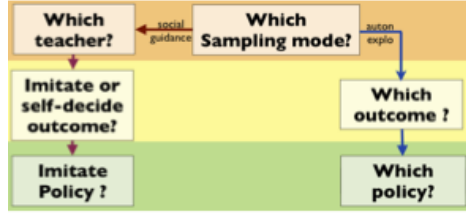

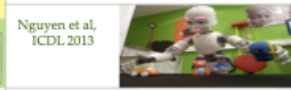
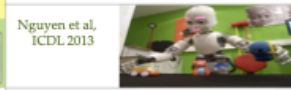

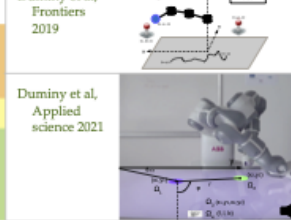

| Active learning | Experimental setup | Hypotheses and Results |
|--|--|---|
| <p>SGIM-D</p>  | <p>Nguyen and Oudeyer, Autonomous Robots 2014</p>  | <ul style="list-style-type: none"> ☹️ Can a combination of autonomous exploration and social guidance bootstrap learning? ☹️ What are the properties of this combination that bootstrap the learning process? ✓ learns as well or better precision, more reliably, faster ✓ uses demonstrations to bias its search in the policy and outcome spaces ✓ uses self-exploration to overcome correspondence problems ✓ uses self-exploration to compensate sparsity of demonstration set |
| <p>SGIM-ACTS</p>  | <p>Nguyen and Oudeyer, JBR 2012</p>  <p>Moulin-Frier et al, 2014</p>  <p>Nguyen et al, ICDL 2013</p>  | <ul style="list-style-type: none"> ☹️ Can competence progress be a measure for an active choice of what, when, whom to imitate? ☹️ Can the active choice scale to a high-dimensional continuous environment with several teachers to learn an open-ended number of tasks and tasks of different types? ✓ interactive learning with several teachers ✓ learn several types of tasks ✓ tested in continuous and discrete spaces ✓ tested in simulation and physical robot ✓ model for child development |
| <p>SGIM-PB</p>  | <p>Duminy et al, Frontiers 2019</p>  <p>Duminy et al, Applied science 2021</p>  | <ul style="list-style-type: none"> ☹️ Can a goal-oriented representation of task decomposition enable the agent to learn the hierarchy between tasks ? ☹️ Can competence progress be a measure for choosing to decompose a task or learn directly the policy? ☹️ Can intrinsic motivation choose to ask teachers different types of demonstrations? ✓ learn the relationship between tasks to enable several types of tasks ✓ tested in continuous and discrete spaces ✓ tested in simulation and physical robot ✓ Policy space exploration for simple tasks and task decomposition for sequential tasks |

Figure 3.1.2: Three implementations of the SGIM framework are presented with various illustrative experiments and their corresponding publication, as well as the main hypotheses and results. Each algorithmic architecture allows the agent to take active control of various aspects of its learning strategy. Each of the results presented allow us to present aspects of the advantages for a fully active system, that can decide on all aspects of its learning strategy.

SGIM-D can produce more varied outcomes in the environment, covering a larger area of the task space, and is robust to large size task spaces, even when the set of reachable tasks covers a small region. This first study **validates the benefits of combining both learning paradigms**, and allowed us to consider more varied interactions.

Second, by assessing the progress thanks to each demonstration, I proposed a formulation of the intrinsic motivation to request for demonstrations from a teacher, and proposed the algorithm SGIM-ACTS to actively collect data by choosing when to explore autonomously and when to request a demonstration; in which case, it chooses what to imitate (what type of demonstration to ask : a goal for emulation or an action for mimicry), and whom to imitate by learning each teacher’s field of expertise [Nguyen and Oudeyer, 2012d]. Thus, SGIM-ACTS can choose the learning strategy and its curriculum. It has been successful at learning parametrised control tasks of different nature from multiple teachers [Nguyen and Oudeyer, 2012d], at recognising 3D objects by collecting data through different manipulation strategies and by requesting the help of a human [Ivaldi et al., 2013].

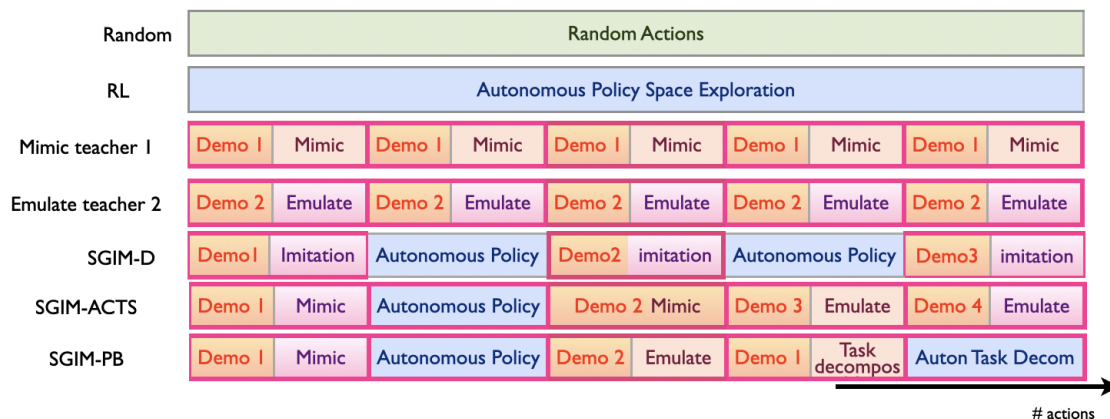


Figure 3.1.3: Comparison of the timeline of several learning algorithms : (a) Random exploration of the action space A ; (b) autonomous exploration with Reinforcement Learning; (c) Mimicry to reproduce a demonstrated action, received at a fixed frequency; (d) Emulation learning to reproduce a demonstrated outcome, received at a fixed frequency; and (e) SGIM-D that requests demonstrations on a fixed frequency; (f) SGIM-ACTS that decides when and who to request demonstrations of actions and outcomes and (g) SGIM-PB that decides when and who to request demonstrations of actions or decomposition into subgoals.

3.2 INFLUENCE OF DEMONSTRATIONS

We could also model the **influence of the mother tongue in the development of vocalisation in infants**, in a collaboration with Clement Moulin-Frier from Flowers Team, Inria [Moulin-Frier et al., 2014] : we simulated how an agent can activate the different muscles of the mouth and vocal tract to learn to produce sounds. Our findings show that the sounds produced during the automatic curriculum bears resemblance with child development described in behavioural psychology, but also that the influence of the mother tongue on the sounds produced can be explained by our formulation of intrinsic motivation to imitate the mother tongue’s sounds. Thus our formulation of intrinsic motivation for interacting with teachers enables better performance for learning robots from an engineering perspective, but also an element of **modelling child development in cognitive science**.

3.3 ROBUSTNESS TO THE QUALITY OF DEMONSTRATIONS

I showed in these works that the SGIM algorithms is **robust to poor quality demonstrations**, comparing a helpful teacher who provides with different demonstrations with a teacher who repeats the same demonstrations from a small demonstration dataset, we show in [Nguyen et al., 2013] that the robot’s performance is not impacted, contrarily to a random choice of strategies : it can choose the right learning strategy to cope with poor quality demonstrations. SGIM can handle correspondence problems that occur when the robot can not move like the teacher, due for instance to correspondance problems, which come from differences in their bodies [Nguyen and Oudeyer, 2012d; Duminy et al., 2016].

3.4 COMPOSITIONAL TASKS WITH ACTIVE IMITATION LEARNING

With the algorithm SGIM-PB (Socially Guided Intrinsic Motivation with Procedure Babbling), Duminy et al. [2021] extends, this intrinsic motivation criteria for requesting from the best teacher either a demonstration of a policy or a task decomposition. It uses the same representation for task decomposition as IM-PB, presented in the previous chapter. The interaction with teachers thus has a supplementary strategy: asking how to decompose a task. SGIM-PB outperforms SGIM-ACTS on the most complex outcome spaces, owing to task decomposition which enables it to learn and exploit the task hierarchy of this experimental setup and previously learned skills.

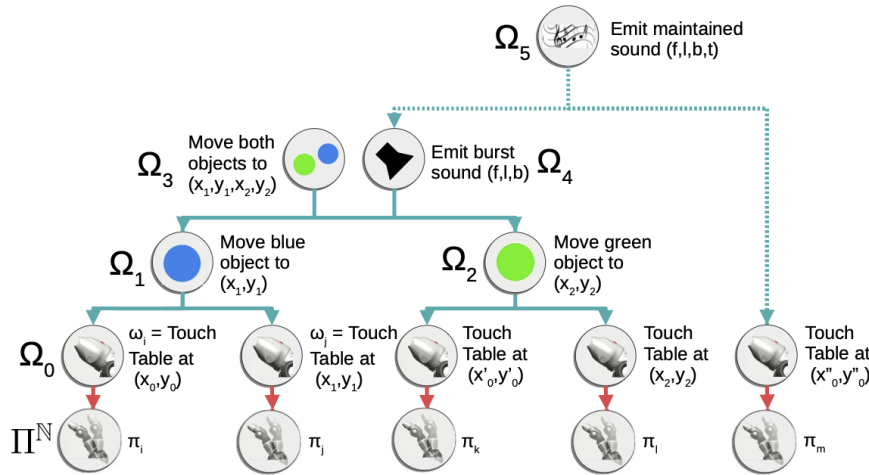


Figure 3.4.1: Task hierarchy of the Yumi experimental setup : blue lines represent task decomposition for the simulation setup, dashed lines for the physical setup, red lines for the direct inverse model. Both task decomposition and inverse model are learned online. Eg. to move a blue object placed initially at (x_0, y_0) to a desired position (x_1, y_1) , the robot can carry out task (ω_i) that consists in moving its end-effector to (x_0, y_0) to pick it, then task (ω_j) that consists in moving the end-effector to (x_1, y_1) . These subtasks are executed with respectively action primitives π^i and π^j . Therefore to move the object, the learning agent can use the sequence of action primitives (π^i, π^j) .

For a 6 DOF robot arm learning hierarchical tasks as represented in Fig. 3.4.1, our results also show that demonstrations of task decomposition can efficiently replace demonstrations of sequential motor policy. Most of all, our robot **learned to ask demonstrations of motor policies for simple tasks and to ask demonstrations of task decomposition for sequential tasks** [Duminy et al., 2018b]. Likewise, for autonomous exploration, it explores directly the policy space for simple tasks and the task decomposition space for sequential tasks. These behaviours derive from the same criteria for active learning : the learning progress. This intrinsic reward is thus a common criteria for choosing tasks and for choosing teachers.

To understand the reasons for the better performance of SGIM-PB on a physical robot setup, let us examine the histogram of the strategies chosen per task in Figure 3.4.2. For the most complex task, Ω_5 , SGIM-PB uses massively task decomposition space exploration (AutonomousProcedure) compared to action space exploration (Autonomous Actins). Owing to autonomous task decomposition exploration, SGIM-PB

can thus learn complex tasks by using the decomposition of tasks into known subtasks. This highlights the **essential role of the task decomposition representation and the task decomposition space exploration by imitation but also by autonomous exploration, in order to learn high-level hierarchy tasks, which have sparse rewards, so as to adapt the complexity of the action sequences to the hierarchy level of the goal task.**

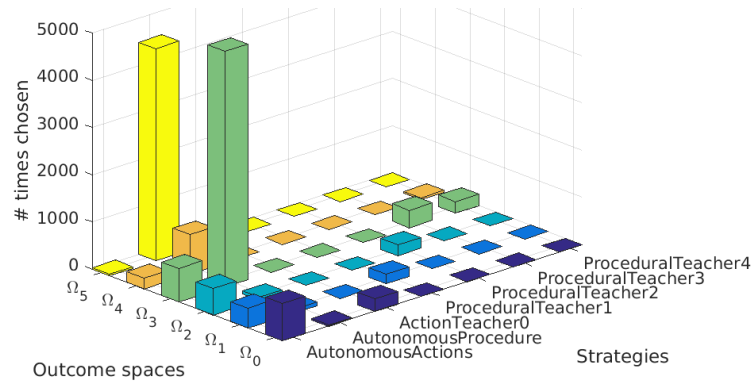


Figure 3.4.2: Choices of strategy and goal outcome for the SGIM-PB learner during the learning process. Strategies are autonomous exploration of the action space, task decomposition (=procedure), mimicry of actions, or imitation of task decomposition (=procedural teacher).

Besides, the evolution of outcome spaces (Figure 3.4.3) and strategies (Figure 3.4.4) chosen during training show the same trend from iterations 0 to 10,000: the easy outcome spaces $\Omega_0, \Omega_1, \Omega_2$ are more explored in the beginning before being neglected after 1000 iterations, to explore the most complex outcome spaces Ω_3, Ω_5 . Imitation is mostly used in the beginning of the learning process, whereas later in the automatic curriculum, autonomous exploration is preferred.

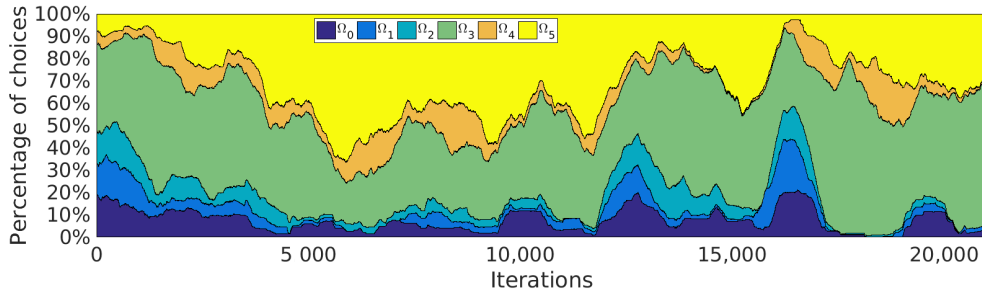


Figure 3.4.3: Automatic Curriculum Learning : Evolution of choices of tasks for the SGIM-PB learner during the learning process on the physical setup.

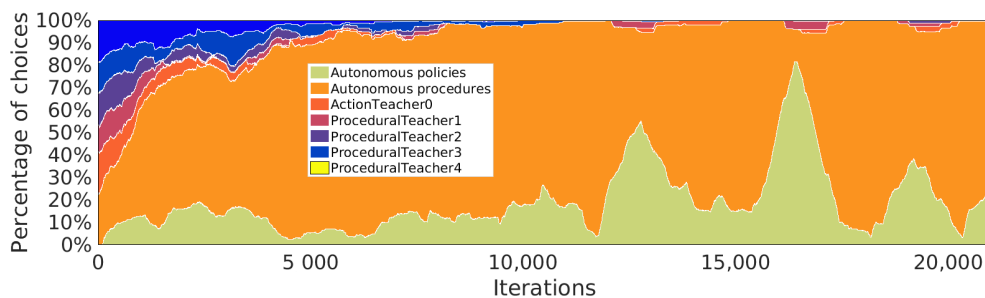


Figure 3.4.4: Evolution of choices of strategies for the SGIM-PB learner during the learning process on the physical setup.

