Interactive Robot Learning: An Overview

Mohamed Chetouani

Interactive Robot Learning: An Overview*

Mohamed CHETOUANI[1 [0000−0002−2920−4539]]

Institute for Intelligent Systems and Robotics, CNRS, UMR7222,
Sorbonne University, Paris, France
mohamed.chetouani@sorbonne-universite.fr

Abstract. How do we teach robots to perform tasks? Here, we focus on main methods and models enabling humans to teach embodied social agents such as social robots, using natural interaction. Humans guide the learning process of such agents by providing various teaching signals, which could take the form of feedback, demonstrations and instructions. This overview describes how human teaching strategies are incorporated within machine learning models. We detail the approaches by providing definitions, technical descriptions, examples and discussions on limitations. We also address natural human biases during teaching. We then present applications such as interactive task learning, robot behavior learning and socially assistive robotics. Finally, we discuss research opportunities and challenges of interactive robot learning.

Keywords: Robot Learning · Interactive Machine Learning · Reinforcement Learning · Learning from Feedback · Learning from Demonstrations · Learning from Instructions · Human Teaching Strategies

1 Introduction

Robot learning deals with algorithms, methods and methodologies allowing a robot to master a new task such as navigation, manipulation and classification of objects. At the intersection of machine learning and robotics, robot learning addresses the challenge of task learning, which is defined by a goal (e.g. grasping an object). The aim is to identify a sequence of actions to achieve this goal. Multi-task learning, transfer learning or life-long learning are also considered for this purpose.

Several trends of robot learning take inspiration from human learning by studying developmental mechanisms [61]. In particular, several of such trends focus *social learning* since human learning often occurs in a social context. The computational approaches of social learning are formulated as an interaction between a tutor/teacher/demonstrator and an artificial learner/student/observer. The aim of the teacher is to influence the behavior of the learning agent by providing various cues such as feedback, demonstrations or instructions. Interactive

* This work has received funding from European Union's Horizon 2020 ICT-48 research and innovation actions under grant agreement No 952026 (HumanE-AI-Net) and from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765955 (ANIMATAS)

task learning [56] aims at translating such interactions into efficient and robust machine learning frameworks. Interactive task learning is usually considered to be an alternative to autonomous learning. The latter requires an evaluation function that defines the objective of the task. The robot autonomously learns the task by continuously evaluating its actions using this function. Interactive learning assumes that a human will be able to assist the robot in the evaluation by providing feedback, guidance and/or showing optimal actions. In this chapter, we describe the fundamental concepts of interactive robot learning with the aim of allowing students, engineers and researchers to get familiar with main definitions, principles, methodologies, applications and challenges.

The chapter is structured as follows. Section 2 presents learning objectives, notations, abbreviations and relevant readings. Section 3 provides a background on reinforcement learning and robot learning. Section 4 describes the types of human interventions in both traditional supervised machine learning and interactive machine learning. Section 5 discusses human teaching strategies in interactive robot learning. In Sections 6 to 8, we provide definitions, describe learning methods and examples as well as limitations of each strategy: feedback (Section 6), demonstrations (Section 7) and instructions (Section 8). Section 9 gives deeper insights about modeling approaches to take into account natural human biases during teaching. Section 10 presents several applications of interactive robot learning: interactive task learning, learning robot behaviors from human demonstrations or instructions, and socially assistive robotics. Finally, in Section 11 sums up main observations and describes several opportunities and challenges of interactive robot learning.

2 Tutorial scope and resources

2.1 Learning objectives

- Awareness of the human interventions in standard machine learning and interactive machine learning.
- Understand human teaching strategies
- Gain knowledge about learning from feedback, demonstrations and instructions.
- Explore ongoing works on how human teaching biases could be modeled.
- Discover applications of interactive robot learning.

2.2 Notations

- s, a: state and action, $s \in S$ and $a \in A$.
- a^*: optimal action.
- $H(s, a)$: Human Feedback at state s for robot action a.
- $D = \{(s_t, a_t^*), (s_{t+1}, a_{t+1}^*)....\}$: Human Demonstrations, a state-action sequence.
- $I(s)$: Human Instruction at state s, $Pr_t(a|i)$.

- $Pr(s'|s,a)$: the probability of going from state s to state s' after executing action a.
- $< S, A, T, R, \gamma >$: State & Action spaces, State-Transition probability function ($Pr(s'|s,a)$), Reward function, and the discount factor ($[0,1]$).
- $r(s,a)$: reward at state s for action a.
- π: agent/robot policy.
- $V^{\pi}(s)$: state-value function.
- $Q^{\pi}(s,a)$: action-value function.

2.3 Acronyms

- AI: Artificial Intelligence
- HRI: Human Robot Interaction
- IML: Interactive Machine Learning
- IRL: Inverse Reinforcement Learning
- ITL: Interactive Task Learning
- LfD: Learning from Demonstrations
- MDP: Markov Decision Process
- ML: Machine Learning
- RL: Reinforcement Learning

2.4 Selected relevant readings

- Robot learning from human teachers, Chernova & Thomaz (2014) [21]
- Interactive task learning, Laird et al. (2017) [56]
- Recent advances in leveraging human guidance for sequential decision-making tasks, Zhang et al. (2021) [100]
- Reinforcement Learning With Human Advice: A Survey, Najar & Chetouani (2020) [67]
- Survey of Robot Learning from Demonstration, Argall et al. (2009) [6]
- Recent advances in robot learning from demonstration, Ravichandar et al. (2020) [80]
- A survey on interactive reinforcement learning: Design principles and open challenges, Cruz et al. (2020) [24]
- On studying human teaching behavior with robots: A review, Vollmer & Schillingmann (2018) [96]
- Cognitive science as a source of forward and inverse models of human decisions for robotics and control, Ho & Griffiths (2022) [40]
- Towards teachable autonomous agents, Sigaud et al. (2022) [85].

3 Background

3.1 Fundamentals of Reinforcement Learning

Reinforcement Learning (RL) is a branch of machine learning concerned with how an *autonomous agent* learns sequential decisions in an uncertain environment by maximizing a cumulative reward [87]. This category of problems is

modeled as a Markov Decision Process (MDP), which is defined by a tuple $< S, A, T, R, \gamma >$ with S the state space, A the action space, $T : S \times A \rightarrow Pr(s'|s, a)$ state-transition probability function, where $Pr(s'|s, a)$ is the probability of going from state s to state s' after executing action a and $R : S \times A \rightarrow R$ the reward function, which represents the reward $r(s, a)$ that the agent gets for performing action a in state s. The reward function R defines the objective of the task. A discount factor γ ($[0, 1]$) controls of the trade-off between immediate reward and delayed reward. In the reinforcement learning framework, the dynamics of the autonomous agent is captured by the transition function T: at time t, the agent performs an action a_t from state s_t, it receives a reward r_t and transitions to state s_{t+1}.

Example: Figure 1 illustrates the concept of RL in robotics in which the state space describes the environment: position of boxes; action space the possible robot actions: arm motion. After each action, the robot receives a binary reward.

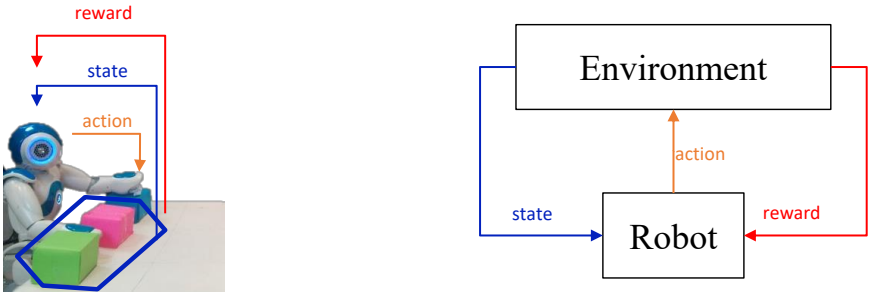


Fig. 1. RL learning framework used to model the environment (state space S), robot actions (action space A) and the task that should be achieved formulated by an MDP (Markov Decision Process). The robot receives a reward R after each action (adapted from [66])

The aim of RL is to find a policy $\pi : S \rightarrow A$ which maps from the state s to a distribution on the action A, this is the decision function. An agent endowed with a policy π will behave in a certain manner to achieve a task. The policy could be either deterministic or stochastic [87]. The optimal policy π^* maximizes agent's long-term rewards. Learning is performed with a trial-and-error strategy. The long-term view is defined by the *state-value function* $V^\pi(s)$, which specifies the expected gain for an agent to be in a particular state s with a policy π. Similarly, the *action-value function* $Q^\pi(s, a)$ defines the expected gain for policy π starting from state s taking action a.

Several algorithms of RL have been proposed (see [87] for an overview). A fundamental problem faced by some of the RL algorithms, is the *dilemma between the exploration and exploitation*. Exploration allows the autonomous agent

to gather new information that may improve future reward and consequently the policy (eg., $\epsilon-$greedy approach). The exploitation refers to making the best decision given the current model. Several modern RL algorithms have been proposed to tackle this critical dilemma of exploration and exploitation.

3.2 Robot Learning

RL and robotics. In [52], the authors discuss the challenges, problems and opportunities of reinforcement learning in the context of robotics. In particular, the physical embodied nature of robots, the nature of tasks and the limited perception of the environment often result in problems that should be represented with high-dimensional continuous states and actions. In addition, these states are not always completely observable. Exploration of actions is costly, difficult to reproduce and sometimes unsafe. The specification of a "good" reward function is not always straightforward and requires a significant amount of domain knowledge.

Human interventions in the robot learning process. In the context of human-robot interaction, reinforcement learning could be employed for several purposes:

- *Interactive Task Learning*. The aim is to learn to perform tasks with a human involved in the learning process by evaluating and guiding the learning process [92].
- *Learning Communicative Behaviors*. The aim is to learn to generate multimodal robot behaviors such as legible motion (transparency) [16] or select appropriate behaviors during interaction in order to adapt to the human partner [64].

The focus of this chapter is Interactive Task Learning. We consider tasks that are performed by embodied agents in the physical world or in a simulation of it. Such agents could use their sensors and effectors to both perform the task and communicate with the human: manipulating an object, pointing to indicate an object. This chapter discusses the role of interaction in the agent/robot learning process.

