



Towards Teachable Autonomous Agents

Olivier Sigaud, Hugo Caselles-Dupré, Cédric Colas, Ahmed Akakzia,
Pierre-Yves Oudeyer, Mohamed Chetouani

► To cite this version:

Olivier Sigaud, Hugo Caselles-Dupré, Cédric Colas, Ahmed Akakzia, Pierre-Yves Oudeyer, et al..
Towards Teachable Autonomous Agents. 2021. hal-03364200

HAL Id: hal-03364200







<https://inria.hal.science/hal-03364200>

Preprint submitted on 4 Oct 2021

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.

TOWARDS TEACHABLE AUTONOMOUS AGENTS

 **Olivier Sigaud**⁽¹⁾,
  **Hugo Caselles-Dupré**^{(2)*},
  **Cédric Colas**^{(3)*},
  **Ahmed Akakzia**⁽¹⁾,
 **Pierre-Yves Oudeyer**⁽³⁾⁺,
  **Mohamed Chetouani**⁽¹⁾⁺

(1) Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, F-75005 Paris, France

(2) ENSTA Paris, Institut Polytechnique de Paris, Palaiseau, France

(3) INRIA Bordeaux Sud-Ouest, équipe FLOWERS

(*) equally contributing authors

(+) equally contributing senior authors

correspondence to Olivier.Sigaud@upmc.fr

ABSTRACT

Autonomous discovery and direct instruction are two extreme sources of learning in children, but educational sciences have shown that intermediate approaches such as assisted discovery or guided play resulted in better acquisition of skills. When turning to Artificial Intelligence, the above dichotomy can be translated into the distinction between autonomous agents, which learn in isolation from their own signals, and interactive learning agents which can be taught by social partners but generally lack autonomy. In between should stand *teachable autonomous agents*: agents that learn from both internal and teaching signals to benefit from the higher efficiency of assisted discovery processes. Designing such agents could result in progress in two ways. First, very concretely, it would offer a way to non-expert users in the real world to drive the learning behavior of agents towards their expectations. Second, more fundamentally, it might be a key step to endow agents with the necessary capabilities to reach general intelligence. The purpose of this paper is to elucidate the key obstacles standing in the way towards the design of such agents. We proceed in four steps. First, we build on a seminal work of Bruner to extract relevant features of the assisted discovery processes happening between a child and a tutor. Second, we highlight how current research on intrinsically motivated agents is paving the way towards teachable and autonomous agents. In particular, we focus on *autotelic* agents, i.e. agents equipped with forms of intrinsic motivations that enable them to represent, self-generate and pursue their own goals. We argue that such autotelic capabilities from the learner side are key in the discovery process. Third, we adopt a social learning perspective on the interaction between a tutor and a learner to highlight some components that are currently missing to these agents before they can be taught by ordinary people using natural pedagogy. Finally, we provide a list of specific research questions that emerge from the perspective of extending these agents with assisted learning capabilities.

INTRODUCTION

From the etymology of the word, being *autonomous* means deciding by oneself (*autos*) of its own rules (*nomos*). More generally, an agent can be said autonomous if it determines its own sensorimotor behavior or if it makes its own decisions. Autonomy matters for Artificial Intelligence (AI). Indeed, at first glance, for an agent to be intelligent, it has to be autonomous in some sense: if we always had to tell an agent what to do at each step of a sequential decision or control process, we would not consider such an agent as intelligent. Let us temporarily consider a radical definition and call “truly autonomous” an agent which would only decide what to do on its own, without any consideration for our needs or any other constraints like expectations of its user or ethics. Most probably, such an agent would

not be useful. So at first glance, *true autonomy* and *usefulness* seem to be contradictory requirements: if an autonomous agent decides what to do only on its own, how can it be useful at all?

This apparently rhetorical question can be turned into a much more practical one: if a truly autonomous agent decides its own, how can we influence it so that it can be useful anyways? A solution, as put forward in Chakraborti et al. (2017), is that autonomous agents should understand and adapt to human behavior much like humans adapt to the behavior of other humans. But how can we obtain such an adaptation? Part of the answer can be found by reading again the conclusion of the seminal paper of Alan Turing widely considered as the first published paper about Artificial Intelligence, though the term does not appear:

“It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach to understand and speak English. That process could follow the normal teaching of a child.”
Turing (1950)

Even if they are not always autonomous in the common sense—they may need caregivers to fulfill their basic living requirements—children are definitely autonomous agents. Nevertheless, we can succeed in influencing their behavior in many ways including through “normal teaching”.