To summarise, SGIM-PB shows the following characteristics :

- Hierarchical RL: it learns online task decomposition on 4 levels of hierarchy using the procedural framework; and it exploits the task decomposition to match the complexity of the sequences of action primitives to the task;
- Automatic curriculum learning: it autonomously switches from simple to complex tasks, and from exploration of actions for simple tasks to exploration of procedures for the most complex tasks;
- Imitation learning: it empirically infers which kind of information is useful for each kind of task and requests just a small amount of demonstrations with the right type of information by choosing between procedural and action teachers;
- Transfer of knowledge: it automatically decides what information, how, when to transfer and which source of information for cross-task and cross-learner transfer learning.

Thus, **our work proposes an active imitation learning algorithm based on intrinsic motivation that uses empirical measures of competence progress to choose at the same time what target task to focus on, which source tasks to reuse and how to transfer knowledge about task decomposition.** Our contributions, grounded in the field of cognitive robotics, are : a new representation of complex actions enabling the exploitation of task decomposition and the proposition for teachers of supplying information on the task hierarchy to learn compound tasks.

3.5 SOCIALLY GUIDED INTRINSIC MOTIVATION

We have presented two algorithms for active imitation learning based on intrinsic motivation. While SGIM-ACTS learns several parametrised tasks by choosing its teachers and the timing of its requests, and SGIM-PB learns uses transfer of knowledge between tasks by learning the relationship between the parametrised tasks with the supplementary strategies of task decomposition by autonomous exploration and task decomposition imitation. The main differences between SGIM-D, SGIM-ACTS and SGIM-PB are outlined and they are contrasted with autonomous learning algorithm IM-PB in Table 3.5.1.

The various implementations and experimental setups show that the robot benefits more from exploring directly the policy space for simple tasks, and by trying to decompose into subtasks for sequential tasks. With

Table 3.5.1: Differences between SGIM-D, SGIM-ACTS, SGIM-PB and IM-PB in the learning strategies used, the imitation content and timing, and the modalities of transfer of knowledge.

| Algorithm | Action Representation | Strategies σ | What to imitate | When to imitate | Who to imitate | Transfer of Knowledge |
|--|---|---|---|--|--|---|
| IM-PB Duminy et al. [2018c] | parametrised continuous actions, task decomposition | Autonomous action exploration, Autonomous outcome space exploration, autonomous task decomposition exploration | NA | NA | None | Cross-task transfer |
| SGIM-D Nguyen and Oudeyer [2014] | parametrised continuous actions | autonomous action space exploration, mimicry of an action teacher, emulation of an outcome teacher | Teacher demo. of actions and outcomes | Predefined frequency | 1 teacher with correspondence problems | Mimicry, Emulation |
| SGIM-ACTS Nguyen and Oudeyer [2012d] | parametrised continuous actions | autonomous action space exploration, mimicry of action teachers, emulation of outcome teachers | Teacher demo. of actions and outcomes | Active request by the learner to the teacher | Several (imperfect) teachers | Mimicry, Emulation |
| SGIM-PB Duminy et al. [2021] | parametrised continuous actions, task decomposition | autonomous action space exploration, autonomous procedural exploration, mimicry of an action teacher, mimicry of a task decomposition teacher | Teacher demo. of actions and task decomposition | Active request by the learner to the teacher | Several teachers | Cross-task transfer, Mimicry, Emulation |

our intrinsic motivation criteria, it can choose at each time step the appropriate exploration strategy, both in autonomous exploration strategies and in imitation learning strategies, and thus automatically composes its learning curriculum.

To summarise, I **extended the reward hypothesis to active interactions with teachers, with a formulation of the intrinsic motivation that is valid both for autonomous exploration and for social guidance, may the demonstrations requested to teachers be low-level policies, goals or decomposition into subgoals, in multi-task learning setting with parametrised goals or hierarchical goals.** Thus, by considering that each choice of teacher and type of demonstration is a learning strategy σ , the intrinsic motivation reward for a learning strategy σ and task ω can be modelled as

$$im(\sigma, \omega) = \kappa(\sigma) * progress(\sigma, \omega) \quad (3.1)$$

where *progress* is the empirical progress measured through the last episodes with strategy σ and goal ω , and $\kappa(\sigma)$ is the cost of the strategy, representing the availability of teachers, their willingness to interact with the robot ...

4

Application to Socially Assistive Robotics

4.1 A ROBOT COACH PROPOSING AN AUTOMATIC CURRICULUM

The algorithms for multi-task learning SGIM, CHIME and GARA automatically design their own learning curriculum, by deciding at each step which task to learn and the appropriate learning strategy. Such a learning curriculum can be used by intelligent tutoring systems for a robot coach to propose exercises personalised to each user. In [Annabi and Nguyen, 2023a] we study an **intelligent tutoring system based on intrinsic motivation that chooses exercises to maximise the progress for each student**. The efficiency of this intelligent tutoring system relies on the knowledge about the relationship between the exercises, and the relationship between the skills. As with our study on hierarchically sequential tasks, some exercises train high-level skills (for instance multiplication in maths) that require other simpler skills to be mastered before (for instance addition in maths is a prerequisite before learning multiplication). We propose an algorithm to **uncover these pre-requisite relationships using causality analysis, based only data of student scores**. We therefore obtain a hierarchical relationship between exercises, but this time with respect to pre-requisite requirements.

These intelligent tutoring systems can be applied to socially assistive robots to personalise training programs for robot coaches.

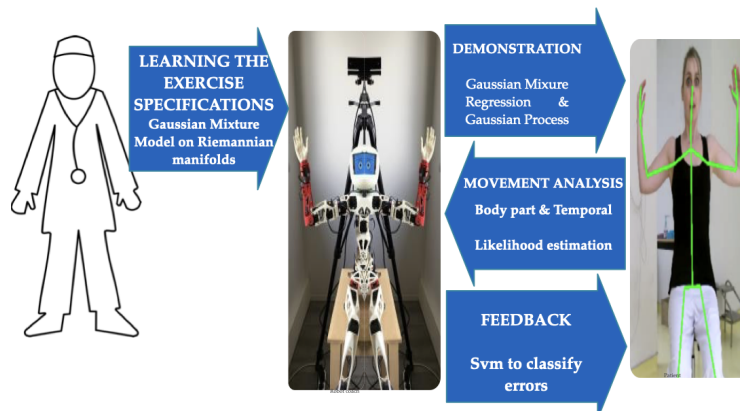


Figure 4.2.1: The robot learns the probabilistic model of exercises from demonstrations using GMM on Riemannian manifolds. The robot assesses the patient’s movements per body part and temporal segment by estimating the likelihood, and gives a feedback after classification of the error.

4.2 ROBOT COACH FOR PHYSICAL REHABILITATION

4.2.1 HUMANOID SOCIALLY ASSISTIVE ROBOT IN A CLINICAL TRIAL

As part of the Keraal experiment, funded by EU FP-7 ECHORD++, for which I was the scientific coordinator, we proposed a **humanoid robot coach for physical rehabilitation**. The consortium includes Olivier Remy Neris, director of the rehabilitation department of CHRU Brest; the company Generation Robots; Mathieu Simonnet, specialist in cognitive ergonomics at LEGO of IMT Atlantique, and Myriam Le Goff-Pronost, health economist at LaTIM, IMT Atlantique. Its goal is to entice motivation in patients while they exercise by themselves. This humanoid robot can entice motivation by its embodied presence and by the feedback it can give. In [Devanne and Nguyen, 2017], we applied Gaussian Mixture Model (GMM) on Riemannian manifolds [Zeestraten et al., 2017] for giving **feedback after analysing patients’ movements** for three low-back pain exercises using a single RGB-D camera. In later works, we tested LSTM auto-encoder [Nguyen et al., 2024] and Spatio-Temporal Graph Convolutional Networks [Marusic et al., 2023a] to analyse the whole body movements.

In order to design our system and our HRI, we first choose an anthropomorphic robot platform, and led a psychological analysis of the target population in a co-design process [Devanne et al., 2018], in collaboration with Gilles Kermarrec, psychologist at European Research Center for Virtual Reality and Research Center for Education, Learning and Didactics, European University of Brittany. Then, this system has been tested with 21 subjects, including 12 patients in **clinical trials** in a longitudinal study over daily sessions over three weeks for each patient. The system is assessed with medical criteria that there is non-inferiority of this system compared to classical care with therapists [Blanchard et al., 2022].

4.2.2 FOUR CHALLENGES FOR A DATASET

This clinical trial enables us to identify four challenges for robot coaches and publish a medical low-back pain physical rehabilitation dataset for human body movement analysis [Nguyen et al., 2024]. With the aim of rehabilitation using human body movement analysis, intelligent tutoring systems (ITS) need to analyse complex full-body exercises that can involve several parts of the body but not necessarily all parts of the body. The assessment algorithm should be able to understand which parts are important, and what are the ranges of freedom that are acceptable. The ITS should encapsulate the tolerated variance for each joint and time frame. Thus, we have identified 4 challenges in rehabilitation movements analysis:

1. Rehabilitation motion assessment. The goal is to assess an observed motion sequence by detecting if the rehabilitation exercise is correctly performed or not. Thus each recording is labeled with a global evaluation: correct or incorrect, and also if the error is significant, medium or small.
2. Rehabilitation error recognition. The goal is to classify the observed error among a set of known errors, so as to explain and give feedback. Each recording has the label of the error and also if the error is significant, medium or small.
3. Spatial localization of the error. In addition to recognizing the error, the goal is also to identify which body part is responsible of the error to draw the patient’s attention to it. Each annotation indicates the body part causing the error.
4. Temporal localization of the error. The goal is to detect the temporal segment where the detected error occurred along the sequence. Each annotation indicates the time window where the error occurs.

While most rehabilitation datasets [Leightley et al., 2015; Ar and Akgul, 2014; Vakanski et al., 2018; Dolatabadi et al., 2017; Miron et al., 2021; Aung et al., 2016; Capecci et al., 2019] only provide annotations for the rehabilitation motion assessment, we propose a medical dataset recruiting 12 clinical patients carrying out low back-pain rehabilitation exercises over 3 weeks of rehabilitation program. The Keraal dataset is composed of 2622 recordings , including 1881 recordings from patients. Each recording is composed of 3D Kinect skeleton positions and orientations, RGB videos, 2D skeleton data, and medical annotations by two medical annotators to assess the correctness, label and timing errors of each movement part.

4.2.3 PERSONALISATION OF THE COACHING TO DIFFERENT MORPHOLOGIES

Our current work seeks to personalise this coaching, by taking into account the physical limitations of each person. We proposed to use shared latent variables to translate the movement representation between the robot and the human motions from patients of different morphologies, representing patients who have different biomechanical constraints because of their pathologies or physical differences. In [Devanne and Nguyen, 2019], the common representation was obtained using a shared Gaussian Process Latent Variable Model (shared GP-LVM). Simulating mechanical constraints of joint angles, we could show that the robot could extend the model to these constraints and adapt the assessment of rehabilitation exercises to each patient’s physical limitation, by updating the inverse mapping matrix W , as shown in Fig. 4.2.2.

More recently, we started investigating how we can learn this common latent space when we do not have recordings of all exercises for the all biomechanical models. Indeed, in the absence of paired data, supervised

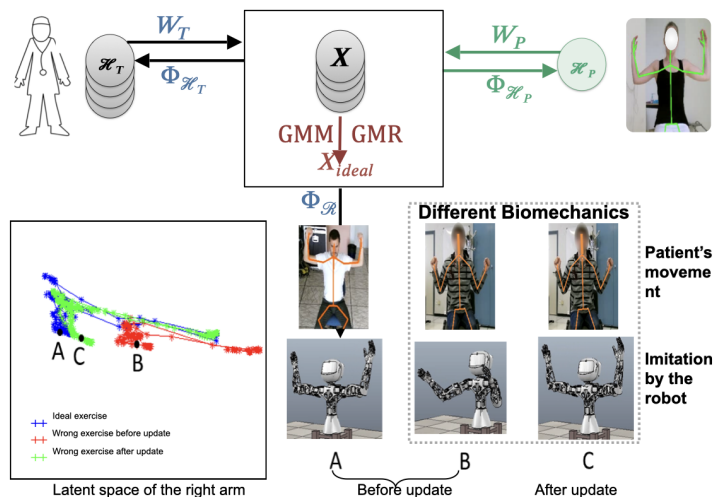


Figure 4.2.2: The shared GP-LVM algorithms learn a latent space X from paired data from teacher’s and a patient’s movements, and the parameters W and Φ to translate between each person’s movement space and the latent space. Using Gaussian Mixture Model (GMM) and Gaussian Mixture Regression (GMR), we translate the movement of a new patient to make the robot imitate. Bottom right: illustration for three movements : the movement of a new patient, the corresponding movement by the robot. Bottom left: representation in the latent space of the right arm of the three movements: movement A is the ideal movement of the exercise by a patient with the same biomechanics as the teacher; movement B is the movement by a patient with a different biomechanics, before the model update.; movement C is the same as B but after the model update. Figure from [Devanne and Nguyen, 2019].

data is not possible. In Annabi et al. [2024], we explored unsuccessfully deep learning methods for unpaired domain-to-domain translation, that we adapt in order to perform human-robot imitation.

The next steps would be to propose an intelligent tutoring system to personalise the training curriculum to each patient. We plan to analyse the exercises proposed by therapists to uncover their prerequisite relationships and to personalise the training curriculum, by applying the algorithms to uncover the prerequisite relationship between exercises proposed in [Annabi and Nguyen, 2023a].

4.3 IMITATION GAME FOR ASD CHILDREN

In collaboration with the psychiatrist Nathalie Collot-Lavenne from of CHRU Brest, we adapted a program of their daily care for **autism spectrum disorder (ASD) children to build a robot coach for imitation games**, building on the same movement analysis algorithms.