Communication in the robot learning process. Human interventions could enrich task learning in the form of teaching signals to guide the learning process such as gaze at objects or spoken language for feedback. Demonstration of a task is an interesting paradigm since it gathers action and communication about it at the same time. Similarly, robot communication using transparency and explainability mechanisms of robots could include both task and communicative goals directed actions. For example, generation of legible motions [28, 98] facilitates human interpretation of the robot task.

The objective of a task could be to communicate. For example to address the symbol grounding problem [88], both physical and social symbol grounding,

there is a need for agent to connect sensory and symbolic representations in order to be able to process them, reason about them, generate new concepts and communicate with humans in particular.

The distinction between task achievement and communication is not always easy and sometimes not relevant at all. However, being able to qualify the goal of learning is important for interactive and autonomous agents. For this purpose, interactive learning is conceptualized as a mutual exchange process using two main communication channels, the social channel and the task channel with valuable interpretations of several learning strategies from: observation, demonstration, instruction or feedback [85].

4 Interactive Machine Learning vs. Machine Learning

The purpose of this section is to introduce the main concepts of interactive machine learning. Reviewing all the mechanisms of interactive machine learning is beyond the scope of this chapter. However, understanding the impact of human interventions during the learning process is important for the content of this chapter. For this purpose and for a sake of clarity, we only discuss the human interventions on both the traditional and interactive machine learning processes.

4.1 Human interventions in the supervised machine learning process

Human interventions are already present in the traditional machine learning process (Figure 2). The obvious case is *human machine interaction*: the objective is to support end-users interaction. However, in most of the machine learning approaches, humans are present at different stages. They could provide data, annotations, design the algorithms, evaluate the model, design the interaction and interact with it (Figure 2). These interventions are made by different profiles: users, domain expert, machine learning expert, human-machine expert and end-users and at different steps corresponding to different time scales in the process: data collection, annotation, data analysis, algorithm design, interaction design, model training and evaluation and final end-user model interaction. The impacts of such interventions are of different natures.

Example: Human emotion recognition systems require data and annotations to build robust and efficient systems. Data collection is a key phase and usual approaches rely either on acted or real-life scenarios. In addition, various methodologies have been proposed to annotate emotions with different representations (discrete vs. dimensional). Self-assessment (e.g, asking the data provider for annotation) or assessment from an external observer [1]. All these methodological and experimental choices impact the design, performance and robustness of the emotion recognition system.

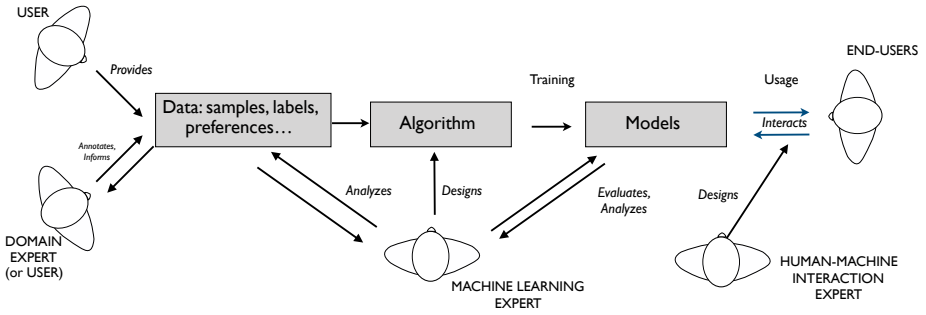


Fig. 2. Humans in the Machine Learning Process. Data Providers, Domain, Machine Learning and Human-Machine Interaction Experts as well as End-Users play a key role in the machine learning process.

4.2 Human interventions in the interactive machine learning process

Interaction with humans is at the core of the Interactive Machine Learning (IML) process [4]: from the design to the usage (Figure 3). There is a growing interest for IML for several applications: web recommendation, e-mail processing, chatbots or rehabilitation. *Interactive Machine Learning* is at the intersection of *Machine Learning* and *Human-Computer/Machine Interaction*.

Human interventions are of different nature. While in traditional ML (Figure 2), it is usual to collect and process very large datasets with multiple users, interactive ML relies on the interaction with the end-user to collect data. The training, evaluation and usage phases are intrinsically linked, which requires specific approaches for the design of both algorithms and interactions.

Interestingly, interactive ML opens new ways of lifelong learning, adaptation and personalization of models, which could improve usage and trust. However, having humans at the center of the process also raises several ethical questions [84, 56] that should be addressed including the definition of requirements to develop human-centric AI that minimizes negative unintended consequences on individuals and on the society as a whole [57, 30].

Interactive Robot Learning approaches are grounded in Interactive ML [21]. The embodied nature of robots make the researchers to also take inspiration from human social learning [92] or even modeling its key mechanisms of child development as done in developmental robotics [61, 85]. In the following sections, we will describe human's teaching strategies and how they are represented and modeled to fit interactive machine learning frameworks.

5 Overview of Human Strategies

In this section, we give an overview of the main strategies employed by humans to teach robots. We consider the situation in which a human provides teaching signals to a learning agent (Figure 4). Teaching signals could take different forms

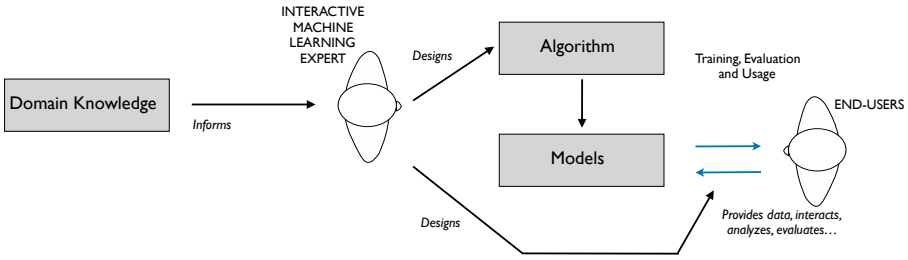


Fig. 3. Humans in the Interactive Machine Learning Process. End-users are both data providers and model evaluators. Designing the interaction is as important as designing the models. The Expert becomes an *Interactive Machine Learning Expert*.

(e.g., demonstration, instruction) and are at the core of the human teaching strategy. Most of these strategies assume that humans are rational intentional agents, optimal with respect to their decisions, actions and behaviors. The concept of intention is used to characterize both the human's actions and mental states [13]. In case of interactive robot learning, this means that humans provide teaching signals with an intention, which is translated into a sequence of actions aiming at influencing the robot learning.

In Table 1, we describe the main teaching signals considered in human-robot interaction: *feedback, demonstration and instruction*. They are respectively used to communicate specific intentions: evaluating/correcting, showing and telling. Social and task channels are employed to communicate intentions (Figure 4). Instructions, gaze or pointing are considered as being conveyed by the social channel. Manipulation of objects to demonstrate a task exploit the task channel. These channels could be combined and exploited by both the human and the robot during the learning process.

The learning agent needs to infer human's intention from the teaching signals. However, intentions are not explicit. Ambiguities could occur during communication. Humans could also intentionally deviate from optimality and apparently behave as non-rational agents (Section 9). The most common strategies are certainly to learn new tasks by providing feedback, demonstration or instructions (Table 1). In the following sections, we propose definitions, mathematical formulations and discuss interpretations and teaching/learning costs of such strategies.

6 Feedback

6.1 Representation

Human feedback $H(s, a)$ is considered as an observation about the reward $r(s, a)$. Binary and Real-valued quantities have been considered in interactive reinforcement learning.

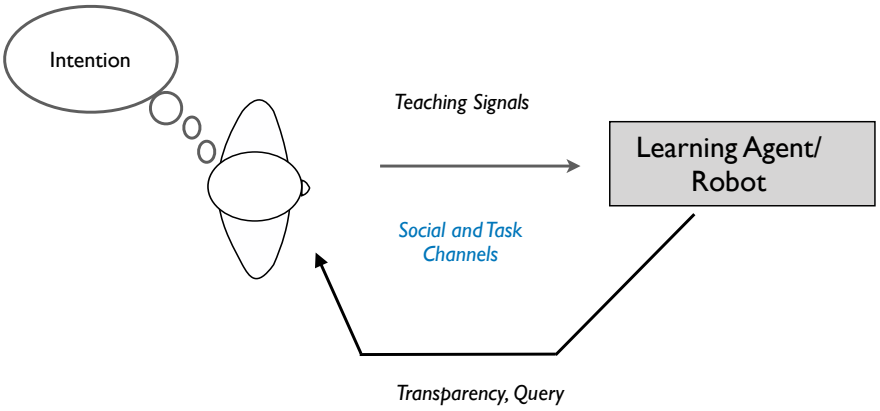


Fig. 4. Interactive Robot Learning Paradigm: The Human Teacher provides Teaching signals using Social and/or Task Channels. Teaching signals are performed with an intention. The Learning Agent/Robot infers the intention from the teaching signal and exploits them to learn a task. Learning agents could improve learning by increasing transparency and/or asking for additional information (e.g., labels, features).

Table 1. Description of main Human Teaching Strategies. Robot action is performed at time-step t. A teaching signal is the physical support of the strategy using social and/or task channels.

Teaching signals		**Feedback**	**Demonstration**	**Instruction**
Nature	**Notation**	$H(s,a)$	$D = \{(s_t, a_t^*), (s_{t+1}, a_{t+1}^*)....\}$	$I_\pi(s) = a_t^*$
	Value	Binary / Scalar	State-Action pairs	Probability of an action
Time-step	**t-1**		✓	✓
	t		✓	
	t+1	✓		
Human	**Intention**	Evaluating / Correcting	Showing	Telling
	Teaching cost	Low	High	Medium
Robot	**Interpretation**	State-Action evaluation Reward-/Value-like	Optimal actions Policy-like	Optimal action Policy-like
	Learning cost	High	Low	High

6.2 Definition

Human feedback H is produced at $t+1$, after the evaluation of the robot's action (Table 1). Human feedback is provided with the intention of evaluating the robot's action (critique). Human Feedback is considered to communicate about the performance of a robot's action. Human feedback is usually termed as *evaluative feedback*: the human observes the states s_t and s_{t+1} and the last robot action a and then gives a reward signal. The value of the evaluative feedback depends on the last action performed by the robot.

Teaching with evaluative feedback is a low cost strategy (Table 1). The actions are performed by the robot using exploration algorithms. The human only delivers feedback on the observed actions. However, intervention at each time-step is not realistic and imposes a significant burden on the human teacher. This calls for new algorithms able to efficiently integrate human feedback during learning.