But again, we cannot teach an agent if it is “truly autonomous” in a radical way. So, to solve the contradiction between the “true autonomy” and the usefulness and teachability requirements, we will relax the constraint on true autonomy. Rather, we will focus on the weaker constraint of agents which can set their own goals and learn to achieve them, a category that we call *autotelic* agents. Though an autotelic agent pursues its own goals, its behavior can still be influenced by external signals. Such an agent could become appropriate and useful if we succeed in convincing it to achieve our goals, rather than goals it has chosen by itself.

Thus this paper proposes to reach useful autonomous agents by developing *teachable autotelic agents*—agents that set their own goals but can still benefit from teaching via natural social interactions.

Putting autonomy aside, research about teachable agents already exists, in the field of interactive learning. This research is driven by the fact that an agent immersed in the everyday life of its users cannot be programmed in advance to meet all its user’s expectations, neither in advance by a designer nor online by its user, as most users generally lack the technical background to do so. Rather, these users should be able to communicate or teach their preferences and expectations in a natural way, as they would do with children Vollmer et al. (2016). But the way to teach agents is currently not natural enough. For instance, in a state-of-the-art work about semi-supervised robot exploration Chen et al. (2020), the user has to collect by hand a set of pictures showing the robot potential target contexts so as to drive its exploration process, which is quite far from natural education. More generally, interactive learning research relies a lot on Interactive Reinforcement Learning (RL), but numerous papers have already shown that the way naive human users tend to teach an agent is far from meeting the expectations of the RL framework Thomaz & Breazeal (2008a;b). Using natural teaching methods such as language description of behaviors, instructions, explanations and both verbal and non-verbal feedback to educate agents would contrast a lot with the current necessary effort to drive their learning processes. In turn, this means that these agents should be equipped with the appropriate capabilities to efficiently learn from their users. We claim that having the capability to represent and pursue goals, to autonomously imagine and select goals and to infer the goals of others and to interact with them about these goals is one of these appropriate capabilities, as these goals can play a pivotal role so that users can influence the behavior of autonomous agents towards their preferences.

Moreover, all the Interactive RL research is restricted to a form of teaching where the learning processes of the agent are directly driven by an external tutor. But children also learn a lot on their own, driven by autonomous discovery processes. It would thus be more efficient to let agents learn from their own curiosity, but to rely on education to drive this

intrinsic tendency to learn in the direction of useful and acceptable behaviors. This would result in assisted discovery processes, which have been shown to result in more efficient skill acquisition than direct instruction Yu et al. (2018).

In themselves, these practical concerns for the feasibility and efficiency of agent teaching are already strong enough incentives for laying the foundations of a dedicated research program. But there are other reasons for doing so.

First, autonomous agents immersed in human societies will need to be endowed with the sociocultural skills that are required in these ecosystems. The only way to acquire these various skills specific to regions or groups of people is by learning them through practical interactions with social partners. For these autonomous agents, being teachable means being equipped with the core capabilities to acquire such skills.

Beyond this, a deeper, more fundamental reason for combining autonomy and teachability stems from the endeavour of strong AI Lake et al. (2017). It might be the case, as put forward by the *social situatedness* vision of researchers like Vygotsky Vygotsky (1978), Bruner Bruner (1990; 1991; 2009b) and a few others Dautenhahn (1995); Zlatev (2001); Tomasello (2009) that, in addition to the capability to pursue their own goals, interactions with caregivers, tutors and mates are themselves necessary conditions for the emergence of sophisticated forms of intelligence in agents, see Lindblom & Ziemke (2003) for a review. In particular, these cultural interactions may play a crucial role in the acquisition of shared cognitive representations making sense for agents and their human partners. From that perspective, to obtain useful autonomous agents in a strong sense, we should focus on questions centered on the role of verbal and nonverbal interactions in the acquisition of representational capabilities compatible with those of social partners Vygotsky (1978); Tomasello (2009).

In this paper, as we said, we ground our view of autonomous systems in research dedicated to *intrinsically motivated* agents, and particularly to the specific category of *autotelic* agents Steels (2004) which determine and learn to pursue their own goals based on intrinsic motivations. In order to ground our investigation in developmental science, we start from a classical work about the normal teaching of a child, namely “The role of tutoring in problem solving” Wood et al. (1976). As put forward in Vollmer & Schillingmann (2018), there are numerous theories and guidelines on optimal teaching, but studies such as Wood et al. (1976) on how teaching is done naturally are relatively scarce.