Indeed, while autism is an neurodevelopmental condition affecting social interactions and motor capacities, it has been shown that robots and ICT technologies can induce social and cognitive stimulation of children with ASD. This seems to be partly because they are more comfortable with explicit and predictive behaviors. Several studies have shown that imitating ASD children can be efficient in enhancing their social behaviour [Nadel, 2005]. These studies have been carried with caregivers as imitators of children’s movements [Sanefuji et al., 2009; Katagiri et al., 2010]. While most imitation studies have focused on facial expressions or sounds [Dautenhahn et al., 2009; Kozima et al., 2009; Pioggia et al., 2007; Jouen et al., 2017], few studies have focused on the imitation of movements. Development of motor control requires forming an internal model

of action relying on the coupling between action (motor commands) and perception (sensory feedback). Critical to the development of social, communicative, and motor coordination behaviors, internal model of action accurately predicts the sensory consequences of motor commands [Krakauer and Shadmehr, 2007]. This is why, we propose to examine how ASD children can learn to improve their motor control by a serious game. We propose to use a robot platform as they have been shown to arise their interest. We therefore implemented a robot for an imitation game based on a nursery rhyme. Through a user experiment, we tested the application of the Gaussian Mixture Model to analyse the movements of ASD children in the setting. The experiments were described in [Nguyen et al., 2019; Vallee et al., 2020]. These studies support the use of a humanoid robot as a therapeutic tool to improve imitation skills in ASD children.

5

The motivation to interact actively with tutors

Our work lies at the intersection of several fields, which can be related as pictured in Fig. 5.0.1. We propose hierarchical action models through the study of activities of daily living and human movement analysis. The latter is applied for socially assistive robotics with a robot coach for physical rehabilitation.

In parallel, hierarchical action models are also harnessed in reinforcement learning with algorithms of hierarchical reinforcement learning algorithms that rely on symbol emergence and the heuristic of intrinsic motivation to structure its learning strategy and derive automatically its learning curriculum. Most of all, we frame the field of socially guided intrinsic motivation, that extends the theory of intrinsic motivation to situations where the agent chooses aspects of its interaction with teachers. We formulate a reward function for choosing when, what, whom to imitate for multi-task learning, hierarchical learning and automatic curriculum learning. We aim to apply this theory for intelligent tutoring systems.

In Chapter 3, through the SGIM framework [Nguyen and Oudeyer, 2012d; Duminy et al., 2021], with my previous advisors and my collaborators, I proposed an **active imitation learning algorithm for the robot to actively decide when, what and from whom to imitate, and when to learn autonomously, based on intrinsic motivation** [Oudeyer and Kaplan, 2009; Schmidhuber, 2010]. Nevertheless, the tasks solved remain quite simple, and have the disadvantage of relying only on low-level demonstrations by kinaesthetic of movements or outcomes, instead of more natural means of communication with humans. In the future, we aim for more complex tasks such as tool use, and explore how this active imitation learning can use natural interactions.

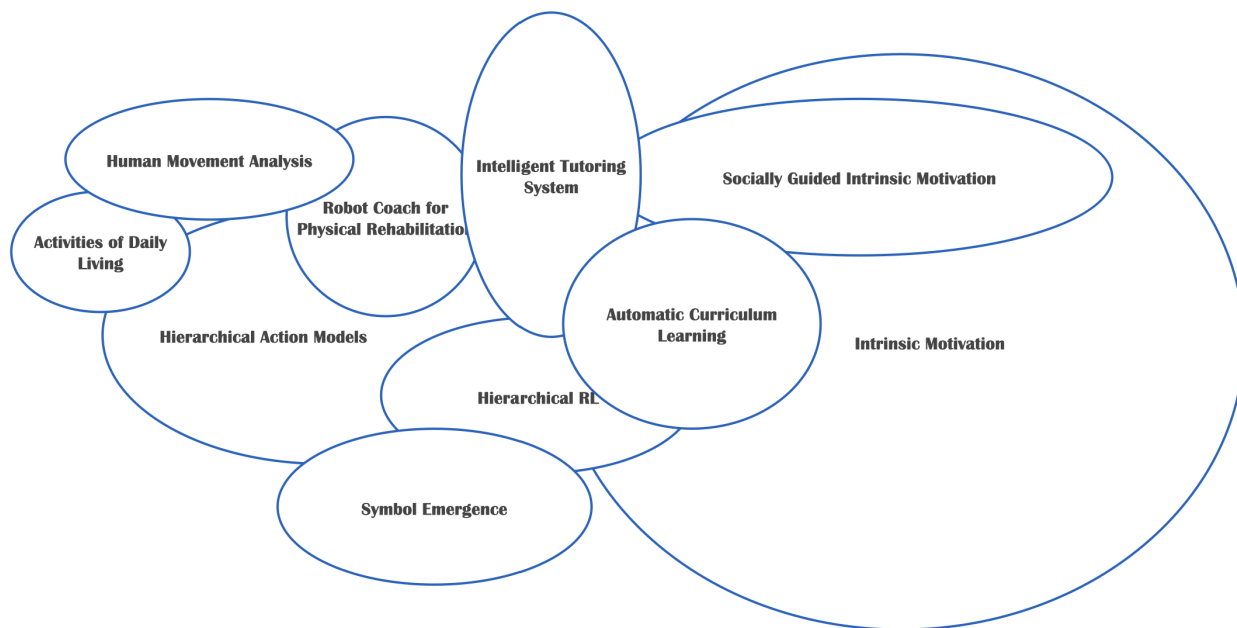


Figure 5.0.1: Fields of our research projects

5.1 RESEARCH PROBLEM

Developmental psychology works such as [Fagard et al., 2016] have started to study the influence of social guidance beyond the learning of simple motor skills, but have researched the learning of hierarchical skills such as tool use. They highlight the existence of a zone of proximal development of children for sequential tasks. Even when each subtask is known, composing the subtasks can be a challenge. In [Fagard et al., 2016], children from 18 months of age succeeded in tool use thanks to demonstrations, while younger children could not benefit from it. Therefore, there is a pre-requisite to social learning. We argue that this pre-requisite is linked to an emergent representation of complex tasks that needs to be known by learning agent to enable it to map the observations to its representational structure.

The purpose of my future research project is to model both how artificial agents can interact efficiently with tutors to learn in open-ended settings, and to model the motivations of humans to interact with tutors in a hierarchical learning setting. We investigate representations of sequential tasks and their emergence, as well as the refinement and the generalisation of my formulation of the motivational drive of learning agents to interact with tutors. This constitutes an extension of the reward hypothesis to social interaction.

We propose to investigate following two axes.

The first axis examines how a representation of complex tasks can enable the learning agent to understand the guidance of tutors. We hypothesise that an emerging discrete representation of our continuous environment and tasks is key to an efficient communication between the learner and the tutor.

The second axis examines the different factors of the motivational drive of learning agents to interact with tutors by analysing human interactions with tutors and by expanding the model of intrinsic motivation

for learning to more drives and factors, integrating human-robot interaction methods, affective computing theories. For instance, we will expand beyond the intrinsic motivation based on learning progress and consider other intrinsic motivations such as homeostatic drives. Besides, we investigate other functions of imitation than learning : imitation can and is often mostly considered as a communication method [Nadel et al., 2004]. The structures underlying our ability to imitate enable the emergence of both learning and communication abilities. For instance a tutor imitating the learner provides salient signals for his learning process, in which case imitation is both a means of a communication and of teaching [Andry et al., 2001].

5.2 REPRESENTATIONS OF TASKS TO ENABLE SOCIAL LEARNING

5.2.1 CHALLENGES

To enable natural communication between robots and humans, we look for a *learning process for robots to acquire a discrete representation of tasks in a continuous environment, and how this discrete representation can be used by the robot to learn more efficiently with human guidance*. Indeed, symbol emergence is key for open-ended learning to tackle the curse of dimensionality and scale up to open-ended high-dimensional sensorimotor space, by allowing symbolic reasoning, compositionality, hierarchical organisation of the knowledge, etc. While symbol emergence has been recently investigated for the sensor data such as images and video, action symbolization can lead to a repertoire of various movement patterns by bottom-up processes, which can be used by top-down processes such as composition to form an action sequence, planning and reasoning for more efficient learning, as reviewed in [Taniguchi et al., 2018], but also social guidance.

We will propose a learning algorithm where the robot learns multiple compositional tasks by deciding itself its learning strategy between reinforcement learning and social guidance and how to efficiently interact with tutors, by means of a symbolic representation of the tasks, emerging from its sensorimotor data. This line of research contributes to human-centered Artificial Intelligence and robotic learning.

5.2.2 LEARNING A SYMBOLIC REPRESENTATION

Human users can give teaching signals to robots by low-level kinaesthetic demonstrations of movements [Billard and Siegward, 2004; Schaal et al., 2003], but can also take a more natural means of communication on a higher level. High-level communication, such as language relies on symbols or a discrete abstract representation of space. How can a robot understand these words or symbols used by human communication ? How can a robot acquire such a discrete abstract representation of the sensorimotor data of the continuous environment ?

Besides, from the point of view open-ended learning, learning algorithms benefit immensely from the use of symbolic methods for goal representation as they offer ways to structure knowledge for efficient and transferable learning. However, the most existing Hierarchical Reinforcement Learning approaches relying on symbolic reasoning are often limited as they require a manual goal representation. The challenge in autonomously discovering a symbolic goal representation is that it must preserve critical information, such as the environment dynamics. Recent works in hierarchical reinforcement learning [Vezhnevets et al., 2017; Nachum et al., 2019; Zhang et al., 2020; Li et al., 2022] have shown that learning an abstract goal representation is key to proposing semantically meaningful subgoals and to solving more complex tasks. In

particular, representations that capture environment dynamics over an abstract temporal scale have been shown to provide interesting properties with regards to bounding the suboptimality of learned policies under abstract goal spaces [Nachum et al., 2019; Abel et al., 2020], as well as efficiently handling continuous control problems. However, temporal abstractions that capture aspects of the environment dynamics [Ghosh et al., 2019; Savinov et al., 2019; Eysenbach et al., 2019; Zhang et al., 2020; Li et al., 2022] still cannot scale to environments where the pairwise state reachability relation is complex. This situation typically occurs when temporally abstract relations take into account more variables in the state space. The main limitation of these approaches is the lack of a spatial abstraction to generalise such relations over states.

Alternatively, other works [Kulkarni et al., 2016; Illanes et al., 2020; Garnelo et al., 2016]) have studied various forms of spatial abstractions for goal spaces. These abstractions effectively group states with similar roles in sets to construct a discrete goal space. The advantage of such representation is a smaller size exploration space that expresses large and long-horizon tasks. In contrast to these algorithms that require varying levels of prior knowledge, our proposed algorithm STAR [Zadem et al., 2024] gradually learns such spatial abstractions by considering reachability relations between sets of states. While these representations have been obtained in simple maze environments, we will scale these algorithms to robotic setups in higher dimensions and with stochasticity.

5.3 DIFFERENT FACTORS OF THE MOTIVATIONAL DRIVE OF LEARNING AGENTS TO INTERACT WITH TUTORS

In the Socially Guided Intrinsic Motivation (SGIM) framework, I proposed a common formulation of the intrinsic motivation valid both any sampling strategy, may it be autonomous exploration or for social guidance, and may the demonstrations requested to tutors be low-level policies, goals or decomposition into subgoals. Thus, the intrinsic motivation for a learning strategy σ and task ω can be modelled as

$$im(\sigma, \omega) = \kappa(\sigma) * progress(\sigma, \omega) \quad (5.1)$$

where *progress* is the empirical progress measured through the last episodes with strategy σ and goal ω , and $\kappa(\sigma)$ is the cost of the strategy. In the previous studies, this is was considered a constant set arbitrarily. However, this cost representing the availability of teachers, their willingness to interact with the robot ... but also other motivational drives of learners to interact with tutors.

To identify the motivational drives of learners to interact with tutors, I will analyse human behaviour in coaching situations in socially assistive robotics (SAR), by taking the use case of a robot coach for physical exercise. This work will take advantage of the framework laid in project Keraal, capitalising on the medical dataset collected during project Keraal. This analysis will enrich the formulation of the cost of each strategy, and will enable a more precise and automatic formulation of the cost coefficient as a function of the other motivational drives. It will enable us to generalise the formulation proposed, and sketch a common theory of the motivations to interact with tutors.

5.4 CONCLUSION

To tackle the problem of open-ended learning, my work in developmental robotics investigates the potential of social guidance to bootstrap the learning. My proposal provides a theoretical model enabling artificial agents to interact with tutors, by first investigating a prerequisite for efficient guidance : an appropriate representation of sequential tasks, and by second precisng the formulation of the motivation criteria based on human-robot interaction studies with the use case of an intelligent tutoring system for a robot coach. This research project is instrumental for study the notion of affordance. compositionality of actions, the co-development of tool-use and language, and the grounding of language. It will entail an alignment of AI to human representations.

6

Selected Articles

- Mehdi Zadem, Sergio Mover, and Sao Mai Nguyen. Reconciling spatial and temporal abstractions for goal representation. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=odY3PkI5VB>
- Sao Mai Nguyen, Maxime Devanne, Olivier Remy-Neris, Mathieu Lempereur, and Andre Thepaut. A medical low-back pain physical rehabilitation dataset for human body movement analysis. In *International Joint Conference on Neural Networks*, 2024
- Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit Leduc, and Ioannis Kanellos. Representation of irregularly sampled time series with generative language models for classification and transfer learning: a case study in activities of daily living recognition. In *Machine Learning for Irregular Time Series Workshop in ECML PKDD*, Turino, Italy, September 2023c. URL <https://hal.science/hal-04437409>
- Nicolas Duminy, Sao Mai Nguyen, Junshuai Zhu, Dominique Duhaut, and Jerome Kerdreux. Intrinsically motivated open-ended multi-task learning using transfer learning to discover task hierarchy. *Applied Sciences*, 11(3), 2021. ISSN 2076-3417. doi: 10.3390/app11030975. URL <https://inria.hal.science/hal-03118190>; <https://www.mdpi.com/2076-3417/11/3/975>
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Active choice of teachers, learning strategies and goals for a socially guided intrinsic motivation learner. *Paladyn Journal of Behavioural Robotics*, 3(3):136–146, 2012d. ISSN 2080-9778. doi: 10.2478/s13230-013-0110-z. URL <https://inria.hal.science/hal-00936932>

7

Publication list of Nguyen Sao Mai

Download available on <https://cv.hal.science/sao-mai-nguyen> and <http://nguyensmai.free.fr/Publications.html>

7.1 BOOK CHAPTER

1. André Thépaut, Sao Mai Nguyen, Maxime Devanne, and Olivier Rémy-Néris. Rééducation fonctionnelle assistée par robot humanoïde. In Serge Tisseron and Frédéric Tordo, editors, *Des robots au service des soignés, et des soignants*. Editions Eres, 2018. doi: 10.3917/eres.tisse.2018.02.0157. URL https://www.cairn.info/load_pdf.php?ID_ARTICLE=ERES_TISSE_2018_02_0157&download=1;https://hal.archives-ouvertes.fr/hal-03604132