6.3 Learning from Evaluative Feedback

An intuitive way is to remove the reward function R from the reinforcement approach for task specification (Section 3). This results in an $MDP \backslash R$, an MDP without a reward function. The reward is then replaced by the human feedback. Understanding human's feedback strategy has been the focus of several works [90, 47, 42, 68]. They all highlight the importance of understanding human's intentions and the design of adequate algorithms for the exploitation of teaching signals. To efficiently learn from human feedback, there is a need to go beyond just considering the human feedback as a reward and eventually combining it with the reward of the environment.

The usual approach is to design models of human feedback that are able to capture specific properties of human strategies. This leads to various integration strategies of Human feedback H into RL systems. There is no clear agreement on the many ways of achieving this integration but the main approaches rely on shaping such as:

- Human feedback H as a reward r: *reward shaping*
- Human feedback H as a value V or Q: *value shaping*
- Human feedback H as a policy π: *policy shaping*

We detail the approaches that result from three different interpretations of human feedback. They all consider shaping as a method to influence the agent behavior towards a desired behavior [49, 69, 68].

6.4 Reward shaping

Human feedback is interpreted as a reward (Table 1). When the agent interacts with its environment, modeled as a Markov Decision Process (MDP), it receives a reward $r(s,a)$ and an additional reward shaping reward $H(s,a)$:

$$r'(s,a) = r(s,a) + \beta * \hat{H}(s,a). \tag{1}$$

where β, a decaying weight factor, controls the contribution of human feedback $H(s,a)$ over the environment reward $r(s,a)$.

The reward $r(s,a)$ is first augmented by the human feedback $\hat{H}(s,a)$. Then, the augmented reward $r'(s,a)$ is used to shape the agent. This is usually considered as an indirect shaping approach [68, 77].

6.5 Value shaping

Human feedback is interpreted as a human value function (Table 1). The human evaluates the action by providing a rating of the current agent's action with respect to some forecast of future behavior [100, 68, 48, 23, 43]. This rating is employed to augment the action-value function $Q(s,a)$. Shaping method have been considered in the literature [68]:

$$Q'(s,a) = Q(s,a) + \beta * \hat{H}(s,a), \tag{2}$$

where β is a decaying weight factor (see also equation 1).

Other approaches consider an estimation of the human value function as done in TAMER [49] (Section 6.7).

6.6 Policy shaping

Human feedback is still interpreted as a value (Table 1) but employed to directly influence the agent's policy [68, 77, 100]. Two main methods have been considered so far [50]:

- Action biasing: The shaping is only performed during the decision-making step. The value function is not directly perturbed by the human feedback augmentation:
$$a^* = \arg\max[Q(s,a) + \beta * \hat{H}(s,a)], \tag{3}$$
- Control sharing: This method arbitrates between the MDP policy and human value function. The human policy derived from feedback is used for action selection given the probability β:
$$Pr\left[a = \arg\max\left(\hat{H}(s,a)\right)\right] = \min(\beta, 1) \tag{4}$$

Other approaches have been proposed in the literature such as combination of policies using multiplication of probability distributions [36, 69, 78]. COACH (Convergent Actor-Critic by Humans) algorithm [62] is motivated by the observation that human policy is influenced by learner's current policy. The authors argue that *the advantage function* [87] is a good model of human feedback. They use actor-critic algorithms to compute an unbiased estimate of the advantage function.

6.7 Example: The TAMER architecture

The TAMER architecture (Training an Agent Manually via Evaluative Reinforcement) [49] assumes that the human has an internal function H that maps observed agent action in a feedback (negative, neutral or positive). The human has in mind a desired policy π_H and wants to communicate it to the agent through feedback. In other words, TAMER estimates human's intention (table 1) from the observation of feedback. The internal function H is called the "Human Reinforcement Function". TAMER approximates the human internal function by a regression model $\hat{H}^*$ trough minimizing a standard squared error loss between $H(s_t, a_t)$ and $\hat{H}(s_t, a_t)$. TAMER is formulated as an MPD without a reward ($MDP \backslash R$). The agent uses this function to perform action selection:

$$\pi(s) = \arg\max_a \hat{H}^*(s, a) \tag{5}$$

TAMER interprets Human feedback as a value. In [51], the authors address delays in human evaluation (human's feedback) through credit assignment, which includes creation of labels from delayed reward signals. TAMER has been successfully combined with Reinforcement learning (TAMER+RL) [50] and recently Deep Learning (Deep TAMER) in order to deal with high-dimensional state spaces [99].

6.8 Limitations

Understanding human feedback strategies is essential in the design and development of learning from evaluative feedback algorithms. Previous works have identified numerous issues such as:

- *Credit assignment problem*: humans provide feedback with a delay by considering actions happened in the past [24].
- *Policy dependent*: humans' feedback strategy is influenced by learner's current policy [62].
- *Positively biased feedback*: positive and negative feedback are not employed in the same way by humans [89, 43]. Whether providing both positive and negative rewards is necessary is an open question [77].
- *Reward hacking*: this describes situations in which non anticipated actions are introduced by the agent in order to obtain a positive human feedback. Several works show that is difficult to anticipate failure behaviors aroused from reward functions [24, 77].
- *Autonomous/self agent exploration vs social interaction*: relying only on human feedback is not efficient and imposes a significant burden on the human teacher. Current trends combine agent exploration and human feedback in order to improve robustness to sparse and/or erroneous teaching signals

Current and future works of the domain are addressing such issues by both conducting human studies as well as developing new machine learning algorithms and human-machine interaction designs.

7 Demonstrations

7.1 Representation:

Human demonstrations D are usually represented by a state-action sequence (Table 1): $\{(s_t, a_t^*).\}$. Where a_t^* is the optimal human action at time-step t given a state s_t.

7.2 Definition

A demonstration D is produced by the human demonstrator with the intention of *showing a state-action sequence to the robot* (Table 1). The paradigm assumes that the human expert shows the optimal actions a_t^* for each state. The state-action sequence is then reproduced by the robot.

The identification of a mapping between the human teacher and robot learner actions, which allows the transfer of information from one to the other, is called the *correspondence problem* [6]. Contrary to evaluative feedback, demonstrations could be provided before (time-step $t-1$) or simultaneously (time-step t) to robot actions (Table 1).

7.3 Methods

Teaching with a demonstration strategy imposes a significant burden on the human teacher. The assumption is that the human teacher is being able to perform the task in order to communicate optimal actions a_t^* through the task channel (see Figure 4). The set of demonstrations are interpreted as a human policy of the task (π_h) (Table 1). For such reasons, *Interaction Design* plays an important role in learning from demonstrations (see Figure 3). Three general methods are usually considered in the literature [80]: *kinesthetic, teleoperation and observation*.

Kinesthetic demonstration is an example of simultaneous teacher-learner interactions through demonstrations, in which the teacher directly manipulates the robot. This approach eliminates the correspondence problem and simplifies the machine learning process. Several industrial robots are now proposing this approach to facilitate task learning (Section 10). Kinesthetic demonstration facilitates the mapping and it facilitates the production of human demonstrations. However, the quality of demonstrations is known to be low as it depends on the dexterity and smoothness of the human demonstrator during the manipulation of the robot [80]. With teleoperation, the human demonstrator can provide demonstrations of robots with high degrees-of-freedom (HOF) as well as facilitating mapping. However, teleoperation requires the development of specific interfaces (including virtual/augmented reality) [80]. Observation of human demonstrations offers a natural interaction mode to the users. However, learning from demonstrations by observation of human actions using a camera or a motion capture is not sufficient to compute the mapping. Several challenges of machine perception are faced including motion tracking, occlusion or high degrees-of-freedom of human motion.

7.4 Learning from demonstrations

Demonstration facilitates engagement of non-expert users in robot programming. They can teach new tasks by showing examples rather than programming them (see Section 10). Interest in learning from demonstrations is shown by the number of publications in recent years (see [80] for an overview), resulting also in various terminologies: *imitation learning, programming by demonstration, behavioral cloning and Learning from Demonstrations (LfD).*

The challenge of learning from demonstrations in robotics has been addressed by methods and models of supervised, unsupervised and reinforcement learning. In [6], the authors argue that LfD could be seen as a subset of Supervised Learning. The robot is presented with human demonstrations (labeled training data) and learns an approximation to the function which produced the data. In the following, we discuss two different perspectives: (i) behavioral cloning (supervised learning) and sequential decision-making (reinforcement learning). Inverse Reinforcement Learning (IRL) is presented as an example in section 7.7.

7.5 Behavioral cloning

Behavioral cloning employs supervised learning methods to determine a robot policy that imitates the human expert policy. This is performed by minimizing the difference between the learned policy and expert demonstrations (state-action pairs) with respect to some metric (see algorithm 1).

Algorithm 1 Behavioral Cloning

1: **procedure** Behavioral Cloning
2: Collect a set of expert demonstrations $D = \{(s_t, a_t^*), (s_{t+1}, a_{t+1}^*)...\}$
3: Select an agent policy representation π_θ
4: Select a loss function L
5: Optimize L using supervised learning: $L\left(a^*, \pi_\theta(s, a)\right)$
6: Return π_θ

A series of recommendations are made in [75] regarding nature of demonstrations (trajectory, action-state space), choice of Loss Functions (quadratic, l_1, log, hinge and Kullback-Leibler divergence) and supervised learning methods (regression, model-free and model-based).

7.6 Imitation learning as a sequential decision-making problem

In [81], the authors argue that imitation learning could not be addressed as a standard supervised learning problem, where it is assumed the training and testing data are independent and identically distributed (i.i.d). They show that imitation learning is a *sequential decision-making problem*. For this reason, reinforcement learning techniques and interactive supervision techniques have been considered to address imitation learning.

7.7 Example: Inverse Reinforcement Learning

Inverse Reinforcement Learning (IRL) [70] is a popular approach in imitation learning. IRL aims to recover an unknown reward function given a set of demonstrations and then to find the optimal policy (conditioned by the learned reward function) using reinforcement learning (see Figure 5).

IRL methods consider the set of expert's demonstrations as *observations of the optimal policy* π^*. Interpretation of demonstrations follows a *policy-like* vision (Table 1). From a set of demonstrations, the goal of IRL is to estimate the unknown reward function parameters of a policy $\hat{\pi}^*$ that imitates expert's policy π^*. The reward function is then analyzed to understand and/or to explain the expert's policy. Standard IRL algorithms consider the reward function as a linear combination (ψ^T) of features $f(s,a)$ of the environment:

$$r_\psi(s,a) = \psi^T f(s,a) \tag{6}$$

This approach assumes that the expert is acting accordingly in the environment. Modern approaches consider non-linear combinations using neural network based estimators of $r_\psi(s,a)$. IRL is an ill-posed problem since there are infinitely many reward functions consistent with the human expert's demonstrations.