Though Wood et al. (1976) is most focused on tutoring processes and provides a lot of hints on how to efficiently teach children, in this paper we focus more on the learner side of the tutoring process. We leverage observations on how natural teaching is performed to extract a list of properties one may expect from teachable autonomous agents. In Section 2, we then describe current research on Interactive Reinforcement Learning from one side and autotelic agents from the other side to show that the former provides teachable agents that lack autonomy, whereas the latter mostly provides non-teachable autonomous agents. We then outline in Section 3 the emergence of a new line of research striving to provide the best of both worlds. In particular, we describe the IMAGINE Colas et al. (2020a) and DECSTR Akakzia et al. (2021) agents as examples of autotelic agent endowed with both problem solving capabilities and primitive interactive learning capabilities in the form of sensitivity to language. In Section 4, we investigate the properties that are missing to such agents in order to become more teachable. In Section 5, we do the same about augmenting the autonomous learning capabilities of these agents. Finally, in Section 6, we extract a list of specific research questions that emerge from the perspective of extending autotelic agents with interactive learning capabilities.

1 THE NORMAL TEACHING OF A CHILD

According to educational sciences, children can learn in three ways: *unassisted discovery*, where children are left on their own to discover new things, *direct instruction* where a tutor explicitly sets the learning goals of children step by step, and *guided or assisted discovery*, where the tutor intervenes on the discovery process of children to make it more fruitful Yu et al. (2018). In this paper, we focus on the latter form of education.

“*The role of tutoring in problem solving*” Wood et al. (1976) is a classical developmental psychology paper where the authors study the latter form of education. They describe children as natural problem solvers and outline the importance for their development of being assisted by social partners who are more skillful than themselves on the tasks they are committing to. As the authors put it, “*tutorial interactions are, in short, a crucial feature of infancy and childhood.*”

In their paper, they design a study of tutoring interactions in which children of age 3, 4 and 5 have to learn how to build a wooden pyramid from a set of wooden blocks with sophisticated pairing constraints. A tutor can assist them in reaching this challenging goal. The task can be achieved in 15 elementary actions and, under assistance, children take 40 actions on average to achieve it. The tutor can intervene mainly in two ways. From one side, she can directly act on the setup, handing blocks to children, or providing *partial demonstrations* by making pairs of blocks or taking blocks apart. From the other side, she can interact with children through language, mostly pointing to errors by asking them to compare two constructions, or motivating children by asking them to do more of what they just did. Other non-verbal signals seem to exist without being explicitly studied in the paper. Indeed, the authors mention the importance of providing a positive atmosphere or refer to the role of encouraging the child to move to next steps without mentioning the way these signals are fleshed out.

From this classical study and a few additional sources from the developmental psychology literature, we highlight properties of the learning children and properties of the tutoring process itself.

1.1 PROPERTIES OF LEARNING CHILDREN

Children are autonomous

An extraordinary property of natural learning in children is that it is **open-ended**: the child is capable of solving new problems of increasing difficulty up to becoming an adult and keeps learning during a whole life. Processes of autonomous learning in infants have several properties that are fundamentally different from many current machine learning systems. Among them is **autonomy**, i.e. the capability to spontaneously explore their environments, driven by an intrinsic motivation to discover and learn new tasks and problems that they imagine and select by themselves Berlyne (1966); Gopnik et al. (1999); Chu & Schulz (2020). Crucially, there is no engineer externally imposing one target goal that they should reach, providing a curriculum for learning or a ready-to-use database of training examples. Rather, children self-select their objectives within a large, potentially open-ended, space of goals they can imagine, and collect training data by physically practicing these goals.

This tendency of children to pursue their own goals is central in Wood et al. (1976). In a first stage of the study, children are left for a time with the blocks and can play with them on their own. Subsequently, the authors list six scaffolding mechanisms which come into play in the tutoring process, three of which explicitly deal with the fact that children tend to attend to their own goals. Namely, these three mechanisms are *recruitment*, *direction maintenance* and *frustration control*, which we further describe in Section 1.2.

Another key sign that children pursue goals is that, as the authors put it, “*children understand goals before being able to produce them.*” To show this, the authors monitor the number of times children take apart blocks that are either correctly or incorrectly paired together during the problem solving process. They show that, though younger children consistently fail to build the pyramid on their own, they can recognize whether a given pair of blocks is a partial solution to the global problem, as they will less often take apart the correct ones than the incorrect ones. The authors claim that their study shows that “*comprehension precedes production*”. They explain that understanding the goal is necessary so that useful feedback can be extracted from trying to achieving it. Furthermore, they emphasize that the role of tutoring is more about helping children understand the goal than about helping them physically achieving it.