7.2 THESIS

2. Sao Mai Nguyen. *A Curious Robot Learner for Interactive Goal-Babbling: Strategically Choosing What, How, When and from Whom to Learn*. PhD thesis, INRIA, 2013
3. Sao Mai Nguyen. Real-time face swapping based on head posture using particle filter towards the understanding of infant self-recognition. Master's thesis, Osaka University, 2010

7.3 EDITORIALS

4. Nicolás Navarro-Guerrero, Sao Mai Nguyen, Erhan Oztop, and Junpei Zhong. Guest editorial special issue on continual unsupervised sensorimotor learning. *IEEE Transactions on Cognitive and Developmental Systems*, 13(2):234–238, 2021. doi: 10.1109/TCDS.2021.3082880

7.4 PEER REVIEWED JOURNALS

5. Damien Bouchabou, Juliette Grosset, Sao Mai Nguyen, Christophe Lohr, and Xavier Puig. A smart home digital twin to support the recognition of activities of daily living. *Sensors*, 23(17):7586, sep 2023a. doi: 10.3390/s23177586 (*Impact Factor = 3.9, top 25 % en Information Systems*)
6. Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Ioannis Kanellos, and Benoit LeDuc. Reconnaissance d’activités de la vie quotidienne au moyen de capteurs domotiques et d’apprentissage profond: lorsque syntaxe, sémantique et contexte se rencontrent. *Revue Ouverte d’Intelligence Artificielle*, 4(1):129–156, 2023b. doi: 10.5802/roia.53. URL <https://roia.centre-mersenne.org/articles/10.5802/roia.53/>
7. Agathe Blanchard, Sao Mai Nguyen, Maxime Devanne, Mathieu Simonnet, Myriam Le Goff-Pronost, and Olivier Remy-Neris. Technical feasibility of supervision of stretching exercises by a humanoid robot coach for chronic low back pain: The r-cool randomized trial. *BioMed Research International*, 2022:1–10, mar 2022. doi: 10.1155/2022/5667223. URL <https://hal.archives-ouvertes.fr/hal-03606239>
8. Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit LeDuc, and Ioannis Kanellos. Using language model to bootstrap human activity recognition ambient sensors based in smart homes. *Electronics*, 10(20), 2021b. ISSN 2079-9292. doi: 10.3390/electronics10202498. URL <https://www.mdpi.com/2079-9292/10/20/2498>; <https://hal.archives-ouvertes.fr/hal-03381762> (*Impact Factor = 2.9*)
9. Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit LeDuc, and Ioannis Kanellos. A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning. *Sensors*, 21(18):6037, sep 2021c. doi: 10.3390/s21186037 (*Impact Factor = 3.9, top 25 % en Information Systems*)
10. Nicolas Duminy, Sao Mai Nguyen, Junshuai Zhu, Dominique Duhaut, and Jerome Kerdreux. Intrinsically motivated open-ended multi-task learning using transfer learning to discover task hierarchy. *Applied Sciences*, 11(3), 2021. ISSN 2076-3417. doi: 10.3390/app11030975. URL <https://inria.hal.science/hal-03118190>; <https://www.mdpi.com/2076-3417/11/3/975> (*Impact Factor = 2.474*)
11. Sao Mai Nguyen, Nicolas Duminy, Alexandre Manoury, Dominique Duhaut, and Cedric Buche. Robots learn increasingly complex tasks with intrinsic motivation and automatic curriculum learning. *KI - Künstliche Intelligenz*, 35:81–90, March 2021. doi: 10.1007/s13218-021-00708-8. URL <https://hal.archives-ouvertes.fr/hal-03157219>

12. Nicolas Duminy, Sao Mai Nguyen, and Dominique Duhaut. Learning a set of interrelated tasks by using a succession of motor policies for a socially guided intrinsically motivated learner. *Frontiers in Neurorobotics*, 12:87, 2019. ISSN 1662-5218. doi: 10.3389/fnbot.2018.00087. URL <https://hal.science/hal-02087995>; <https://www.frontiersin.org/article/10.3389/fnbot.2018.00087> (*Impact Factor = 2.6*)
13. Sao Mai Nguyen, Christophe Lohr, Philippe Tanguy, and Yiqiao Chen. Plug and play your robot into your smart home: Illustration of a new framework. *KI - Künstliche Intelligenz*, pages 1–7, 2017. ISSN 1610-1987. doi: 10.1007/s13218-017-0494-8. URL <http://dx.doi.org/10.1007/s13218-017-0494-8> (*h5 index = 15*)
14. Paul-Ariel Kenigsberg, Jean-Pierre Aquino, Alain Bérard, François Brémond, Kevin Charras, Tom Dening, Rose-Marie Dröes, Fabrice Gzil, Ben Hicks, Anthea Innes, Sao Mai Nguyen, Louise Nygård, Maribel Pino, Guillaume Sacco, Eric Salmon, Henriëtte van der Roest, Hervé Villet, Marion Villez, Philippe Robert, and Valeria Manera. Assistive technologies to address capabilities of people with dementia: from research to practice. *Dementia: the international journal of social research and practice*, 2017. doi: <https://doi.org/10.1177/1471301217714093> (*Impact Factor = 1.7*)
15. André Thépaut, Sao Mai Nguyen, and Christophe Lohr. Télé-rééducation fonctionnelle dans le cadre du projet VITAAL. *Actualités en médecine physique et de réadaptation : actualités en rééducation fonctionnelle et réadaptation*, (1):18–19, Mar 2017
16. Clément Moulin-Frier, Sao Mai Nguyen, and Pierre-Yves Oudeyer. Self-organization of early vocal development in infants and machines: The role of intrinsic motivation. *Frontiers in Psychology*, 4 (1006), 2014. ISSN 1664-1078. doi: 10.3389/fpsyg.2013.01006. URL <https://inria.hal.science/hal-00927940>; http://www.frontiersin.org/cognitive_science/10.3389/fpsyg.2013.01006/abstract (*Impact Factor = 1.1, top 25 % en psychologie*)
17. Sao Mai Nguyen and Pierre-Yves Oudeyer. Socially guided intrinsic motivation for robot learning of motor skills. *Autonomous Robots*, 36(3):273–294, 2014. doi: 10.1007/s10514-013-9339-y. URL <https://inria.hal.science/hal-00936938> (*Impact Factor = 3.7*)
18. Serena Ivaldi, Sao Mai Nguyen, Natalia Lyubova, Alain Droniou, Vincent Padois, David Filliat, Pierre-Yves Oudeyer, and Olivier Sigaud. Object learning through active exploration. *Transactions on Autonomous Mental Development*, PP(99):1–1, 2013. ISSN 1943-0604. doi: 10.1109/TAMD.2013.2280614. URL <https://hal.science/hal-00919694> (*Impact Factor = 5, top 25% en IA*)
19. Sao Mai Nguyen and Pierre-Yves Oudeyer. Active choice of teachers, learning strategies and goals for a socially guided intrinsic motivation learner. *Paladyn Journal of Behavioural Robotics*, 3(3):136–146, 2012d. ISSN 2080-9778. doi: 10.2478/s13230-013-0110-z. URL <https://inria.hal.science/hal-00936932> (*Impact Factor = 2.5*)

7.5 CONFERENCES WITH PROCEEDINGS

20. Mehdi Zadem, Sergio Mover, and Sao Mai Nguyen. Reconciling spatial and temporal abstractions for goal representation. In *The Twelfth International Conference on Learning Representations*, 2024.

- URL <https://openreview.net/forum?id=odY3PkI5VB> (rank A*)
21. Sao Mai Nguyen, Maxime Devanne, Olivier Remy-Neris, Mathieu Lempereur, and Andre Thepaut. A medical low-back pain physical rehabilitation dataset for human body movement analysis. In *International Joint Conference on Neural Networks*, 2024 (rank A)
 22. Louis Annabi, Ziqi Ma, and Sao Mai Nguyen. Unsupervised motion retargeting for human-robot imitation. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '24, pages 587–591. Association for Computing Machinery, March 2024. doi: 3610978.3640588. URL doi.acm.org?doi=3610978.3640588 (HRI is rank A*)
 23. Damien Bouchabou and Sao Mai Nguyen. *Generative Pretrained Embedding and Hierarchical Irregular Time Series Representation for Daily Living Activity Recognition*, pages 4764 – 4771. IOS Press, October 2024. ISBN 9781643685489. doi: 10.3233/faia241075. URL <http://dx.doi.org/10.3233/FAIA241075> accepted (ECAI is rank A)
 24. Aleksa Marusic, Sao Mai Nguyen, and Adriana Tapus. Physioformer: A spatio-temporal transformer for physical rehabilitation assessments. In *ICSR*, 2024
 25. Ivelina Stoyanova, Nicolas Museux, Sao Mai Nguyen, and David Filliat. Open the chests: An environment for activity recognition and sequential decision problems using temporal logic. In *LIPICs, TIME*, volume 318, 2024
 26. Qi Gan, Sao Mai Nguyen, Mounim A. El-Yacoubi, Eric Fenaux, and Stephan Clemencon. Human pose estimation based biomechanical feature extraction for long jumps. In *Human System Interaction*, 2024
 27. Mehdi Zadem, Sergio Mover, and Sao Mai Nguyen. Goal space abstraction in hierarchical reinforcement learning via set-based reachability analysis. In *2023 IEEE International Conference on Development and Learning (ICDL)*, pages 423–428, Nov 2023a. doi: 10.1109/ICDL55364.2023.10364473. URL <https://arxiv.org/abs/2309.07675>
 28. Louis Annabi and Sao Mai Nguyen. Prerequisite structure discovery in intelligent tutoring systems. In *2023 IEEE International Conference on Development and Learning (ICDL)*, pages 176–181, Nov 2023a. doi: 10.1109/ICDL55364.2023.10364416. URL <https://inria.hal.science/hal-04401880>
 29. Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit Leduc, and Ioannis Kanellos. Representation of irregularly sampled time series with generative language models for classification and transfer learning: a case study in activities of daily living recognition. In *Machine Learning for Irregular Time Series Workshop in ECML PKDD*, Turino, Italy, September 2023c. URL <https://hal.science/hal-04437409> (ECML is rank A)
 30. Aleksa Marusic, Louis Annabi, Sao Mai Nguyen, and Adriana Tapus. Analyzing data efficiency and performance of machine learning algorithms for assessing low back pain physical rehabilitation exercises. In ACM/IEEE, editor, *European Conference on Mobile Robots*, 2023a. doi: 10.1109/ECMR59166.2023.10256318. URL <https://ieeexplore.ieee.org/document/10256318>
 31. Aleksa Marusic, Sao Mai Nguyen, and Adriana Tapus. Evaluating kinect, openpose and blazepose for human body movement analysis on a low back pain physical rehabilitation dataset. In *Companion of*

- the 2023 ACM/IEEE International Conference on Human-Robot Interaction, HRI '23*, pages 587–591, New York, NY, USA, March 2023b. Association for Computing Machinery. ISBN 9781450399708. doi: 10.1145/3568294.3580153. URL <https://doi.org/10.1145/3568294.3580153> (HRI is rank A*)
32. Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit LeDuc, and Ioannis Kanellos. Fully convolutional network bootstrapped by word encoding and embedding for activity recognition in smart homes. In Xiaoli Li, Min Wu, Zhenghua Chen, and Le Zhang, editors, *Deep Learning for Human Activity Recognition*, pages 111–125, Singapore, February 2021a. Springer Singapore. ISBN 978-981-16-0575-8. doi: https://doi.org/10.1007/978-981-16-0575-8_9. URL <https://hal.archives-ouvertes.fr/hal-03032449> (IJCAI is rank A*)
 33. A. Mitriakov, P. Papadakis, Sao Mai Nguyen, and S. Garlatti. Staircase negotiation learning for articulated tracked robots with varying degrees of freedom. In *2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 394–400, Nov 2020. doi: 10.1109/SSRR50563.2020.9292594
 34. Andrei Mitriakov, Panagiotis Papadakis, Sao Mai Nguyen, and Serge Garlatti. Staircase traversal via reinforcement learning for active reconfiguration of assistive robots. In *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8, July 2020. doi: 10.1109/FUZZ48607.2020.9177581
 35. Maxime Devanne, Panagiotis Papadakis, and Sao Mai Nguyen. Recognition of activities of daily living via hierarchical long-short term memory networks. In *International Conference on Systems, Man and Cybernetics (SMC)*, pages 3318–3324. IEEE, July 2019. doi: 10.1109/SMC.2019.8914457. URL <https://hal-imt-atlantique.archives-ouvertes.fr/hal-02194928>
 36. Alexandre Manoury, Sao Mai Nguyen, and Cédric Buche. Hierarchical affordance discovery using intrinsic motivation. In *Proceedings of the 7th International Conference on Human-Agent Interaction, HAI '19*, pages 186–193, New York, NY, USA, 2019b. Association for Computing Machinery. ISBN 9781450369220. doi: 10.1145/3349537.3351898. URL <https://hal.science/hal-02283820>
 37. Alexandre Manoury, Sao Mai Nguyen, and Cédric Buche. Chime: An adaptive hierarchical representation for continuous intrinsically motivated exploration. In IEEE, editor, *International Robotics Conference*, pages 167–170, February 2019a. doi: 10.1109/IRC.2019.00032
 38. Nicolas Duminy, Sao Mai Nguyen, and Dominique Duhaut. Learning a set of interrelated tasks by using sequences of motor policies for a strategic intrinsically motivated learner. In *Proceedings of IEEE International Conference on Robotic Computing*, 2018c. doi: 10.1109/IRC.2018.00061. URL <https://ieeexplore.ieee.org/document/8329925>
 39. Maxime Devanne and Sao Mai Nguyen. Generating shared latent variables for robots to imitate human movements and understand their physical limitations. In Laura Leal-Taixé and Stefan Roth, editors, *Computer Vision – ECCV 2018 Workshops*, pages 190–197, Cham, 2019. Springer International Publishing. ISBN 978-3-030-11012-3. doi: 10.1007/978-3-030-11012-3_15 (ECCV is rank A*)
 40. Maxime Devanne, Sao Mai Nguyen, Olivier Remy-Neris, Béatrice Le Gales-Garnett, Gilles Kermarrec, and André Thepaut. A co-design approach for a rehabilitation robot coach for physical rehabilitation based on the error classification of motion errors. In *IEEE International Conference on Robotic*