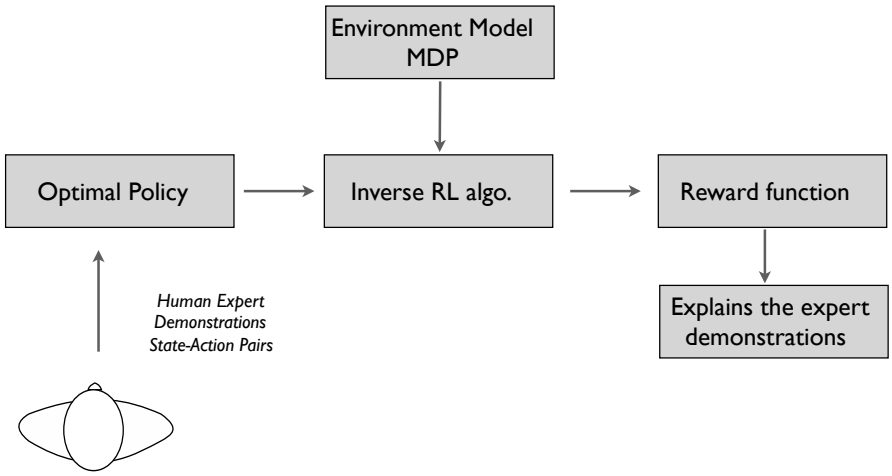


Fig. 5. Inverse Reinforcement Learning (IRL): Human expert provides demonstrations (state-action pairs). Demonstrations are interpreted as a policy. The goal of IRL is to compute the underlying reward function from the demonstrations.

7.8 Limitations

LfD methods assume that the demonstrations, state-action pairs (s_t, a_t^*), are the only teaching signals available to the learner (Table 1). After human demon-

strations recording, the learner attempts to learn the task using these teaching signals. This approach has been formulated as an MDP without a reward function ($MDP\backslash R$). As previously mentioned, demonstrations could be interpreted as a *human policy* π_h (Table 1). Similarly to learning from evaluative feedback, a current challenge of LfD is to *include other forms of rewards* in the learning, in particular those computed from environment and/or intrinsic motivations [85].

As discussed in [80], the *choice of learning methods* is compelling: (i) the optimal behavior cannot be scripted, (ii) the optimal behavior is not always easily defined in the form of a reward function, and (iii) the optimal behavior is only available through human teacher demonstrations. This leads to several challenges at the intersection of machine learning (e.g. learning methods), robotics (e.g. control of physical robots) and human-robot interaction (e.g., human factors, interaction design). In [75], a series of questions summarizes the current challenges of imitation learning:

- *Why and when should imitation learning be used?*
- *Who should demonstrate?*
- *How should we record data of the expert demonstrations?*
- *What should we imitate?*
- *How should we represent the policy?*
- *How should we learn the policy?*

Answering these questions is required for the design of imitation learning based systems (more details are available in [75]).

8 Instructions

8.1 Representation:

Human instructions are usually represented as a probability distribution over actions: $Pr_t(a|i)$ and $a \in A$.

8.2 Definition:

An instruction is produced by the human with the intention of communicating the action to be performed in a given task state (Table 1). This could be formulated as *telling*, which is a language based perspective of the interaction. Examples of instructions could be *turn left*, *pick up the object* or *go forward*.

Telling offers other opportunities to the teacher compared to *evaluating* (feedback) and *showing* (demonstration). A recent work of [86] studies how humans teach concepts using either demonstrations or language. The results suggest that language communicates more complex concepts by directly transmitting abstract rules (e.g., shapes and colors), while demonstrations transmit positive examples (e.g. manipulation of objects) and feedback evaluates actions. Inferring rules from demonstrations or feedback is supposed to be more complex.

Instructions are of different nature and are produced with different intentions grouped in the notion of *advice* [68]: guidance, critique, action advice. In [54],

two different methods are compared: critique and action advice. In the critique based teaching method, the learning agent is trained using positive and negative verbal critiques, such as *good job*, and *don't do that*. As described in section 6, such verbal critiques are binary feedback. In the action advice based teaching method, the learner is trained using action advice such as *move right* and *go left*. The authors show that action advice creates a better user experience compared to an agent that learns from binary critique in terms of frustration, perceived performance, transparency, immediacy, and perceived intelligence.

8.3 Learning from instructions

Learning from instructions is formulated as *mapping instructions (often in natural language) to a sequence of executable actions* [12, 60]. Mutual understanding (human/learning agent) of the meaning of instructions is usually assumed, which obviously facilitates the *instruction-to-action mapping*. The usual approach is to pre-define the mapping, which raises with several issues such as engineering and calibration phases [37, 68], adaptation and flexibility [37, 68], intermediate semantic representation [88, 63, 2], and reward inference from language [58].

We could distinguish between methods that focus on simple commands and the ones addressing sequence of instructions. The latter are employed to compose complex robot/agent behaviors with the interventions of a human teacher [82, 27, 76]. In [27], the authors describe a Spoken Language Programming (SLP) approach that allows the user to guide the robot through an arbitrary, task relevant, motor sequence via spoken commands, and to store this sequence as a re-usable macro. Most of these approaches model the task through a graph [88].

Learning from instructions has been combined with learning from demonstrations [72, 82] and evaluative feedback (critique) [54]. In [25], the authors combine reinforcement learning, inverse reinforcement learning and instruction based learning using policy shaping through action selection guidance. During training, an external teacher is able to formulate verbal instructions that will change a selected action to be performed in the environment. The results indicate that interaction helps to increase the learning speed, even with an impoverished Automatic Speech Recognition system.

8.4 Example: The TICS Architecture

The TICS architecture (Task-Instruction-Contingency-Shaping) [69] combines different information sources: a predefined reward function, human evaluative feedback and unlabeled instructions. Dealing with unlabeled instructions denotes that the meaning of the teaching signal is unknown to the robot. There is a lack of mutual understanding. TICS focuses on grounding the meaning of teaching signals (instructions).

TICS architecture enables a human teacher to shape a robot behavior by interactively providing it with unlabeled instructions. Grounding is performed during the task-learning process, and used simultaneously for guiding the latter (see Figure 6).

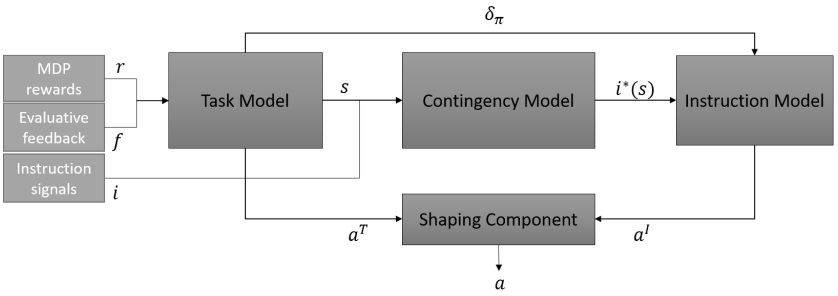


Fig. 6. The TICS architecture includes four main components: a Task Model learns the task, a Contingency Model associates task states with instruction signals, an Instruction Model interprets instructions, and a Shaping Component combines the outputs of the Task Model and the Instruction Model for decision-making (adapted from [69]).

Finally, TICS combines:

- standard RL with a predefined reward function (MDP rewards),
- learning from evaluative feedback based on policy shaping approach,
- learning from instructions, which are represented as a policy.

The Contingency Model plays a key role in grounding and it is defined for a given state as a probability distribution over detected teaching signals. In [69], a co-occurrence matrix is employed to estimate this probability distribution. TICS enables human teachers to employ unlabeled instructions as well as evaluative feedback during the interactive learning process. The robot is also learning from its own experience using the predefined reward function. The results show that in addition to the acceleration of learning, TICS offers more adaptability to the preferences of the teacher.

8.5 Limitations

Learning from instructions suffers from similar limitations as learning from evaluative feedback and demonstrations. Here, we report some specific limitations:

- *Speech recognition*: Instructions are usually provided through natural spoken language. Speech recognition is challenging in human-robot interaction contexts. Most of the approaches are impacted by the performance of speech recognition systems [25, 76, 27].
- *Instruction-to-action mapping*: Learning from instructions is mainly formulated as an instruction-to-action mapping. Going beyond engineering based approaches requires to address the Symbol Grounding problem [39], which facilitates the correspondence between natural language instructions and agent actions [69, 2]

- *Descriptions and Explanation*: Instructions explicitly communicate about robot/agent actions by either advise or critique. Language allows other forms of teaching such as explanations [93, 74] and descriptions [22, 71]. Counterfactual explanations evaluate what would happen if an alternative action is performed. There is a less restrictive nature in a description.
- *Pragmatics*: Natural language conveys more than the literal interpretation of words [35]. There is an increasing interest in the introduction of pragmatics in interactive learning, often inspired by research in linguistics and cognitive science [34, 41, 38, 18].
- *Emotions*: Spoken language is also employed to communicate emotions, which have to be taken into account for interpretation of instructions [53].

Most of these limitations call for methods able to handle uncertainty of spoken natural language in order to build efficient interactive robot/agent learning methods.