- Computing (IRC)*, pages 352–357, Jan 2018. doi: 10.1109/IRC.2018.00074. URL <https://ieeexplore.ieee.org/document/8329938>
41. Maxime Devanne and Sao Mai Nguyen. Multi-level motion analysis for physical exercises assessment in kinaesthetic rehabilitation. In IEEE, editor, *International Conference on Humanoid Robots (Humanoids)*, 2017. doi: 10.1109/HUMANOIDS.2017.8246923. URL <https://hal.science/hal-01886757>; <https://ieeexplore.ieee.org/document/8246923>
 42. Nicolas Duminy, Sao Mai Nguyen, and Dominique Duhaut. Strategic and interactive learning of a hierarchical set of tasks by the Poppy humanoid robot. In *ICDL-EPIROB 2016 : 6th Joint IEEE International Conference Developmental Learning and Epigenetic Robotics*, pages 204–209, Sep 2016. doi: 10.1109/DEVLRN.2016.7846820. URL <https://hal.science/hal-01547074>
 43. Sao Mai Nguyen, Philippe Tanguy, and Olivier Rémy-Néris. Computational Architecture of a Robot Coach for Physical Exercises in Kinesthetic Rehabilitation. In *RO-MAN 2016 : IEEE International Symposium on Human and Robot Interactive Communication*, pages 1138–1143, Aug 2016. doi: 10.1109/ROMAN.2016.7745251
 44. Nicolas Duminy, Sao Mai Nguyen, and Dominique Duhaut. Strategic and interactive learning of a hierarchical set of tasks by the Poppy humanoid robot. In *ICDL-EPIROB 2016 : 6th Joint IEEE International Conference Developmental Learning and Epigenetic Robotics*, pages 204–209, Sep 2016. doi: 10.1109/DEVLRN.2016.7846820. URL <https://hal.science/hal-01547074>
 45. Sao Mai Nguyen and Pierre-Yves Oudeyer. Strategic robot learner for interactive goal-babbling. In *Reinforcement Learning and Decision Making*, Princeton, New Jersey, U.S.A, 2013a
 46. Sao Mai Nguyen, Serena Ivaldi, Natalia Lyubova, Alain Droniou, Damien Gerardeaux-Viret, David Filliat, Vincent Padois, Olivier Sigaud, and Pierre-Yves Oudeyer. Learning to recognize objects through curiosity-driven manipulation with the icub humanoid robot. In *IEEE International Conference on Development and Learning - Epirob*, number 1-8, 2013. doi: 10.1109/DevLrn.2013.6652525. URL <https://hal.science/hal-00919674> :
 47. Sao Mai Nguyen and Pierre-Yves Oudeyer. Properties for efficient demonstrations to a socially guided intrinsically motivated learner. In *21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012a. doi: 10.1109/ROMAN.2012.6343819. URL <https://inria.hal.science/hal-00762758>
 48. Sao Mai Nguyen and Pierre-Yves Oudeyer. Whom will an intrinsically motivated robot learner choose to imitate from? In Joanna Szufnarowska, editor, *Proceedings of the Post-Graduate Conference on Robotics and Development of Cognition*, pages 32–35, September 2012e. doi: 10.2390/biecoll-robotdoc2012-12
 49. Sao Mai Nguyen and Pierre-Yves Oudeyer. Interactive learning gives the tempo to an intrinsically motivated robot learner. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, pages 645–652, 2012c. doi: 10.1109/HUMANOIDS.2012.6651588
 50. Sao Mai Nguyen, Adrien Baranes, and Pierre-Yves Oudeyer. Bootstrapping intrinsically motivated learning with human demonstrations. In *IEEE International Conference on Development and Learning*, volume 2, pages 1–8, Frankfurt, Germany, Aug 2011a. doi: 10.1109/DEVLRN.2011.6037329

51. Sao Mai Nguyen, Adrien Baranes, and Pierre-Yves Oudeyer. Constraining the size growth of the task space with socially guided intrinsic motivation using demonstrations. In *IJCAI Workshop on Agents Learning Interactively from Human Teachers*, 2011b (IJCAI is rank A*)
52. Sao Mai Nguyen, Masaki Ogino, and Minoru Asada. Real-time face swapping as a tool for understanding infant self-recognition. In *Proceedings of the 10th International Conference on Epigenetic Robotics*, pages pp.171–172, Glumslov, Sweden, 2010

7.6 INTERNATIONAL CONFERENCES WITH PEER REVIEW

53. Zihe Ji, Sao Mai Nguyen, and Mehdi Zadem. Bridging natural language and emergent representation in hierarchical reinforcement learning. In *Intrinsically-Motivated and Open-Ended Learning Workshop @NeurIPS2024*, 2024a. URL <https://openreview.net/forum?id=HMJJ10SAFy>
54. Zihe Ji, Mehdi Zadem, and Sao Mai Nguyen. Llm to bridge human instructions with a dynamic symbolic representation in hierarchical reinforcement learning. In *NeurIPS 2024 Workshop on Behavioral Machine Learning*, 2024b. URL <https://openreview.net/forum?id=GNVfcHACC6>
55. Zihe Ji, Mehdi Zadem, and Sao Mai Nguyen. Dynamic symbolic representation and LLM to enhance task abstraction in hierarchical reinforcement learning. In *NeurIPS 2024 Workshop on Compositional Learning: Perspectives, Methods, and Paths Forward*, 2024c. URL <https://openreview.net/forum?id=DM6q3YPWqG>
56. Mehdi Zadem, Sergio Mover, and Sao Mai Nguyen. Emergence of a symbolic goal representation with an intelligent tutoring system based on intrinsic motivation. In *Intrinsically Motivated and Open-Ended Learning Workshop @NeurIPS2023*, 2023c. URL <https://openreview.net/forum?id=UawMOafW3f> (Neurips is rank A*)
57. Sao Mai Nguyen. Online learning a symbolic abstraction of actions in hierarchical rl with formal methods. In *ActSynt at ECAI*, 2024 (ECAI is rank A)
58. Louis Annabi and Sao Mai Nguyen. Prerequisite structure discovery for an intelligent tutoring system based on intrinsic motivation. In *Intrinsically Motivated Open-ended Learning*, Paris, France, September 2023b. URL <https://hal.science/hal-04201624v1>
59. Mehdi Zadem, Sergio Mover, and Sao Mai Nguyen. Goal space abstraction in hierarchical reinforcement learning via reachability analysis. In *Intrinsically Motivated Open-ended Learning*, Tübingen, Germany, September 2023b. URL <https://hal.science/hal-03600799>
60. Qi Gan, Sao Mai Nguyen, Eric Fenaux, Stephan Clemencon, Mounim A. El-Yacoubi, and Ons Jelassi. Reconstruction of trajectories of athletes using computer vision models and kinetic analysis. In *MIT Sloan Sports Analytics Conference*, 2023
61. Mehdi Zadem, Sergio Mover, Sao Mai Nguyen, and Sylvie Putot. Towards automata-based abstraction of goals in hierarchical reinforcement learning. In *Intrinsically Motivated Open-ended Learning IMOL 2022*, Tübingen, Germany, April 2022. URL <https://hal.science/hal-03600799>

62. Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit Leduc, and Ioannis Kanellos. What if time, or past context, improved human activity recognition in smart homes? In *IEEE ICCE*, Da Nang, Viet Nam, 2022
63. Linda Vallee, Sao Mai Nguyen, Christophe Lohr, Ioannis Kanellos, and Olivier Asseu. How an automated gesture imitation game can improve social interactions with teenagers with asd. In *IEEE ICRA workshop on Social Robotics for Neurodevelopmental Disorders*, 06 2020. doi: 10.13140/RG.2.2.10961.25444. URL <https://icra2020ndd.wordpress.com>
64. Sao Mai Nguyen, Nathalie Collot-Lavenne, Christophe Lohr, Sebastien Guillon, Patricio Tula, Alvaro Paez, Mouad Bouaida, Arthus Anin, and Saad El Qacemi. An implementation of an imitation game with asd children to learn nursery rhymes. In *HAI Workshop on Clinical Use of Technology for Individuals with Autism Spectrum Disorder*, 2019. URL <https://hal.science/hal-02499990>
65. Andrei Mitriakov, Panagiotis Papadakis, Sao Mai Nguyen, and Serge Garlatti. Learning-based modelling of physical interaction for assistive robots. Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JFPDA), July 2019. URL <https://hal-imt-atlantique.archives-ouvertes.fr/hal-02341202>. Poster
66. Nicolas Duminy, Alexandre Manoury, Sao Mai Nguyen, Cédric Buche, and Dominique Duhaut. Learning sequences of policies by using an intrinsically motivated learner and a task hierarchy. In *Workshop on Continual Unsupervised Sensorimotor Learning, ICDL-EpiRob*, Tokyo, Japan, September 2018a. URL <https://hal.archives-ouvertes.fr/hal-01887073>;<https://youtu.be/US84HjUuUtg>
67. Sao Mai Nguyen and Philippe Tanguy. Cognitive architecture of a humanoid robot for coaching physical exercises in kinaesthetic rehabilitation. In *International Workshop on Cognitive Robotics*, 2016
68. Sao Mai Nguyen and Pierre-Yves Oudeyer. Data collection of an interactive learner for strategy, outcome and policy exploration. In *Workshop in Hierarchical Learning at Robotic Science and Systems*, Berlin, Germany, 2013c (RSS is rank A)
69. Sao Mai Nguyen and Pierre-Yves Oudeyer. Strategic robot learner for active goal-babbling. In *Workshop in Active Learning in Robotics at Robotic Science and Systems*, Berlin, Germany, 2013b (RSS is rank A)
70. Clement Moulin-Frier, Sao Mai Nguyen, and Pierre-Yves Oudeyer. The role of intrinsic motivations in early vocal development: a computational study. In *Workshop on Intrinsic Motivations*, Rome, Italy, June 2013
71. Sao Mai Nguyen and Pierre-Yves Oudeyer. Socially guided intrinsically motivated learner. In *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on*, pages 1 –2, nov. 2012b. doi: 10.1109/DevLrn.2012.6400809
72. Sao Mai Nguyen and Pierre-Yves Oudeyer. Socially guided intrinsic motivation. In *Spring School on Developmental Robotics and Cognitive Bootstrapping*, 2012f

73. Sao Mai Nguyen. Role of default network in the understanding of subjectivity of desire. In *NIPS International Workshop for Scientific Study of Consciousness at Neuroscience*, volume NIPS International Workshop for Scientific Study of Consciousness at Neuroscience, 2009

References

- David Abel, Nate Umbanhowar, Khimya Khetarpal, Dilip Arumugam, Doina Precup, and Michael Littman. Value preserving state-action abstractions. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 1639–1650. PMLR, 26–28 Aug 2020. URL <https://proceedings.mlr.press/v108/abel20a.html>.
- P. Andry, P. Gaussier, S. Moga, J.P. Banquet, and J. Nadel. Learning and communication via imitation: an autonomous robot perspective. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 31(5):431–442, Sep. 2001. ISSN 1558-2426. doi: 10.1109/3468.952717.
- Louis Annabi and Sao Mai Nguyen. Prerequisite structure discovery in intelligent tutoring systems. In *2023 IEEE International Conference on Development and Learning (ICDL)*, pages 176–181, Nov 2023a. doi: 10.1109/ICDL55364.2023.10364416. URL <https://inria.hal.science/hal-04401880>.
- Louis Annabi and Sao Mai Nguyen. Prerequisite structure discovery for an intelligent tutoring system based on intrinsic motivation. In *Intrinsically Motivated Open-ended Learning*, Paris, France, September 2023b. URL <https://hal.science/hal-04201624v1>.
- Louis Annabi, Ziqi Ma, and Sao Mai Nguyen. Unsupervised motion retargeting for human-robot imitation. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, HRI '24*, pages 587–591. Association for Computing Machinery, March 2024. doi: 3610978.3640588. URL doi.acm.org?doi=3610978.3640588.
- Ilktan Ar and Yusuf Akgul. A computerized recognition system for the home-based physiotherapy exercises using an rgbd camera. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22, 05 2014. doi: 10.1109/TNSRE.2014.2326254.
- M. Arbib. *Handbook of Physiology: The Nervous System, II. Motor Control*, chapter Perceptual structures and distributed motor control, pages pp. 1448–1480. MIT Press, Cambridge, MA, USA, 1981.
- Brenna D. Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469 – 483, 2009. ISSN 0921-8890. doi: 10.1016/j.robot.2008.10.024. URL <http://www.sciencedirect.com/science/article/pii/S0921889008001772>.
- Brenna D. Argall, B. Browning, and Manuela Veloso. Teacher feedback to scaffold and refine demonstrated motion primitives on a mobile robot. *Robotics and Autonomous Systems*, 59(3-4):243–255, 2011.
- Christian Arzate Cruz and Takeo Igarashi. A survey on interactive reinforcement learning: Design principles and open challenges. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference, DIS '20*,

- pages 1195–1209, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450369749. doi: 10.1145/3357236.3395525. URL <https://doi.org/10.1145/3357236.3395525>.
- Minoru Asada, Karl F. MacDorman, Hiroshi Ishiguro, and Yasuo Kuniyoshi. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems*, 37(2-3):185–193, 2001. ISSN 09218890. doi: 10.1016/S0921-8890(01)00157-9.
- Minoru Asada, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development*, 1(1):12–34, 2009. URL <http://www.er.ams.eng.osaka-u.ac.jp/Paper/paper-jp.html>.
- Parviz Asghari, Elnaz Soelimani, and Ehsan Nazerfard. Online human activity recognition employing hierarchical hidden markov models, 2019.
- M. S. H. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, M. Valstar, H. Meng, A. Kemp, M. Shafizadeh, A. C. Elkins, N. Kanakam, A. de Rothschild, N. Tyler, P. J. Watson, A. C. d. C. Williams, M. Pantic, and N. Bianchi-Berthouze. The automatic detection of chronic pain-related expression: Requirements, challenges and the multimodal emopain dataset. *IEEE Transactions on Affective Computing*, 7(4):435–451, Oct 2016. ISSN 1949-3045. doi: 10.1109/TAFFC.2015.2462830.
- Albert Bandura. Social learning theory, 1971.
- Andrew G. Barto and Sridhar Mahadevan. Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems*, 13(1):41–77, Jan 2003. ISSN 1573-7594. doi: 10.1023/A:1022140919877. URL <https://doi.org/10.1023/A:1022140919877>.
- Andrew G Barto, George Konidaris, and Christopher Vigorito. Behavioral hierarchy: exploration and representation. In *Computational and robotic models of the hierarchical organization of behavior*, pages 13–46. Springer, 2013.
- Katarina Begus and Victoria Southgate. *Active Learning from Infancy to Childhood*, chapter Curious Learners: How Infants’ Motivation to Learn Shapes and Is Shaped by Infants’ Interactions with the Social World, pages 13–37. Springer International Publishing, Cham, 2018. ISBN 978-3-319-77182-3. doi: 10.1007/978-3-319-77182-3_2. URL https://doi.org/10.1007/978-3-319-77182-3_2.
- Katarina Begus, Teodora Gliga, and Victoria Southgate. Infants’ preferences for native speakers are associated with an expectation of information. *Proceedings of the National Academy of Sciences*, 113(44):12397–12402, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1603261113. URL <http://www.pnas.org/content/113/44/12397>.
- Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML ’09*, pages 41–48, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-516-1. doi: 10.1145/1553374.1553380. URL <http://doi.acm.org/10.1145/1553374.1553380>.
- Aude Billard and R. Siegward. Robot learning from demonstration. *Robotics and Autonomous Systems*, 47(2-3):65–67, June 2004.
- Aude Billard, Sylvain Calinon, Ruediger Dillmann, and Stefan Schaal. *Handbook of Robotics*, chapter Robot Programming by Demonstration. Number 59. MIT Press, 2007.
- Agathe Blanchard, Sao Mai Nguyen, Maxime Devanne, Mathieu Simonnet, Myriam Le Goff-Pronost, and