9 Modeling Human Teaching Strategies

Interactive robot learning is a joint activity allowing robots to learn a task by exploiting observable human behaviors (Figure 3). Most of the methods assume the observation of *overt teaching signals* such as feedback, demonstrations and/or instructions. The *decoding phase*, i.e. interpretation of teaching signals (Table 1), is usually pre-specified by designers and engineers. This approach does not consider *covert communication* such as intentions, beliefs, goals, emotions or attitudes. Humans change their behavior in response to the actions of the robot they are interacting with. Several works have shown that people modify their tutoring behavior in robot-directed interaction [95, 90, 91, 47, 14]. When humans demonstrate a task to another human or agent, the demonstrations are directed not just towards the objects that are manipulated (*instrumental action*), but they are also accompanied by *ostensive communicative cues* such as eye gaze and/or modulations of the demonstrations in the space–time dimensions (*belief-directed action*). This modulation results in behaviors that might appear to be sub-optimal, such as pause, repetition and exaggeration, while they are the result of simultaneous instrumental and belief-directed actions, i.e. *performing the action and communicating about it.*

To address such limitations, several research directions have been proposed. For example, *computer vision and signal processing techniques* are largely employed in human-robot interaction for the analysis and modeling non-verbal cues (e.g. tone of the voice, gesture, facial expressions, gaze) in order to infer human's mental states, intentions, engagement, emotions and attitudes [94, 5]. Another research direction focuses on *human-decision making* by studying forward models of human decision-making and inverse models of how humans think about others decision-making [40]. The approaches draw inspiration from research on how humans interpret observed behaviors as *goal-directed actions* and address the *challenge of communication in action* [40]: actor intends to not just perform

the ordinary action, but also to convey something about it. In [40], a general mathematical framework of probabilistic inference and decision-making has been proposed to characterize the underlying beliefs, intentions and goals of communicative demonstrations. The framework is largely inspired by language in which literal interpretation and pragmatic inference of linguistic and paralinguistic content are performed. These approaches exploit models and methodologies of computational cognitive science and behavioral economics making them relevant for interactive machine learning. How humans make decisions reveal their beliefs, intentions and goals. In the following, we describe some examples of natural human teaching biases (section 9.1) and a Bayesian approach of modeling human decisions (section 9.2).

9.1 Natural human teaching biases

Human teaching biases have been observed in several situations. In section 6.8, we described several natural biases of human feedback strategies, which include delay in feedback delivery [24], influence of robot actions [62] and a tendency to generate more positive feedback than negative ones [89]. Natural deviations to optimal behaviors also occur during communicative demonstrations. In [42], the authors showed the differences in behavior when a human trainer is intentionally teaching (*showing*) versus merely *doing* the task. All these studies call for new methods that go beyond naive and literal interpretations of teaching signals (Table 1) such as *evaluating/correcting, showing and telling*.

Learning situations with children have inspired robotics researchers for the study of natural human biases during teaching others. We describe two dimensions of interest in human-robot interaction: (i) *modulation of non-verbal teacher behaviors in the space–time dimensions* (ii) *teacher training strategy*.

Non-verbal teacher behaviors modulation. Human teaching situations are characterized by significant changes in various adult behaviors such as prosody (*motherese* [83]) or motion (*motionese*). In [95], the authors compared Adult-Child / Adult-Adult / Adult-Robot Interactions. They identified significant differences in hand movement velocity, motion pauses, range of motion, and eye gaze in an Adult-Child Interaction, opposed to an Adult-Adult Interaction. This decrease is even higher in the Adult-Robot Interaction.

There is a large body of research in cognitive science showing that humans are specifically efficient in the communication of generic knowledge to other individuals. This ability has been described in [26] as a communication system called the *'Natural Pedagogy'*. This work and others show that humans seem inherently sensitive to *ostensive communicative signals* such as eye gaze, gesture as well as tone of the voice. In particular, these works show that contingency of ostensive signals is a natural characteristic of social interaction in humans, which has inspired several studies of human teaching behavior [96].

Teacher training strategy. Interactive learning requires that human teachers organize the way they provide training examples to the robot. They select

which examples to present and in which order to present them to the robot in the form of instructions, demonstrations and/or feedback. Organization of a training strategy is termed *curriculum learning* [8], which is a key element in human teaching. Several human curriculum learning strategies have been observed and often result in gradually increasing the level of task complexity, i.e. presentation of simple examples then more complex examples. AI and Machine learning techniques have been derived from this notion of curriculum learning with the idea that guiding training will significantly increase the learning speed of artificial agents. Several works are focusing on the question of how to effectively teach agents with the emergence of *computational machine teaching* as an inverse problem of machine learning [101].

Example: In [46], the authors conduct a study in which participants are asked to teach a robot the concept of "graspability", i.e. if an object can be grasped or not with one hand. To teach the binary task (graspable vs. not graspable), the participants are provided several cards with photos of common objects (e.g., food, furniture, and animals). They have to pick up the cards from the table and show them to the robot while teaching them the concept of "graspability". The authors observed three different human teaching strategies [46]: (1) the extreme strategy, which starts with objects with extreme ratings and gradually moves toward the decision boundary; (2) the linear strategy, which follows a prominent left-to-right or right-to-left sequence; and (3) the positive-only strategy, which involves only positively labeled examples. Building up on such observations, they propose a computational framework as a potential explanation for the teaching strategy that follows a curriculum learning principle.

In [91], observation of various human teaching strategies raised the following question: can we influence humans to teach optimally? The authors developed Teaching Guidance algorithms that allow robots to generate instructions for the human teacher in order to improve their input. They performed experiments to compare human teaching with and without teaching guidance and show that Teaching Guidance substantially improves the data provided by teachers. The experiments demonstrate that humans are not spontaneously as good as computational teachers.

9.2 A noisily-rational decision model

A more accurate model of human behavior would help in interpreting, anticipating and predicting behaviors, in particular when they are not optimal. A common approach is to formalize human intent via a reward function, and assume that the human will act rationally with regard to the reward function [45, 55]. The Boltzmann noisily-rational decision model [59] is often employed for this purpose. This model assumes that people choose trajectories in proportion to their exponentiated reward.

The Boltzmann noisily-rational decision model quantifies the likelihood that a human will select any particular option $o \in O$ (e.g. any teaching signal or

example). If each option o has an underlying reward $R(o)$, the Boltzmann model computes the desirability of an option as:

$$P(o) = \frac{e^{R(o)}}{\sum_{i \in O} e^{R(i)}} \tag{7}$$

Interactive robot learning usually considers a sequence of human states and actions, called a trajectory $\tau = (s_1, a_1, ..., s_T, a_T)$. Boltzmann noisily-rational decision model is usually approximated to estimate the probability the human will take a trajectory is proportional to exponentiated return times a "rationality coefficient" β ("inverse temperature"):

$$P(\tau) \approx \exp \left\{ \beta \sum_{t=1}^{T} \gamma^t R(s_t, a_t) \right\} \tag{8}$$

The rationality coefficient β captures how good an optimizer the human is. The following values are usually considered:

- $\beta = 0$ would yield the uniform distribution capturing a random human type;
- $\beta \to \infty$ would yield a perfectly rational human type.

The goal of a noisily-rational decision mode is to draw inference from the observation of human actions. In [45], the authors introduce a formalism exploiting such a framework to interpret different types of human behaviors called *the reward-rational choice*. Within this framework, the robot is able to interpret a large range of teaching signals such as demonstrations, reward/punishment and instructions. However, several works have demonstrated the limits of the Boltzmann based modeling approach (see for example [9]): needs to identify alternatives of an option (equation 7), suited for trajectories but not policies [55], and only one parameter β is employed to model rationality. In addition, the approach does not take into account human reactions to machine actions (i.e. continuous adaption of human partner to the robot). As described in [40], there is a need to build better *Human Models of Machines* to address such issues by exploiting recent research in cognitive science.

10 Applications

In this section, we illustrate how Interactive Robot Learning Algorithms have been applied to several domains such as *Interactive Task Learning, Behavior Generation/ Composition*, and *Socially Assistive Robotics*. In all these applications, it is necessary to define the role of the human, the agent and their interactions during the learning and use phases.

10.1 Interactive Task Learning

Interactive Task Learning (ITL) [56] is one of the main areas of applications of methods and algorithms presented in this chapter. The aim of ITL is to develop

agents able to learn a task through natural interaction with a human instructor. Numerous applications have been considered and some of them transferred to industry.

A use-case, *object sorting task*, is presented in figure 7. The robot has to learn to sort two types of objects: *Plain* (left) and *Pattern* (right), with two different sizes and three colors. In [69], the TICS architecture (Section 8.4) has been employed to achieve this learning with a human teacher. The results show that the interactive learning perspective facilitates *programming* by including real-time interaction with humans, and improves flexibility and adaptability of communication.

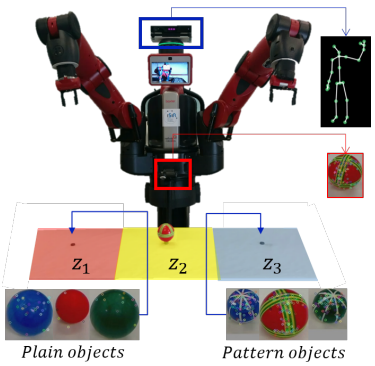


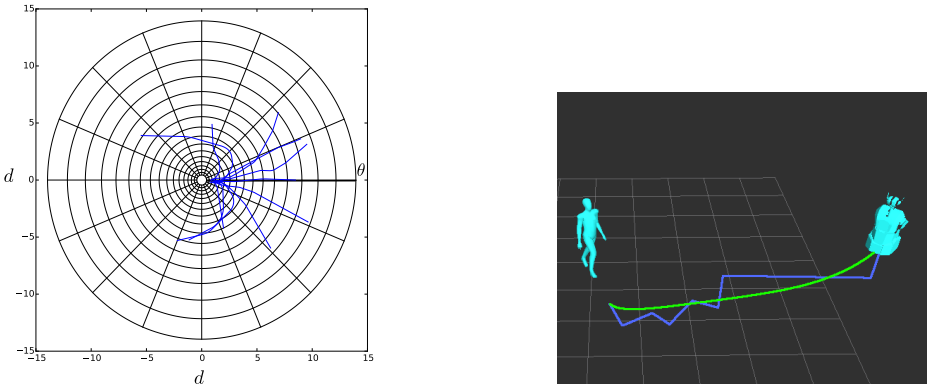
Fig. 7. Teaching object sorting (adapted from [66]).

10.2 Learning robot behaviors from human demonstrations

Figure 8 describes how Inverse Reinforcement Learning (IRL) is employed to learn robot behaviors: how a mobile robot approaches a moving human?. As described in section 7.7, IRL starts with the recording of human demonstrations. In [79], demonstrations collection is performed off-line: humans were asked to teleoperate a mobile robot with the aim of approaching humans (Figure 8a). Based on such demonstration, IRL is then used to learn a reward function that could replicate navigation behaviors. This work contributes to the challenge of generating legible robot motion [28] that should be addressed with a human-in-the-loop perspective [97].

10.3 Learning robot behaviors from human instructions

Spoken language based programming of robots has been described in section 8.3. Humans simply describe actions and behaviors through natural language, which take the form of instructions. The ambition is to go beyond simple commands



(a) Human demonstrations of the robot approaching the target person.

(b) Robot approaching solution using IRL. A smooth trajectory is generated using a Bézier curve.