- Olivier Remy-Neris. Technical feasibility of supervision of stretching exercises by a humanoid robot coach for chronic low back pain: The r-cool randomized trial. *BioMed Research International*, 2022:1–10, mar 2022. doi: 10.1155/2022/5667223. URL <https://hal.archives-ouvertes.fr/hal-03606239>.
- Damien Bouchabou and Sao Mai Nguyen. *Generative Pretrained Embedding and Hierarchical Irregular Time Series Representation for Daily Living Activity Recognition*, pages 4764 – 4771. IOS Press, October 2024. ISBN 9781643685489. doi: 10.3233/faia241075. URL <http://dx.doi.org/10.3233/FAIA241075>.
- Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit LeDuc, and Ioannis Kanellos. Fully convolutional network bootstrapped by word encoding and embedding for activity recognition in smart homes. In Xiaoli Li, Min Wu, Zhenghua Chen, and Le Zhang, editors, *Deep Learning for Human Activity Recognition*, pages 111–125, Singapore, February 2021a. Springer Singapore. ISBN 978-981-16-0575-8. doi: https://doi.org/10.1007/978-981-16-0575-8_9. URL <https://hal.archives-ouvertes.fr/hal-03032449>.
- Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit LeDuc, and Ioannis Kanellos. Using language model to bootstrap human activity recognition ambient sensors based in smart homes. *Electronics*, 10(20), 2021b. ISSN 2079-9292. doi: 10.3390/electronics10202498. URL <https://www.mdpi.com/2079-9292/10/20/2498>; <https://hal.archives-ouvertes.fr/hal-03381762>.
- Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit LeDuc, and Ioannis Kanellos. A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning. *Sensors*, 21(18):6037, sep 2021c. doi: 10.3390/s21186037.
- Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit Leduc, and Ioannis Kanellos. What if time, or past context, improved human activity recognition in smart homes? In *IEEE ICCE*, Da Nang, Viet Nam, 2022.
- Damien Bouchabou, Juliette Grosset, Sao Mai Nguyen, Christophe Lohr, and Xavier Puig. A smart home digital twin to support the recognition of activities of daily living. *Sensors*, 23(17):7586, sep 2023a. doi: 10.3390/s23177586.
- Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Ioannis Kanellos, and Benoit LeDuc. Reconnaissance d’activités de la vie quotidienne au moyen de capteurs domotiques et d’apprentissage profond: lorsque syntaxe, sémantique et contexte se rencontrent. *Revue Ouverte d’Intelligence Artificielle*, 4(1):129–156, 2023b. doi: 10.5802/roia.53. URL <https://roia.centre-mersenne.org/articles/10.5802/roia.53/>.
- Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit Leduc, and Ioannis Kanellos. Representation of irregularly sampled time series with generative language models for classification and transfer learning: a case study in activities of daily living recognition. In *Machine Learning for Irregular Time Series Workshop in ECML PKDD*, Turino, Italy, September 2023c. URL <https://hal.science/hal-04437409>.
- Cynthia Breazeal and B. Scassellati. Robots that imitate humans. *Trends in Cognitive Sciences*, 6(11): 481–487, 2002.
- R. A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–160, 1991.
- Sylvain Calinon. *Robot Programming by Demonstration: A Probabilistic Approach*. EPFL/CRC Press, 2009. EPFL Press ISBN 978-2-940222-31-5, CRC Press ISBN 978-1-4398-0867-2.
- J. Call and M. Carpenter. *Imitation in animals and artifacts*, chapter Three sources of information in social learning, pages 211–228. Cambridge, MA: MIT Press., 2002.

- Angelo Cangelosi and Minoru Asada. *Cognitive Robotics*. The MIT Press, 2022. doi: <https://doi.org/10.7551/mitpress/13780.001.0001>.
- Angelo Cangelosi and Matthew Schlesinger. *Developmental robotics: From babies to robots*. MIT press, Cambridge, MA, USA, 2015. URL <https://idoc.pub/documents/developmental-robotics-from-babies-to-robots-d47ey87vq2n2>.
- Marianna Capecci, Maria Gabriella Ceravolo, Francesco Ferracuti, Sabrina Iarlori, Andrea Monteriù, Luca Romeo, and Federica Verdini. The kimore dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(7):1436–1448, July 2019. ISSN 1558-0210. doi: 10.1109/TNSRE.2019.2923060.
- Sonia Chernova and Manuela Veloso. Interactive policy learning through confidence-based autonomy. *Journal of Artificial Intelligence Research*, 34(1):1, 2009.
- Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf.
- Jeffrey Allen Clouse. *On Integrating Apprentice Learning and Reinforcement Learning*. PhD thesis, University of Massachusetts Amherst, 1996.
- D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. *J. Artif. Intell. Res.*, 4:129–145, 1996.
- Cédric Colas, Pierre Fournier, Mohamed Chetouani, Olivier Sigaud, and Pierre-Yves Oudeyer. CURIOUS: Intrinsically motivated modular multi-goal reinforcement learning. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1331–1340, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL <http://proceedings.mlr.press/v97/colas19a.html>.
- Diane J Cook, Aaron S Crandall, Brian L Thomas, and Narayanan C Krishnan. Casas: A smart home in a box. *Computer*, 46(7):62–69, 2012.
- Gergely Csibra. Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431):447, 2003.
- Kerstin Dautenhahn and Chrystopher L. Nehaniv. *Imitation in Animals and Artifacts*. MIT Press, 2002.
- Kerstin Dautenhahn, Chrystopher L Nehaniv, Michael L Walters, Ben Robins, Hatice Kose-Bagci, N Assif Mirza, and Mike Blow. Kaspar—a minimally expressive humanoid robot for human–robot interaction research. *Applied Bionics and Biomechanics*, 6(3-4):369–397, 2009.
- Peter Dayan and Geoffrey E Hinton. Feudal reinforcement learning. In S. Hanson, J. Cowan, and C. Giles, editors, *Advances in Neural Information Processing Systems*, volume 5. Morgan-Kaufmann, 1992. URL <https://proceedings.neurips.cc/paper/1992/file/d14220ee66aeec73c49038385428ec4c-Paper.pdf>.
- Maxime Devanne and Sao Mai Nguyen. Multi-level motion analysis for physical exercises assessment in kinaesthetic rehabilitation. In IEEE, editor, *International Conference on Humanoid Robots (Humanoids)*,

2017. doi: 10.1109/HUMANOIDS.2017.8246923. URL <https://hal.science/hal-01886757>;<https://ieeexplore.ieee.org/document/8246923>.
- Maxime Devanne and Sao Mai Nguyen. Generating shared latent variables for robots to imitate human movements and understand their physical limitations. In Laura Leal-Taixé and Stefan Roth, editors, *Computer Vision – ECCV 2018 Workshops*, pages 190–197, Cham, 2019. Springer International Publishing. ISBN 978-3-030-11012-3. doi: 10.1007/978-3-030-11012-3_15.
- Maxime Devanne, Sao Mai Nguyen, Olivier Remy-Neris, Béatrice Le Gales-Garnett, Gilles Kermarrec, and André Thepaut. A co-design approach for a rehabilitation robot coach for physical rehabilitation based on the error classification of motion errors. In *IEEE International Conference on Robotic Computing (IRC)*, pages 352–357, Jan 2018. doi: 10.1109/IRC.2018.00074. URL <https://ieeexplore.ieee.org/document/8329938>.
- Maxime Devanne, Panagiotis Papadakis, and Sao Mai Nguyen. Recognition of activities of daily living via hierarchical long-short term memory networks. In *International Conference on Systems, Man and Cybernetics (SMC)*, pages 3318–3324. IEEE, July 2019. doi: 10.1109/SMC.2019.8914457. URL <https://hal-imt-atlantique.archives-ouvertes.fr/hal-02194928>.
- Elham Dolatabadi, Ying Xuan Zhi, Bing Ye, Marge Coahran, Giorgia Lupinacci, Alex Mihailidis, Rosalie Wang, and Babak Taati. The toronto rehab stroke pose dataset to detect compensation during stroke rehabilitation therapy. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '17, pages 375–381, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450363631. doi: 10.1145/3154862.3154925. URL <https://www.kaggle.com/derekdb/toronto-robot-stroke-posture-dataset?>
- Stephane Doncieux, David Filliat, Natalia Díaz-Rodríguez, Timothy Hospedales, Richard Duro, Alexandre Coninx, Diederik M. Roijers, Benoît Girard, Nicolas Perrin, and Olivier Sigaud. Open-ended learning: A conceptual framework based on representational redescription. *Frontiers in Neurorobotics*, 12, sep 2018. doi: 10.3389/fnbot.2018.00059. URL <https://doi.org/10.3389/fnbot.2018.00059>.
- Nicolas Duminy, Sao Mai Nguyen, and Dominique Duhaut. Strategic and interactive learning of a hierarchical set of tasks by the Poppy humanoid robot. In *ICDL-EPIROB 2016 : 6th Joint IEEE International Conference Developmental Learning and Epigenetic Robotics*, pages 204–209, Sep 2016. doi: 10.1109/DEVLRN.2016.7846820. URL <https://hal.science/hal-01547074>.
- Nicolas Duminy, Alexandre Manoury, Sao Mai Nguyen, Cédric Buche, and Dominique Duhaut. Learning sequences of policies by using an intrinsically motivated learner and a task hierarchy. In *Workshop on Continual Unsupervised Sensorimotor Learning, ICDL-EpiRob*, Tokyo, Japan, September 2018a. URL <https://hal.archives-ouvertes.fr/hal-01887073>;<https://youtu.be/US84HjUuUtg>.
- Nicolas Duminy, Sao Mai Nguyen, and Dominique Duhaut. Effects of social guidance on a robot learning sequences of policies in hierarchical learning. In IEEE, editor, *International Conference on Systems Man and Cybernetics*, 2018b. doi: 10.1109/SMC.2018.00636. URL <https://hal.science/hal-02066931>.
- Nicolas Duminy, Sao Mai Nguyen, and Dominique Duhaut. Learning a set of interrelated tasks by using sequences of motor policies for a strategic intrinsically motivated learner. In *Proceedings of IEEE International Conference on Robotic Computing*, 2018c. doi: doi:10.1109/IRC.2018.00061. URL <https://ieeexplore.ieee.org/document/8329925>.

- Nicolas Duminy, Sao Mai Nguyen, and Dominique Duhaut. Learning a set of interrelated tasks by using a succession of motor policies for a socially guided intrinsically motivated learner. *Frontiers in Neurobotics*, 12:87, 2019. ISSN 1662-5218. doi: 10.3389/fnbot.2018.00087. URL <https://hal.science/hal-02087995>; <https://www.frontiersin.org/article/10.3389/fnbot.2018.00087>.
- Nicolas Duminy, Sao Mai Nguyen, Junshuai Zhu, Dominique Duhaut, and Jerome Kerdreux. Intrinsically motivated open-ended multi-task learning using transfer learning to discover task hierarchy. *Applied Sciences*, 11(3), 2021. ISSN 2076-3417. doi: 10.3390/app11030975. URL <https://inria.hal.science/hal-03118190>; <https://www.mdpi.com/2076-3417/11/3/975>.
- Maria K Eckstein and Anne G E Collins. How the mind creates structure: Hierarchical learning of action sequences. In Cognitive Science Society, editor, *CogSci Conference of the Cognitive Science Society*, volume 43, pages 618–624, 2021.
- Ben Eysenbach, Ruslan Salakhutdinov, and Sergey Levine. Search on the replay buffer: Bridging planning and reinforcement learning. In *NeurIPS*, pages 15220–15231, 2019.
- Jacqueline Fagard, Lauriane Rat-Fischer, Rana Esseily, Eszter Somogyi, and J. K. O’Regan. What does it take for an infant to learn how to use a tool by observation? *Frontiers in Psychology*, 7, March 2016. ISSN 1664-1078. doi: 10.3389/fpsyg.2016.00267. URL <http://dx.doi.org/10.3389/fpsyg.2016.00267>.
- V. Fedorov. *Theory of Optimal Experiment*. Academic Press, Inc., New York, NY, 1972.
- Pierre Fournier, Cédric Colas, Olivier Sigaud, and Mohamed Chetouani. Clic: Curriculum learning and imitation for object control in non-rewarding environments. *IEEE Transactions on Cognitive and Developmental Systems*, pages 1–1, 2019. ISSN 2379-8920. doi: 10.1109/TCDS.2019.2933371. URL <https://arxiv.org/abs/1901.09720>.
- Qi Gan, Sao Mai Nguyen, Eric Fenaux, Stephan Clemencon, Mounim A. El-Yacoubi, and Ons Jelassi. Reconstruction of trajectories of athletes using computer vision models and kinetic analysis. In *MIT Sloan Sports Analytics Conference*, 2023.
- Qi Gan, Sao Mai Nguyen, Mounim A. El-Yacoubi, Eric Fenaux, and Stephan Clemencon. Human pose estimation based biomechanical feature extraction for long jumps. In *Human System Interaction*, 2024.
- Marta Garnelo, Kai Arulkumaran, and Murray Shanahan. Towards deep symbolic reinforcement learning, 2016.
- Dibya Ghosh, Abhishek Gupta, and Sergey Levine. Learning actionable representations with goal conditioned policies. In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=Hye9lnCct7>.
- J.J. Gibson. *The Ecological Approach to Visual Perception*, chapter The Theory of Affordances, pages 67–82. Houghton Mifflin, Boston, MA, USA, 1979.
- Jacqueline Gottlieb, Pierre-Yves Oudeyer, Manuel Lopes, and Adrien Baranes. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, 17(11):585–593, 10 2013. doi: 10.1016/j.tics.2013.09.001. URL <https://doi.org/10.1016/j.tics.2013.09.001>.
- Scott T. Grafton and Antonia F. de C. Hamilton. Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, 26(4):590–616, 2007. ISSN 0167-9457. doi: <https://doi.org/10.1016/j.humov.2007.05.009>. URL <https://www.sciencedirect.com/science/article/pii/S0167945707000504>. European Workshop on Movement Science 2007.