Fig. 8. Learning how to approach humans using Inverse Reinforcement Learning (adapted from [79]).

and enable end-users to create new personalized behaviors through natural interactions. In [76], a cognitive architecture has been developed for this purpose (Figure 9). This approach has been successfully transferred to industry and implemented in a Pepper robot with SoftBank Robotics. The cognitive architecture has been evaluated within the Pepper@Home program. The results demonstrate that end-users were able to create their own behaviors and share them with other end-users.

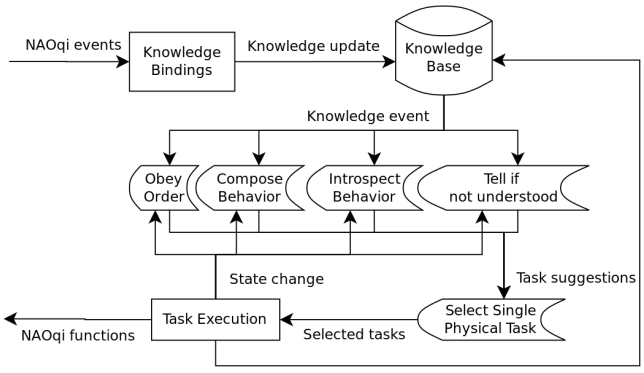


Fig. 9. Teaching new robot behaviors using natural language (adapted from [76]).

10.4 Socially Assistive Robotics

There is a substantial body of work in the design and development of socially assistive robotics where interactive robot learning plays an important role. The research works explore whether a robot could facilitate reciprocal social interaction in cases in which the robot was more predictable, attractive and simple [29, 32, 11]. In figure 10, imitation learning has been employed with children with Autism Spectrum Disorders. The children are asked to teach robots new postures, which is unusual in rehabilitation. Having the children teaching the robot has been shown to be relevant in rehabilitation and education (*protégé effect*) (see [33] for a use-case on dysgraphia). Contrary to the usual assumption of interactive robot learning, the challenge is to learn with children who do not fully master the task. The impact of the human teacher has been analyzed and exploited to capture *individual social signatures* [10] (Figure 10). Using different imitation experiments (posture, facial expressions, avatar-robot), we were able to assess the impact of individual partners in the learning.

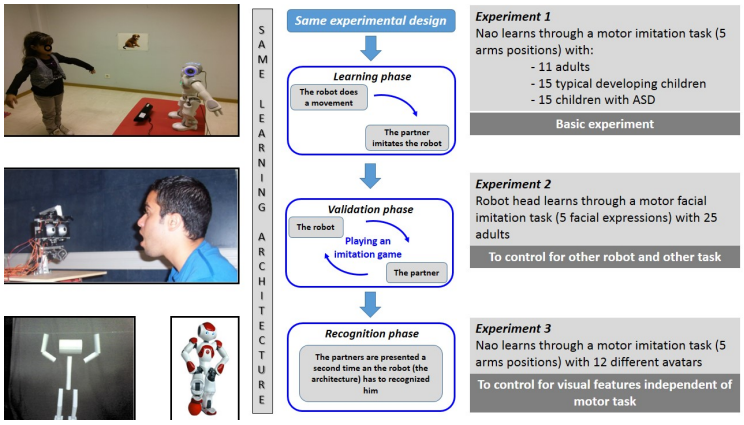


Fig. 10. Exploitation of Imitation Learning in Psychiatry, Developmental and Cognitive Sciences (adapted from [10]).

11 Conclusions, challenges and perspectives

In this chapter, we introduced the main concepts of Interactive Robot Learning, which were illustrated by several examples and applications. The chapter shows that the design of human interventions is as important as the design of algorithms in the efficiency of interactive robot learning. There is a substantial body of literature documenting interactive robot/machine learning methods and models (see also relevant readings in Section 2.4).

In the following, we report some relevant challenges of the domain shared by all human intervention strategies.

Autonomous - Interactive Learning. Collecting human teaching signals is expensive and a limited number of teaching signals could be collected. This issue refers to sample efficiency of algorithms and most of the interactive approaches are sample inefficient since they require a large number of human interventions. In addition, introducing human interventions in algorithms not designed for such purpose raises new issues not always expected (e.g., credit assignment, reward hacking, non-optimal human behaviors, implicit feedback or emotion). There has been several attempts aiming at combining rewards from multiple sources such as from the human and the environment.

This leads to challenges at the core of Human Centered AI, in particular how to build hybrid teams that are more efficient than each team member. In addition, following a Human Centered AI perspective, the aim is to augment the human and not to replace it. In [85], we discuss a range of reinforcement learning agents equipped with different skills that could handle such situations (including intrinsic motivations [77] and curriculum learning).

Evaluation and Replicability. Comparing, selecting and adapting models is essential in Interactive Robot Learning. Machine Learning and Robotics have set-up evaluation metrics that could be employed there. Similar to Human-Robot Interaction (HRI) methodology, there is a need to complement the evaluation by *human oriented metrics*, which could include questionnaires (see for example engagement evaluation in HRI [73]) as well as as assessment of interaction load (e.g., number of interactions [69]).

Another important factor allowing to obtain consistent results is replicability (repeatability + reproducibility). Compared to standard machine learning, the data are collected during training through interaction with humans. In Interactive Robot Learning, data collection is about collecting data from both humans & robots. As mentioned in section 7, recording of teaching signals is part of the interaction design and exploit various modalities: speech signals, gesture, robot motion... Consequently standardization of data collection is challenging. In addition, as mentioned in section 6.8, reciprocal (human-robot) interdependence is observed during interactive robot learning scenarios, which impacts repeatability of experiments. This calls for new ways of collecting and sharing data for improving reproducibility of works.

Multiple and Multimodal Human Interventions. The most common strategies are certainly to learn new tasks by providing feedback, demonstration or instructions (Table 1). However, many other interventions have been considered in the literature such as human preference given in regards to pairwise agent behaviors [77, 100], (joint) attention to states and/or actions during learning [100, 31] and expressing hierarchy of a task [100].

Humans are per definition multimodal and should be enabled to produce multimodal teaching signals. Most of the current works consider only one modality such as facial expressions [15, 10], gesture [69], social signals [65], or physiological signals [3].

Grounding multiple and multimodal teaching signals into actions is required to facilitate interpretation and learning. How to design interactions, interfaces, methods and models able to handle multimodal teaching signals is still an open question.

Mutual Understanding. The ability of agents to predict others and to be predicted by others is usually referred to as *mutual understanding*, which plays an important role in collaborative human-robot settings [44]. This could be achieved by transparency/explainability of robots [19, 98] through the generation of verbal and non-verbal cues including gaze [7], legible motion [28], emotional expressions [16], queries/questions and dialog [20, 17].

Being able to understand humans is also required and this will result in better *Computational Human Models of Machines*. In such direction, several recent contributions deal with *inferential social learning* to understand how humans think, plan, and act during interactive robot/agent learning [45, 58, 18, 97, 40]. This calls for new interdisciplinary research grounded in Cognitive Science, Machine Learning and Robotics [40, 85].

References

1. Aigrain, J., Spodenkiewicz, M., Dubuisson, S., Detyniecki, M., Cohen, D., Chetouani, M.: Multimodal stress detection from multiple assessments. IEEE Transactions on Affective Computing **9**(4), 491–506 (2018). https://doi.org/10.1109/TAFFC.2016.2631594
2. Akakzia, A., Colas, C., Oudeyer, P.y., Chetouani, M., Sigaud, O.: Grounding Language to Autonomously-Acquired Skills via Goal Generation. In: ICLR 2021 - Ninth International Conference on Learning Representation. Vienna / Virtual, Austria (May 2021)
3. Akinola, I., Wang, Z., Shi, J., He, X., Lapborisuth, P., Xu, J., Watkins-Valls, D., Sajda, P., Allen, P.: Accelerated robot learning via human brain signals. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). pp. 3799–3805 (2020). https://doi.org/10.1109/ICRA40945.2020.9196566
4. Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T.: Power to the people: The role of humans in interactive machine learning. AI Magazine **35**(4), 105–120 (Dec 2014). https://doi.org/10.1609/aimag.v35i4.2513
5. Anzalone, S.M., Boucenna, S., Ivaldi, S., Chetouani, M.: Evaluating the engagement with social robots. International Journal of Social Robotics **7**(4), 465–478 (2015)
6. Argall, B.D., Chernova, S., Veloso, M., Browning, B.: A Survey of Robot Learning from Demonstration. Robot. Auton. Syst. **57**(5), 469–483 (May 2009). https://doi.org/10.1016/j.robot.2008.10.024
7. Belkaid, M., Kompatsiari, K., Tommaso, D.D., Zablith, I., Wykowska, A.: Mutual gaze with a robot affects human neural activity and delays decision-making processes. Science Robotics **6**(58), eabc5044 (2021). https://doi.org/10.1126/scirobotics.abc5044
8. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: Curriculum learning. In: Proceedings of the 26th Annual International Conference on Machine Learning. p. 41–48. ICML '09, Association for Computing Machinery, New York, NY, USA (2009). https://doi.org/10.1145/1553374.1553380

9. Bobu, A., Scobee, D.R.R., Fisac, J.F., Sastry, S.S., Dragan, A.D.: Less is more: Rethinking probabilistic models of human behavior. In: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. p. 429–437. HRI '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3319502.3374811
10. Boucenna, S., D., C., Gaussier, P., Meltzoff, A.N., Chetouani, M.: Robots learn to recognize individuals from imitative encounters with people and avatars. Scientific Reports (Nature Publishing Group) **srep19908** (2016)
11. Boucenna, S., Anzalone, S., Tilmont, E., Cohen, D., Chetouani, M.: Learning of social signatures through imitation game between a robot and a human partner. IEEE Transactions on Autonomous Mental Development **6**(3), 213–225 (2014). https://doi.org/10.1109/TAMD.2014.2319861
12. Branavan, S.R.K., Chen, H., Zettlemoyer, L.S., Barzilay, R.: Reinforcement Learning for Mapping Instructions to Actions. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1. pp. 82–90. ACL '09, Association for Computational Linguistics, Stroudsburg, PA, USA (2009)
13. Bratman, M.E.: Intention and personal policies. Philosophical Perspectives **3**, 443–469 (1989)
14. Breazeal, C., Thomaz, A.L.: Learning from human teachers with Socially Guided Exploration. In: 2008 IEEE International Conference on Robotics and Automation. pp. 3539–3544 (May 2008). https://doi.org/10.1109/ROBOT.2008.4543752
15. Broekens, J.: Emotion and reinforcement: Affective facial expressions facilitate robot learning. In: Huang, T.S., Nijholt, A., Pantic, M., Pentland, A. (eds.) Artifical Intelligence for Human Computing. pp. 113–132. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
16. Broekens, J., Chetouani, M.: Towards transparent robot learning through tdrl-based emotional expressions. IEEE Transactions on Affective Computing **12**(2), 352–362 (2021). https://doi.org/10.1109/TAFFC.2019.2893348
17. Cakmak, M., Thomaz, A.L.: Designing robot learners that ask good questions. In: 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp. 17–24 (Mar 2012). https://doi.org/10.1145/2157689.2157693
18. Caselles-Dupré, H., Sigaud, O., Chetouani, M.: Pragmatically learning from pedagogical demonstrations in multi-goal environments (2022). https://doi.org/10.48550/arxiv.2206.04546
19. Chakraborti, T., Kulkarni, A., Sreedharan, S., Smith, D.E., Kambhampati, S.: Explicability? legibility? predictability? transparency? privacy? security? the emerging landscape of interpretable agent behavior. Proceedings of the International Conference on Automated Planning and Scheduling **29**(1), 86–96 (May 2018)
20. Chao, C., Cakmak, M., Thomaz, A.L.: Transparent active learning for robots. In: 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp. 317–324 (Mar 2010). https://doi.org/10.1109/HRI.2010.5453178
21. Chernova, S., Thomaz, A.L.: Robot learning from human teachers. Synthesis Lectures on Artificial Intelligence and Machine Learning **8**(3), 1–121 (2014)
22. Colas, C., Karch, T., Lair, N., Dussoux, J.M., Moulin-Frier, C., Dominey, P.F., Oudeyer, P.Y.: Language as a cognitive tool to imagine goals in curiosity-driven exploration. arXiv preprint arXiv:2002.09253 (2020)
23. Colombetti, M., Dorigo, M., Borghi, G.: Behavior analysis and training-a methodology for behavior engineering. IEEE Transactions on Systems,

Man, and Cybernetics, Part B (Cybernetics) **26**(3), 365–380 (Jun 1996). https://doi.org/10.1109/3477.499789
24. Cruz, C.A., Igarashi, T.: A survey on interactive reinforcement learning: Design principles and open challenges. Proceedings of the 2020 ACM Designing Interactive Systems Conference (2020)
25. Cruz, F., Twiefel, J., Magg, S., Weber, C., Wermter, S.: Interactive reinforcement learning through speech guidance in a domestic scenario. In: 2015 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (Jul 2015). https://doi.org/10.1109/IJCNN.2015.7280477
26. Csibra, G., Gergely, G.: Natural pedagogy. Trends in Cognitive Sciences **13**, 148 – 153 (2009)
27. Dominey, P., Mallet, A., Yoshida, E.: Real-time spoken-language programming for cooperative interaction with a humanoid apprentice. I. J. Humanoid Robotics **6**, 147–171 (06 2009). https://doi.org/10.1142/S0219843609001711
28. Dragan, A.D., Lee, K.C., Srinivasa, S.S.: Legibility and predictability of robot motion. In: 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp. 301–308. IEEE (2013). https://doi.org/10.1109/HRI.2013.6483603
29. Duquette, A., Michaud, F., Mercier, H.: Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. Autonomous Robots **24**(2), 147–157 (2008)
30. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E.: Ai4people—an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. Minds and Machines **28**(4), 689–707 (2018). https://doi.org/10.1007/s11023-018-9482-5
31. Fournier, P., Sigaud, O., Chetouani, M.: Combining artificial curiosity and tutor guidance for environment exploration. In: Workshop on Behavior Adaptation, Interaction and Learning for Assistive Robotics at IEEE RO-MAN 2017. Lisbon, Portugal (2017), https://hal.archives-ouvertes.fr/hal-01581363
32. Fujimoto, I., Matsumoto, T., De Silva, P.R.S., Kobayashi, M., Higashi, M.: Mimicking and evaluating human motion to improve the imitation skill of children with autism through a robot. International Journal of Social Robotics **3**(4), 349–357 (2011)
33. Gargot, T., Asselborn, T., Zammouri, I., Brunelle, J., Johal, W., Dillenbourg, P., Archambault, D., Chetouani, M., Cohen, D., Anzalone, S.M.: "it is not the robot who learns, it is me." treating severe dysgraphia using child-robot interaction. Frontiers in Psychiatry **12** (2021). https://doi.org/10.3389/fpsyt.2021.596055
34. Goodman, N.D., Frank, M.C.: Pragmatic language interpretation as probabilistic inference. Trends in Cognitive Sciences **20**(11), 818–829 (2016). https://doi.org/10.1016/j.tics.2016.08.005
35. Grice, H.P.: Logic and conversation. In: Cole, P., Morgan, J.L. (eds.) Syntax and Semantics: Vol. 3: Speech Acts, pp. 41–58. Academic Press, New York (1975)
36. Griffith, S., Subramanian, K., Scholz, J., Isbell, C.L., Thomaz, A.: Policy Shaping: Integrating Human Feedback with Reinforcement Learning. In: Proceedings of the 26th International Conference on Neural Information Processing Systems. pp. 2625–2633. NIPS'13, Curran Associates Inc., USA (2013)
37. Grizou, J., Iturrate, I., Montesano, L., Oudeyer, P.Y., Lopes, M.: Interactive Learning from Unlabeled Instructions. In: Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence. pp. 290–299. UAI'14, AUAI Press, Arlington, Virginia, United States (2014)

38. Gweon, H.: Inferential social learning: cognitive foundations of human social learning and teaching. Trends in Cognitive Sciences (2021)
39. Harnad, S.: The symbol grounding problem. Physica D **42**, 335–346 (1990)
40. Ho, M., Griffiths, T.: Cognitive science as a source of forward and inverse models of human decisions for robotics and control. Annual Review of Control, Robotics, and Autonomous Systems **5** (05 2022). https://doi.org/10.1146/annurev-control-042920-015547
41. Ho, M.K., Cushman, F., Littman, M.L., Austerweil, J.L.: Communication in action: Planning and interpreting communicative demonstrations (2019)
42. Ho, M.K., Littman, M.L., Cushman, F., Austerweil, J.L.: Teaching with Rewards and Punishments: Reinforcement or Communication? In: Proceedings of the 37th Annual Meeting of the Cognitive Science Society (Jul 2015)
43. Ho, M.K., MacGlashan, J., Littman, M.L., Cushman, F.: Social is special: A normative framework for teaching with and learning from evaluative feedback. Cognition **167**, 91–106 (2017)
44. Jacq, A.D., Magnan, J., Ferreira, M.J., Dillenbourg, P., Paiva, A.: Sensitivity to perceived mutual understanding in human-robot collaborations. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. p. 2233–2235. AAMAS '18, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2018)
45. Jeon, H.J., Milli, S., Dragan, A.: Reward-rational (implicit) choice: A unifying formalism for reward learning. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS'20, Curran Associates Inc., Red Hook, NY, USA (2020)
46. Khan, F., Zhu, X., Mutlu, B.: How do humans teach: On curriculum learning and teaching dimension. In: Proceedings of the 24th International Conference on Neural Information Processing Systems. p. 1449–1457. NIPS'11, Curran Associates Inc., Red Hook, NY, USA (2011)
47. Knox, W.B., Stone, P.: Reinforcement learning from human reward: Discounting in episodic tasks. In: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. pp. 878–885 (Sep 2012). https://doi.org/10.1109/ROMAN.2012.6343862
48. Knox, W.B., Breazeal, C., Stone, P.: Learning from feedback on actions past and intended. In: In Proceedings of 7th ACM/IEEE International Conference on Human-Robot Interaction, Late-Breaking Reports Session (HRI 2012) (Mar 2012)
49. Knox, W.B., Stone, P.: Interactively Shaping Agents via Human Reinforcement: The TAMER Framework. In: Proceedings of the Fifth International Conference on Knowledge Capture. pp. 9–16. K-CAP '09, ACM, New York, NY, USA (2009). https://doi.org/10.1145/1597735.1597738
50. Knox, W.B., Stone, P.: Combining Manual Feedback with Subsequent MDP Reward Signals for Reinforcement Learning. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1. pp. 5–12. AAMAS '10, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2010)
51. Knox, W.B., Stone, P., Breazeal, C.: Training a Robot via Human Feedback: A Case Study. In: Proceedings of the 5th International Conference on Social Robotics - Volume 8239. pp. 460–470. ICSR 2013, Springer-Verlag New York, Inc., New York, NY, USA (2013)
52. Kober, J., Bagnell, J.A., Peters, J.: Reinforcement Learning in Robotics: A Survey. Int. J. Rob. Res. **32**(11), 1238–1274 (Sep 2013). https://doi.org/10.1177/0278364913495721