- Xin Hong, Chris Nugent, Maurice Mulvenna, Sally McClean, Bryan Scotney, and Steven Devlin. Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive and Mobile Computing*, 5(3):236 – 252, 2009. ISSN 1574-1192. doi: <https://doi.org/10.1016/j.pmcj.2008.05.002>. URL <http://www.sciencedirect.com/science/article/pii/S157411920800045X>. Pervasive Health and Wellness Management.
- León Illanes, Xi Yan, Rodrigo Toro Icarte, and Sheila A. McIlraith. Symbolic plans as high-level instructions for reinforcement learning. *Proceedings of the International Conference on Automated Planning and Scheduling*, 30:540–550, jun 2020. doi: 10.1609/icaps.v30i1.6750.
- Serena Ivaldi, Sao Mai Nguyen, Natalia Lyubova, Alain Droniou, Vincent Padois, David Filliat, Pierre-Yves Oudeyer, and Olivier Sigaud. Object learning through active exploration. *Transactions on Autonomous Mental Development*, PP(99):1–1, 2013. ISSN 1943-0604. doi: 10.1109/TAMD.2013.2280614. URL <https://hal.science/hal-00919694>.
- Zihe Ji, Sao Mai Nguyen, and Mehdi Zadem. Bridging natural language and emergent representation in hierarchical reinforcement learning. In *Intrinsically-Motivated and Open-Ended Learning Workshop @NeurIPS2024*, 2024a. URL <https://openreview.net/forum?id=HMJJ10SAFy>.
- Zihe Ji, Mehdi Zadem, and Sao Mai Nguyen. Llm to bridge human instructions with a dynamic symbolic representation in hierarchical reinforcement learning. In *NeurIPS 2024 Workshop on Behavioral Machine Learning*, 2024b. URL <https://openreview.net/forum?id=GNVfcHACC6>.
- Zihe Ji, Mehdi Zadem, and Sao Mai Nguyen. Dynamic symbolic representation and LLM to enhance task abstraction in hierarchical reinforcement learning. In *NeurIPS 2024 Workshop on Compositional Learning: Perspectives, Methods, and Paths Forward*, 2024c. URL <https://openreview.net/forum?id=DM6q3YPWqG>.
- Mark Johnson. *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press, 1987.
- Anne-Lise Jouen, Antonio Narzisi, Jean Xavier, Elodie Tilmont, Nicolas Bodeau, Valentina Bono, Nabila Ketem-Premel, Salvatore Anzalone, Koushik Maharatna, Mohamed Chetouani, Filippo Muratori, and David Cohen. Goliah (gaming open library for intervention in autism at home): a 6-month single blind matched controlled exploratory study. *Child and Adolescent Psychiatry and Mental Health*, 11(1):17, Mar 2017. ISSN 1753-2000. doi: 10.1186/s13034-017-0154-7. URL <https://doi.org/10.1186/s13034-017-0154-7>.
- Kshitij Judah, Alan P. Fern, Thomas G. Dietterich, and Prasad Tadepalli. Active imitation learning: Formal and practical reductions to i.i.d. learning. *Journal of Machine Learning Research*, 15(120):4105–4143, 2014. URL <http://jmlr.org/papers/v15/judah14a.html>.
- Masatoshi Katagiri, Naoko Inada, and Yoko Kamio. Mirroring effect in 2- and 3-year-olds with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 4(3):474 – 478, 2010. ISSN 1750-9467. doi: <https://doi.org/10.1016/j.rasd.2009.11.004>. URL <http://www.sciencedirect.com/science/article/pii/S1750946709001354>.
- Sidney Katz. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. *Journal of the American Geriatrics Society*, 31(12):721–727, 1983.
- Paul-Ariel Kenigsberg, Jean-Pierre Aquino, Alain Bérard, François Brémond, Kevin Charras, Tom Dening,

- Rose-Marie Droës, Fabrice Gzil, Ben Hicks, Anthea Innes, Sao Mai Nguyen, Louise Nygård, Maribel Pino, Guillaume Sacco, Eric Salmon, Henriëtte van der Roest, Hervé Villet, Marion Villez, Philippe Robert, and Valeria Manera. Assistive technologies to address capabilities of people with dementia: from research to practice. *Dementia: the international journal of social research and practice*, 2017. doi: <https://doi.org/10.1177/1471301217714093>.
- W. Bradley Knox, Peter Stone, and Cynthia Breazeal. *Training a Robot via Human Feedback: A Case Study*, pages 460–470. Springer International Publishing, 2013. ISBN 9783319026756. doi: 10.1007/978-3-319-02675-6_46. URL http://dx.doi.org/10.1007/978-3-319-02675-6_46.
- Hideki Kozima, Marek P Michalowski, and Cocoro Nakagawa. Kepon. *International Journal of Social Robotics*, 1(1):3–18, 2009.
- John W. Krakauer and Reza Shadmehr. Towards a computational neuropsychology of action. In Paul Cisek, Trevor Drew, and John F. Kalaska, editors, *Computational Neuroscience: Theoretical Insights into Brain Function*, volume 165 of *Progress in Brain Research*, pages 383 – 394. Elsevier, 2007. doi: [https://doi.org/10.1016/S0079-6123\(06\)65024-3](https://doi.org/10.1016/S0079-6123(06)65024-3). URL <http://www.sciencedirect.com/science/article/pii/S0079612306650243>.
- Thibaut Kulak and Sylvain Calinon. Combining social and intrinsically motivated learning for multitask robot skill acquisition. *IEEE Transactions on Cognitive and Developmental Systems*, 15(2):385–394, June 2023. ISSN 2379-8939. doi: 10.1109/TCDS.2021.3069341.
- Tejas D Kulkarni, Karthik Narasimhan, Ardavan Saeedi, and Josh Tenenbaum. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29, pages 3675–3683. Curran Associates, Inc., 2016. URL <http://papers.nips.cc/paper/6233-hierarchical-deep-reinforcement-learning-integrating-temporal-abstraction-and-intrinsic-motivation.pdf>.
- Daniel Leightley, Moi Hoon Yap, Jessica Coulson, Yoann Barnouin, and Jamie S McPhee. Benchmarking human motion analysis using kinect one: An open source dataset. In *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1–7. IEEE, 2015.
- Siyuan Li, Jin Zhang, Jianhao Wang, Yang Yu, and Chongjie Zhang. Active hierarchical exploration with stable subgoal representation learning. In *International Conference on Learning Representations*, 2022. URL <https://openreview.net/forum?id=sNuFKTMktcY>.
- Xuefeng Liu, Takuma Yoneda, Rick Stevens, Matthew Walter, and Yuxin Chen. Blending imitation and reinforcement learning for robust policy improvement. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=eJ0dzPJq1F>.
- Manuel Lopes, Francisco Melo, Luis Montesano, and Jose Santos-Victor. *From Motor to Interaction Learning in Robots*, chapter Abstraction Levels for Robotic Imitation: Overview and Computational Approaches. Springer, 2009.
- M. Lungarella, Giorgio Metta, R. Pfeifer, and Giulio Sandini. Developmental robotics: a survey. *Connection Science*, 15(4):151–190, 2003.
- Ajay Mandlekar, Danfei Xu, Josiah Wong, Soroush Nasiriany, Chen Wang, Rohun Kulkarni, Li Fei-Fei,

- Silvio Savarese, Yuke Zhu, and Roberto Martín-Martín. What matters in learning from offline human demonstrations for robot manipulation. *arXiv preprint arXiv:2108.03298*, 2021.
- Alexandre Manoury, Sao Mai Nguyen, and Cédric Buche. Chime: An adaptive hierarchical representation for continuous intrinsically motivated exploration. In IEEE, editor, *International Robotics Conference*, pages 167–170, February 2019a. doi: 10.1109/IRC.2019.00032.
- Alexandre Manoury, Sao Mai Nguyen, and Cédric Buche. Hierarchical affordance discovery using intrinsic motivation. In *Proceedings of the 7th International Conference on Human-Agent Interaction, HAI '19*, pages 186–193, New York, NY, USA, 2019b. Association for Computing Machinery. ISBN 9781450369220. doi: 10.1145/3349537.3351898. URL <https://hal.science/hal-02283820>.
- Aleksa Marusic, Louis Annabi, Sao Mai Nguyen, and Adriana Tapus. Analyzing data efficiency and performance of machine learning algorithms for assessing low back pain physical rehabilitation exercises. In ACM/IEEE, editor, *European Conference on Mobile Robots*, 2023a. doi: 10.1109/ECMR59166.2023.10256318. URL <https://ieeexplore.ieee.org/document/10256318>.
- Aleksa Marusic, Sao Mai Nguyen, and Adriana Tapus. Evaluating kinect, openpose and blazepose for human body movement analysis on a low back pain physical rehabilitation dataset. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, HRI '23*, pages 587–591, New York, NY, USA, March 2023b. Association for Computing Machinery. ISBN 9781450399708. doi: 10.1145/3568294.3580153. URL <https://doi.org/10.1145/3568294.3580153>.
- Aleksa Marusic, Sao Mai Nguyen, and Adriana Tapus. Physioformer: A spatio-temporal transformer for physical rehabilitation assessment. In *ICSR*, 2024.
- Alina Dana Miron, Nouredin M Sadawi, Waidah Ismail, Hafez Hussain, and Crina Grosan. Intellirehabds (irds) - a dataset of physical rehabilitation movements. *Data*, 6:46, 2021.
- A. Mitriakov, P. Papadakis, Sao Mai Nguyen, and S. Garlatti. Staircase negotiation learning for articulated tracked robots with varying degrees of freedom. In *2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 394–400, Nov 2020. doi: 10.1109/SSRR50563.2020.9292594.
- Andrei Mitriakov, Panagiotis Papadakis, Sao Mai Nguyen, and Serge Garlatti. Learning-based modelling of physical interaction for assistive robots. *Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JFPDA)*, July 2019. URL <https://hal-imt-atlantique.archives-ouvertes.fr/hal-02341202>. Poster.
- Andrei Mitriakov, Panagiotis Papadakis, Sao Mai Nguyen, and Serge Garlatti. Staircase traversal via reinforcement learning for active reconfiguration of assistive robots. In *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8, July 2020. doi: 10.1109/FUZZ48607.2020.9177581.
- Clement Moulin-Frier, Sao Mai Nguyen, and Pierre-Yves Oudeyer. The role of intrinsic motivations in early vocal development: a computational study. In *Workshop on Intrinsic Motivations*, Rome, Italy, June 2013.
- Clément Moulin-Frier, Sao Mai Nguyen, and Pierre-Yves Oudeyer. Self-organization of early vocal development in infants and machines: The role of intrinsic motivation. *Frontiers in Psychology*, 4(1006), 2014. ISSN 1664-1078. doi: 10.3389/fpsyg.2013.01006. URL <https://inria.hal.science/hal-00927940>; http://www.frontiersin.org/cognitive_science/10.3389/fpsyg.2013.01006/abstract.

- Ofir Nachum, Shixiang Gu, Honglak Lee, and Sergey Levine. Near-optimal representation learning for hierarchical reinforcement learning. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL <https://openreview.net/forum?id=H1emus0qF7>.
- J. Nadel, A. Revel, P. Andry, and Ph. Gaussier. Toward communication: First imitations in infants, low-functioning children with autism and robots. *Interaction Studies*, 5(1):45–74, 2004. URL <http://publi-etis.ensea.fr/2004/NRAG04>.
- Jacqueline Nadel. *L'autisme*, chapter Imitation et austisme. Apr 2005.
- Nicolás Navarro-Guerrero, Sao Mai Nguyen, Erhan Oztop, and Junpei Zhong. Guest editorial special issue on continual unsupervised sensorimotor learning. *IEEE Transactions on Cognitive and Developmental Systems*, 13(2):234–238, 2021. doi: 10.1109/TCDS.2021.3082880.
- Chrystopher L Nehaniv and Kerstin Dautenhahn. *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*. Cambridge Univ. Press, Cambridge, March 2007. URL <file:///Users/mai/Sites/bibblywiki.html>.
- Andrew Y. Ng and Stuart Russell. Algorithms for inverse reinforcement learning. In *in Proc. 17th International Conf. on Machine Learning*, pages 663–670. Morgan Kaufmann, 2000.
- Sao Mai Nguyen. Role of default network in the understanding of subjectivity of desire. In *NIPS International Workshop for Scientific Study of Consciousness at Neuroscience*, volume NIPS International Workshop for Scientific Study of Consciousness at Neuroscience, 2009.
- Sao Mai Nguyen. Real-time face swapping based on head posture using particle filter towards the understanding of infant self-recognition. Master’s thesis, Osaka University, 2010.
- Sao Mai Nguyen. *A Curious Robot Learner for Interactive Goal-Babbling: Strategically Choosing What, How, When and from Whom to Learn*. PhD thesis, INRIA, 2013.
- Sao Mai Nguyen. Online learning a symbolic abstraction of actions in hierarchical rl with formal methods. In *ActSynt at ECAI*, 2024.
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Properties for efficient demonstrations to a socially guided intrinsically motivated learner. In *21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012a. doi: 10.1109/ROMAN.2012.6343819. URL <https://inria.hal.science/hal-00762758>.
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Socially guided intrinsically motivated learner. In *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on*, pages 1–2, nov. 2012b. doi: 10.1109/DevLrn.2012.6400809.
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Interactive learning gives the tempo to an intrinsically motivated robot learner. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, pages 645–652, 2012c. doi: 10.1109/HUMANOIDS.2012.6651588.
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Active choice of teachers, learning strategies and goals for a socially guided intrinsic motivation learner. *Paladyn Journal of Behavioural Robotics*, 3(3):136–146, 2012d. ISSN 2080-9778. doi: 10.2478/s13230-013-0110-z. URL <https://inria.hal.science/hal-00936932>.
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Whom will an intrinsically motivated robot learner choose to