53. Krening, S., Harrison, B., Feigh, K.M., Isbell, C.L., Riedl, M., Thomaz, A.: Learning from explanations using sentiment and advice in rl. IEEE Transactions on Cognitive and Developmental Systems **9**(1), 44–55 (March 2017). https://doi.org/10.1109/TCDS.2016.2628365
54. Krening, S., Feigh, K.M.: Interaction algorithm effect on human experience with reinforcement learning. J. Hum.-Robot Interact. **7**(2) (oct 2018). https://doi.org/10.1145/3277904
55. Laidlaw, C., Dragan, A.D.: The boltzmann policy distribution: Accounting for systematic suboptimality in human models. ArXiv **abs/2204.10759** (2022)
56. Laird, J.E., Gluck, K., Anderson, J., Forbus, K.D., Jenkins, O.C., Lebiere, C., Salvucci, D., Scheutz, M., Thomaz, A., Trafton, G., Wray, R.E., Mohan, S., Kirk, J.R.: Interactive task learning. IEEE Intelligent Systems **32**(4), 6–21 (2017). https://doi.org/10.1109/MIS.2017.3121552
57. Lepri, B., Oliver, N., Pentland, A.: Ethical machines: The human-centric use of artificial intelligence. iScience **24**(3), 102249 (2021). https://doi.org/doi.org/10.1016/j.isci.2021.102249
58. Lin, J., Fried, D., Klein, D., Dragan, A.: Inferring rewards from language in context (2022). https://doi.org/10.48550/arxiv.2204.02515
59. Luce, R.D.: The choice axiom after twenty years. Journal of Mathematical Psychology **15**, 215–233 (1977)
60. Luketina, J., Nardelli, N., Farquhar, G., Foerster, J.N., Andreas, J., Grefenstette, E., Whiteson, S., Rocktäschel, T.: A survey of reinforcement learning informed by natural language. In: Kraus, S. (ed.) Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019. pp. 6309–6317. ijcai.org (2019). https://doi.org/10.24963/ijcai.2019/880
61. Lungarella, M., Metta, G., Pfeifer, R., Sandini, G.: Developmental robotics: a survey. Connection Science **15**(4), 151–190 (2003). https://doi.org/10.1080/09540090310001655110
62. MacGlashan, J., Ho, M.K., Loftin, R., Peng, B., Wang, G., Roberts, D.L., Taylor, M.E., Littman, M.L.: Interactive learning from policy-dependent human feedback. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. pp. 2285–2294. JMLR. org (2017)
63. Matuszek, C., Herbst, E., Zettlemoyer, L., Fox, D.: Learning to Parse Natural Language Commands to a Robot Control System. In: Desai, J.P., Dudek, G., Khatib, O., Kumar, V. (eds.) Experimental Robotics: The 13th International Symposium on Experimental Robotics, pp. 403–415. Springer International Publishing, Heidelberg (2013)
64. Mitsunaga, N., Smith, C., Kanda, T., Ishiguro, H., Hagita, N.: Robot behavior adaptation for human-robot interaction based on policy gradient reinforcement learning. In: 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 218–225 (2005). https://doi.org/10.1109/IROS.2005.1545206
65. Moerland, T.M., Broekens, J., Jonker, C.M.: Emotion in reinforcement learning agents and robots: A survey. Mach. Learn. **107**(2), 443–480 (feb 2018). https://doi.org/10.1007/s10994-017-5666-0
66. Najar, A.: Shaping robot behaviour with unlabeled human instructions. Ph.D. thesis, Paris 6 (2017)
67. Najar, A., Chetouani, M.: Reinforcement learning with human advice. a survey. arXiv preprint arXiv:2005.11016 (2020)

68. Najar, A., Chetouani, M.: Reinforcement Learning With Human Advice: A Survey. Frontiers in Robotics and AI (Jun 2021). https://doi.org/10.3389/frobt.2021.584075
69. Najar, A., Sigaud, O., Chetouani, M.: Interactively shaping robot behaviour with unlabeled human instructions. Auton. Agents Multi Agent Syst. **34**(2), 35 (2020). https://doi.org/10.1007/s10458-020-09459-6
70. Ng, A.Y., Russell, S.J.: Algorithms for Inverse Reinforcement Learning. In: Proceedings of the Seventeenth International Conference on Machine Learning. pp. 663–670. ICML '00, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2000)
71. Nguyen, K., Misra, D., Schapire, R.E., Dudak, M., Shafto, P.: Interactive learning from activity description. In: 2021 International Conference on Machine Learning (July 2021)
72. Nicolescu, M.N., Mataric, M.J.: 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. pp. 241–248. AAMAS '03, ACM (2003). https://doi.org/10.1145/860575.860614
73. Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C., Peters, C.: Engagement in human-agent interaction: An overview. Frontiers in Robotics and AI **7**, 92 (2020). https://doi.org/10.3389/frobt.2020.00092
74. Olson, M.L., Khanna, R., Neal, L., Li, F., Wong, W.K.: Counterfactual state explanations for reinforcement learning agents via generative deep learning. Artificial Intelligence **295**, 103455 (2021). https://doi.org/10.1016/j.artint.2021.103455
75. Osa, T., Pajarinen, J., Neumann, G., Bagnell, J.A., Abbeel, P., Peters, J.: An algorithmic perspective on imitation learning. Foundations and Trends in Robotics **7**(1-2), 1–179 (2018). https://doi.org/10.1561/2300000053
76. Paléologue, V., Martin, J., Pandey, A.K., Chetouani, M.: Semantic-based interaction for teaching robot behavior compositions using spoken language. In: Social Robotics - 10th International Conference, ICSR 2018, Qingdao, China, November 28-30, 2018, Proceedings. pp. 421–430 (2018). https://doi.org/10.1007/978-3-030-05204-1_41
77. Poole, B., Lee, M.: Towards intrinsic interactive reinforcement learning (2021). https://doi.org/10.48550/ARXIV.2112.01575
78. Pradyot, K.V.N., Manimaran, S.S., Ravindran, B., Natarajan, S.: Integrating Human Instructions and Reinforcement Learners: An SRL Approach. Proceedings of the UAI workshop on Statistical Relational AI (2012)
79. Ramírez, O.A.I., Khambhaita, H., Chatila, R., Chetouani, M., Alami, R.: Robots learning how and where to approach people. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 347–353 (2016). https://doi.org/10.1109/ROMAN.2016.7745154
80. Ravichandar, H., Polydoros, A.S., Chernova, S., Billard, A.: Recent advances in robot learning from demonstration. Annual Review of Control, Robotics, and Autonomous Systems **3**(1), 297–330 (2020). https://doi.org/10.1146/annurev-control-100819-063206
81. Ross, S., Bagnell, D.: Efficient reductions for imitation learning. In: Teh, Y.W., Titterington, M. (eds.) Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 9, pp. 661–668. PMLR, Chia Laguna Resort, Sardinia, Italy (13–15 May 2010)

82. Rybski, P.E., Yoon, K., Stolarz, J., Veloso, M.M.: Interactive robot task training through dialog and demonstration. In: 2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp. 49–56 (Mar 2007). https://doi.org/10.1145/1228716.1228724
83. Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M.C., Cohen, D.: Motherese in interaction: At the cross-road of emotion and cognition? (a systematic review). PLOS ONE **8**(10), null (10 2013). https://doi.org/10.1371/journal.pone.0078103
84. Scheutz, M.: The case for explicit ethical agents. AI Magazine **38**(4), 57–64 (Dec 2017). https://doi.org/10.1609/aimag.v38i4.2746
85. Sigaud, O., Caselles-Dupré, H., Colas, C., Akakzia, A., Oudeyer, P., Chetouani, M.: Towards teachable autonomous agents. CoRR **abs/2105.11977** (2021), https://arxiv.org/abs/2105.11977
86. Sumers, T.R., Ho, M.K., Griffiths, T.L.: Show or tell? demonstration is more robust to changes in shared perception than explanation (2020). https://doi.org/10.48550/ARXIV.2012.09035, https://arxiv.org/abs/2012.09035
87. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press (1998)
88. Tellex, S., Kollar, T., Dickerson, S., Walter, M.R., Banerjee, A.G., Teller, S., Roy, N.: Approaching the symbol grounding problem with probabilistic graphical models. AI Magazine **32**(4), 64–76 (Dec 2011). https://doi.org/10.1609/aimag.v32i4.2384
89. Thomaz, A.L., Breazeal, C.: Asymmetric Interpretations of Positive and Negative Human Feedback for a Social Learning Agent. In: RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication. pp. 720–725 (Aug 2007). https://doi.org/10.1109/ROMAN.2007.4415180
90. Thomaz, A.L., Breazeal, C.: Reinforcement Learning with Human Teachers: Evidence of Feedback and Guidance with Implications for Learning Performance. In: Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1. pp. 1000–1005. AAAI'06, AAAI Press, Boston, Massachusetts (2006)
91. Thomaz, A.L., Breazeal, C.: Teachable robots: Understanding human teaching behavior to build more effective robot learners. Artificial Intelligence **172**(6), 716–737 (2008). https://doi.org/10.1016/j.artint.2007.09.009
92. Thomaz, A.L., Cakmak, M.: Learning About Objects with Human Teachers. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction. pp. 15–22. HRI '09, ACM, New York, NY, USA (2009). https://doi.org/10.1145/1514095.1514101
93. Tulli, S., Melo, F., Paiva, A., Chetouani, M.: Learning from explanations with maximum likelihood inverse reinforcement learning (2022). https://doi.org/10.21203/rs.3.rs-1439366/v1
94. Vinciarelli, A., Esposito, A., André, E., Bonin, F., Chetouani, M., Cohn, J.F., Cristani, M., Fuhrmann, F., Gilmartin, E., Hammal, Z., Heylen, D., Kaiser, R., Koutsombogera, M., Potamianos, A., Renals, S., Riccardi, G., Salah, A.A.: Open challenges in modelling, analysis and synthesis of human behaviour in human–human and human–machine interactions. Cognitive Computation **7**(4), 397–413 (2015). https://doi.org/10.1007/s12559-015-9326-z
95. Vollmer, A.L., Lohan, K.S., Fischer, K., Nagai, Y., Pitsch, K., Fritsch, J., Rohlfing, K.J., Wredek, B.: People modify their tutoring behavior in robot-directed interaction for action learning. In: 2009 IEEE 8th International Conference on Development and Learning. pp. 1–6 (2009). https://doi.org/10.1109/DEVLRN.2009.5175516

96. Vollmer, A., Schillingmann, L.: On studying human teaching behavior with robots: A review. Review of Philosophy and Psychology **9**(4), 863–903 (2018). https://doi.org/10.1007/s13164-017-0353-4
97. Wallkötter, S., Chetouani, M., Castellano, G.: Slot-v: Supervised learning of observer models for legible robot motion planning in manipulation. In: SLOT-V: Supervised Learning of Observer Models for Legible Robot Motion Planning in Manipulation (2022)
98. Wallkötter, S., Tulli, S., Castellano, G., Paiva, A., Chetouani, M.: Explainable embodied agents through social cues: A review. ACM Transactions on Human-Robot Interaction **10**(3) (Jul 2021). https://doi.org/10.1145/3457188
99. Warnell, G., Waytowich, N., Lawhern, V., Stone, P.: Deep tamer: Interactive agent shaping in high-dimensional state spaces. Proceedings of the AAAI Conference on Artificial Intelligence **32**(1) (Apr 2018). https://doi.org/10.1609/aaai.v32i1.11485
100. Zhang, R., Torabi, F., Warnell, G., Stone, P.: Recent advances in leveraging human guidance for sequential decision-making tasks. Autonomous Agents and Multi-Agent Systems **35**(2), 31 (2021). https://doi.org/10.1007/s10458-021-09514-w
101. Zhu, X.: Machine teaching: An inverse problem to machine learning and an approach toward optimal education. Proceedings of the AAAI Conference on Artificial Intelligence **29**(1) (Mar 2015). https://doi.org/10.1609/aaai.v29i1.9761