- imitate from? In Joanna Szufnarowska, editor, *Proceedings of the Post-Graduate Conference on Robotics and Development of Cognition*, pages 32–35, September 2012e. doi: 10.2390/biecoll-robotdoc2012-12.
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Socially guided intrinsic motivation. In *Spring School on Developmental Robotics and Cognitive Bootstrapping*, 2012f.
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Strategic robot learner for interactive goal-babbling. In *Reinforcement Learning and Decision Making*, Princeton, New Jersey, U.S.A, 2013a.
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Strategic robot learner for active goal-babbling. In *Workshop in Active Learning in Robotics at Robotic Science and Systems*, Berlin, Germany, 2013b.
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Data collection of an interactive learner for strategy, outcome and policy exploration. In *Workshop in Hierarchical Learning at Robotic Science and Systems*, Berlin, Germany, 2013c.
- Sao Mai Nguyen and Pierre-Yves Oudeyer. Socially guided intrinsic motivation for robot learning of motor skills. *Autonomous Robots*, 36(3):273–294, 2014. doi: 10.1007/s10514-013-9339-y. URL <https://inria.hal.science/hal-00936938>.
- Sao Mai Nguyen and Philippe Tanguy. Cognitive architecture of a humanoid robot for coaching physical exercises in kinaesthetic rehabilitation. In *International Workshop on Cognitive Robotics*, 2016.
- Sao Mai Nguyen, Masaki Ogino, and Minoru Asada. Real-time face swapping as a tool for understanding infant self-recognition. In *Proceedings of the 10th International Conference on Epigenetic Robotics*, pages pp.171–172, Glumslöv, Sweden, 2010.
- Sao Mai Nguyen, Adrien Baranes, and Pierre-Yves Oudeyer. Bootstrapping intrinsically motivated learning with human demonstrations. In *IEEE International Conference on Development and Learning*, volume 2, pages 1–8, Frankfurt, Germany, Aug 2011a. doi: 10.1109/DEVLRN.2011.6037329.
- Sao Mai Nguyen, Adrien Baranes, and Pierre-Yves Oudeyer. Constraining the size growth of the task space with socially guided intrinsic motivation using demonstrations. In *IJCAI Workshop on Agents Learning Interactively from Human Teachers*, 2011b.
- Sao Mai Nguyen, Serena Ivaldi, Natalia Lyubova, Alain Droniou, Damien Gerardeaux-Viret, David Filliat, Vincent Padois, Olivier Sigaud, and Pierre-Yves Oudeyer. Learning to recognize objects through curiosity-driven manipulation with the icub humanoid robot. In *IEEE International Conference on Development and Learning - Epirob*, number 1-8, 2013. doi: 10.1109/DevLrn.2013.6652525. URL <https://hal.science/hal-00919674>.
- Sao Mai Nguyen, Philippe Tanguy, and Olivier Rémy-Néris. Computational Architecture of a Robot Coach for Physical Exercises in Kinesthetic Rehabilitation. In *RO-MAN 2016 : IEEE International Symposium on Human and Robot Interactive Communication*, pages 1138–1143, Aug 2016. doi: 10.1109/ROMAN.2016.7745251.
- Sao Mai Nguyen, Christophe Lohr, Philippe Tanguy, and Yiqiao Chen. Plug and play your robot into your smart home: Illustration of a new framework. *KI - Künstliche Intelligenz*, pages 1–7, 2017. ISSN 1610-1987. doi: 10.1007/s13218-017-0494-8. URL <http://dx.doi.org/10.1007/s13218-017-0494-8>.
- Sao Mai Nguyen, Nathalie Collot-Lavenne, Christophe Lohr, Sebastien Guillon, Patricio Tula, Alvaro Paez, Mouad Bouaida, Arthus Anin, and Saad El Qacemi. An implementation of an imitation game with asd

- children to learn nursery rhymes. In *HAI Workshop on Clinical Use of Technology for Individuals with Autism Spectrum Disorder*, 2019. URL <https://hal.science/hal-02499990>.
- Sao Mai Nguyen, Nicolas Duminy, Alexandre Manoury, Dominique Duhaut, and Cedric Buche. Robots learn increasingly complex tasks with intrinsic motivation and automatic curriculum learning. *KI - Künstliche Intelligenz*, 35:81–90, March 2021. doi: 10.1007/s13218-021-00708-8. URL <https://hal.archives-ouvertes.fr/hal-03157219>.
- Sao Mai Nguyen, Maxime Devanne, Olivier Remy-Neris, Mathieu Lempereur, and Andre Thepaut. A medical low-back pain physical rehabilitation dataset for human body movement analysis. In *International Joint Conference on Neural Networks*, 2024.
- M.N. Nicolescu and M.J. Mataric. Natural methods for robot task learning: Instructive demonstrations, generalization and practice. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 241–248. ACM, 2003.
- Pierre-Yves Oudeyer and Frederic Kaplan. What is intrinsic motivation? a typology of computational approaches. *Frontiers in Neurobotics*, 1(6), 2009.
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf.
- Erhan Oztop, Mitsuo Kawato, and Michael Arbib. Mirror neurons and imitation: A computationally guided review. *Neural networks*, 19(3):254–271, 2006.
- Rolf Pfeifer and C. Scheier. *Understanding Intelligence*. MIT Press, Cambridge, MA, 1999.
- Jean Piaget. *The origins of intelligence in children*, volume 8. International Universities Press New York, 1952. doi: 10.1037/11494-000.
- Giovanni Pioggia, ML Sica, Marcello Ferro, Roberta Iglizzi, Filippo Muratori, Arti Ahluwalia, and Danilo De Rossi. Human-robot interaction in autism: Face, an android-based social therapy. In *RO-MAN 2007-the 16th IEEE international symposium on robot and human interactive communication*, pages 605–612. IEEE, 2007.
- Diane Poulin-Dubois, Ivy Brooker, and Alexandra Polonia. Infants prefer to imitate a reliable person. *Infant Behavior and Development*, 34(2):303 – 309, 2011. ISSN 0163-6383. doi: <https://doi.org/10.1016/j.infbeh.2011.01.006>. URL <http://www.sciencedirect.com/science/article/pii/S0163638311000221>.
- Bob Price and Craig Boutilier. Implicit imitation in multiagent reinforcement learning. In *ICML*, pages 325–334, 1999.
- Harish Ravichandar, Athanasios S. Polydoros, Sonia Chernova, and Aude Billard. Recent advances in robot learning from demonstration. *Annual Review of Control, Robotics, and Autonomous Systems*, 3(1):297–330, 2020. doi: 10.1146/annurev-control-100819-063206. URL <https://doi.org/10.1146/annurev-control-100819-063206>.

- Mark Bishop Ring. *Continual learning in reinforcement environments*. PhD thesis, University of Texas at Austin Austin, Texas 78712, 1994.
- Giacomo Rizzolatti, Corrado Sinigaglia, and Frances Anderson. *Mirrors in the brain: How our minds share actions and emotions*. Oxford University Press, 2008.
- Stephane Ross, Geoffrey J. Gordon, and J. Andrew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. 11 2010. URL <https://arxiv.org/pdf/1011.0686.pdf>.
- N. Roy and A. McCallum. Towards optimal active learning through sampling estimation of error reduction. In *Proc. 18th Int. Conf. Mach. Learn.*, volume 1, pages 143–160, 2001.
- Richard M. Ryan and Edward L. Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1):54 – 67, 2000.
- Wakako Sanefuji, Hiroshi Yamashita, and Hidehiro Ohgami. Shared minds: Effects of a mother’s imitation of her child on the mother–child interaction. *Infant Mental Health Journal: Official Publication of The World Association for Infant Mental Health*, 30(2):145–157, 2009.
- Nikolay Savinov, Anton Raichuk, Damien Vincent, Raphael Marinier, Marc Pollefeys, Timothy Lillicrap, and Sylvain Gelly. Episodic curiosity through reachability. In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=SkeK3s0qKQ>.
- Stefan Schaal. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, 3(6): 233 – 242, 1999. ISSN 1364-6613. doi: [http://dx.doi.org/10.1016/S1364-6613\(99\)01327-3](http://dx.doi.org/10.1016/S1364-6613(99)01327-3). URL <http://www.sciencedirect.com/science/article/pii/S1364661399013273>.
- Stefan Schaal, Auke Jan Ijspeert, and Aude Billard. Computational approaches to motor learning by imitation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 358 (1431), 03 2003. URL <http://ukpmc.ac.uk/abstract/MED/12689379>.
- J. Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, Sep. 2010. ISSN 1943-0612. doi: 10.1109/TAMD.2010.2056368.
- A.P. Shon, D. Verma, and Rajesh PN Rao. Active imitation learning. In *American Association for Artificial Intelligence*, volume 22, page 756. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007.
- Ivelina Stoyanova, Nicolas Museux, Sao Mai Nguyen, and David Filliat. Open the chests: An environment for activity recognition and sequential decision problems using temporal logic. In *LIPICs, TIME*, volume 318, 2024.
- Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: an Introduction*. MIT Press, Cambridge, MA, USA, 1998. URL <http://webdocs.cs.ualberta.ca/~sutton/book/the-book.html>.
- Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: an Introduction*. MIT Press, Cambridge, MA, USA, 2018. URL <http://incompleteideas.net>.
- Richard S. Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112:181 – 211, 1999. ISSN 0004-3702. doi: [http://dx.doi.org/10.1016/S0004-3702\(99\)00052-1](http://dx.doi.org/10.1016/S0004-3702(99)00052-1). URL <http://www.sciencedirect.com/science/article/pii/S0004370299000521>.

- Tadahiro Taniguchi, Emre Ugur, Matej Hoffmann, Lorenzo Jamone, Takayuki Nagai, Benjamin Rosman, Toshihiko Matsuka, Naoto Iwahashi, Erhan Oztop, Justus Piater, and Florentin Wörgötter. Symbol emergence in cognitive developmental systems: a survey. *IEEE Transactions on Cognitive and Developmental Systems*, PP, 01 2018. doi: 10.1109/TCDS.2018.2867772.
- André Thépaut, Sao Mai Nguyen, and Christophe Lohr. Télé-rééducation fonctionnelle dans le cadre du projet VITAAL. *Actualités en médecine physique et de réadaptation : actualités en rééducation fonctionnelle et réadaptation*, (1):18–19, Mar 2017.
- André Thépaut, Sao Mai Nguyen, Maxime Devanne, and Olivier Rémy-Néris. Rééducation fonctionnelle assistée par robot humanoïde. In Serge Tisseron and Frédéric Tordo, editors, *Des robots au service des soignés, et des soignants*. Editions Eres, 2018. doi: 10.3917/eres.tisse.2018.02.0157. URL https://www.cairn.info/load_pdf.php?ID_ARTICLE=ERES_TISSE_2018_02_0157&download=1; <https://hal.archives-ouvertes.fr/hal-03604132>.
- Andrea L. Thomaz. *Socially Guided Machine Learning*. PhD thesis, MIT, 5 2006. URL <http://www.cc.gatech.edu/~athomaz/pubs.html>.
- Andrea L. Thomaz and Cynthia Breazeal. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence Journal*, 172:716–737, 2008. URL <http://www.cc.gatech.edu/~athomaz/pubs.html>.
- Andrea Lockerd Thomaz, Guy Hoffman, and Cynthia Breazeal. Real-time interactive reinforcement learning for robots. In *AAAI 2005 workshop on human comprehensible machine learning*, volume 3, page 1, 2005.
- Aleksandar Vakanski, Hyung-pil Jun, David Paul, and Russell Baker. A data set of human body movements for physical rehabilitation exercises. *Data*, 3(1):2, Jan 2018. ISSN 2306-5729. doi: 10.3390/data3010002. URL <https://www.kaggle.com/derekdb/toronto-robot-stroke-posture-dataset?>
- Linda Vallee, Sao Mai Nguyen, Christophe Lohr, Ioannis Kanellos, and Olivier Asseu. How an automated gesture imitation game can improve social interactions with teenagers with asd. In *IEEE ICRA workshop on Social Robotics for Neurodevelopmental Disorders*, 06 2020. doi: 10.13140/RG.2.2.10961.25444. URL <https://icra2020ndd.wordpress.com>.
- F. Varela, E. Thompson, and E. Rosch. *The embodied mind : cognitive science and human experience*. MIT Press, 1991.
- Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, and Koray Kavukcuoglu. Feudal networks for hierarchical reinforcement learning. *CoRR*, abs/1703.01161, 2017. URL <http://arxiv.org/abs/1703.01161>.
- Lev S Vygotsky. *Thought and Language*. MIT press, 1934.
- Lev S Vygotsky. *Mind in society : The development of higher mental processes*. Cambridge, MA: Harvard University Press, 1978. ISBN 9780674576292. URL <https://www.hup.harvard.edu/catalog.php?isbn=9780674576292>.
- LuKun Wang and RuYue Liu. Human activity recognition based on wearable sensor using hierarchical deep lstm networks. *Circuits, Systems, and Signal Processing*, 39(2):837–856, 2020. doi: 10.1007/s00034-019-01116-y. URL <https://doi.org/10.1007/s00034-019-01116-y>.
- R. White. Motivation reconsidered: The concept of competence. *Psychological review*, (66):297–333, 1959.

- Mehdi Zadem, Sergio Mover, Sao Mai Nguyen, and Sylvie Putot. Towards automata-based abstraction of goals in hierarchical reinforcement learning. In *Intrinsically Motivated Open-ended Learning IMOL 2022*, Tübingen, Germany, April 2022. URL <https://hal.science/hal-03600799>.
- Mehdi Zadem, Sergio Mover, and Sao Mai Nguyen. Goal space abstraction in hierarchical reinforcement learning via set-based reachability analysis. In *2023 IEEE International Conference on Development and Learning (ICDL)*, pages 423–428, Nov 2023a. doi: 10.1109/ICDL55364.2023.10364473. URL <https://arxiv.org/abs/2309.07675>.
- Mehdi Zadem, Sergio Mover, and Sao Mai Nguyen. Goal space abstraction in hierarchical reinforcement learning via reachability analysis. In *Intrinsically Motivated Open-ended Learning*, Tübingen, Germany, September 2023b. URL <https://hal.science/hal-03600799>.
- Mehdi Zadem, Sergio Mover, and Sao Mai Nguyen. Emergence of a symbolic goal representation with an intelligent tutoring system based on intrinsic motivation. In *Intrinsically Motivated and Open-Ended Learning Workshop @NeurIPS2023*, 2023c. URL <https://openreview.net/forum?id=UawMOafW3f>.
- Mehdi Zadem, Sergio Mover, and Sao Mai Nguyen. Reconciling spatial and temporal abstractions for goal representation. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=odY3PkI5VB>.
- Martijn JA Zeestraten, Ioannis Havoutis, Joao Silvério, Sylvain Calinon, and Darwin G Caldwell. An approach for imitation learning on riemannian manifolds. *IEEE Robotics and Automation Letters*, 2017.
- Ruohan Zhang, Faraz Torabi, Lin Guan, Dana H. Ballard, and Peter Stone. Leveraging human guidance for deep reinforcement learning tasks. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 6339–6346. International Joint Conferences on Artificial Intelligence Organization, 7 2019. doi: 10.24963/ijcai.2019/884.
- Tianren Zhang, Shangqi Guo, Tian Tan, Xiaolin Hu, and Feng Chen. Generating adjacency-constrained subgoals in hierarchical reinforcement learning. 06 2020. URL <https://arxiv.org/pdf/2006.11485.pdf>.