Children are few shot learners

Beyond the specific study of Wood et al. (1976), other developmental studies have revealed that children explore goals in an organized manner, attributing to them values of interest that evolve with time, and allowing them to self-define a learning curriculum that developmental psychologists call a developmental trajectory Piaget (1977); Thelen & Smith (1996); Smith & Gasser (2005). This self-generated learning curriculum makes them very **sample efficient**: they avoid spending too much time on goals that are either too easy or too difficult, focusing on goals of the right level of complexity at the right time. Besides, while children are learning, some tasks that were previously too hard become easier because children can transfer what they learned from solving one task to solving another. Based on this transfer process, children are also **few shot learners**: they can leverage what they learned on previous tasks to learn in very few attempts new tasks that are now easier. Thus these mechanisms allow children to discover highly complex skills such as biped locomotion, block stacking or tool use, which would have been extremely difficult to learn if they had directly addressed these goals before mastering simpler skills. Wood et al. (1976) do not focus on these learning efficiency aspects as their main messages are more about the tutoring process.

Children are hierarchical learners

The idea that children are hierarchical learners is pervasive in the developmental psychology literature Eppe et al. (2020). The elementary skills mastered by children are often stepping stones for discovering how to learn other skills of increasing complexity. As Bruner says “*the acquisition of skill in the human child can be fruitfully conceived as a hierarchical program in which component skills are combined into ‘higher skills’ by appropriate orchestration to meet new, more complex task requirements.*” Bruner (1973). Again, though this is not the focus of Wood et al. (1976), the task used in their study is itself hierarchical in that it involves several repetitions of the same assembly processes.

1.2 PROPERTIES OF THE TUTORING PROCESS

Up to now, we have focused on the properties of children as efficient autonomous learners as revealed by the studies of Wood et al. (1976) and beyond. We now investigate the properties of the tutoring process itself as revealed from the same study. The purpose of these investigations is not to model the tutoring process to design an artificial tutoring agent, but to extract from this perspective the properties required from a teachable agent so that it could respond to natural tutoring signals, coming either from a human or another agent.

Showing in tutoring is not for being perfectly imitated

Wood et al. (1976) provide several cues showing that natural tutoring interactions do not rely much on learning from demonstrations. They observe that, among the 30 children of their study, “*there was not a single instance of what might be called blind matching behaviour*”. Blind matching behavior is what would be observed if children were replaying the tutor’s trajectory without understanding the goal. Second, the authors mention that the only acts that children imitate are those they can already perform fairly well. That is, imitating the tutor is not a way to learn how to perform the tutor’s current action, it is more probably a way to move to the next step of the problem solving process.

Demonstrations help correcting imperfect attempts

Beyond showing what to do next, it seems that demonstrations can still play a role in learning new skills. If they are not used for blind imitation, how do they help? Looking more closely at instances of demonstrations helps understanding that they are intended to communicate an explanation of how to perform the task. “*Demonstrating or ‘modelling’ solutions to a task, [...] involves an ‘idealization’ of the act to be performed and it may involve completion or even explication of a solution already partially executed by the tutee himself*” Wood et al. (1976).

The authors further explain that the tutor is ‘imitating’ in idealized form an attempted solution tried (or assumed to be tried) by children in the expectation that they will then

‘imitate’ it back in a more appropriate form. That is, the tutor builds on the current knowledge of the learner to communicate on the parts of the problem solving process that are still inadequate.

It is tempting to conclude from these observations that providing demonstration is in fact a non-verbal way to communicate about intermediate goals rather than about the way to achieve them.

Tutoring is regulating the motivational system

Wood et al. (1976) outline that most tutoring interactions are in charge of regulating the motivational system of children to keep them engaged in targeting the goal they are expected to reach. In more detail, they identify three such processes. Through *recruitment*, the tutor should find a way so that the child engages into building the pyramid rather than any of its other own goals. Through *direction maintenance*, she should ensure that the child keeps committing to that specific goal, providing incentives to make further progress towards the goal. Finally, through *frustration control*, she should avoid that the child gets discouraged and gives up pursuing the goal.

For efficient tutoring, the tutor needs a model of the tutee

The last point that we extract from Wood et al. (1976) is that, for performing well-chosen demonstrations or efficiently regulating the motivational system of children, the tutor needs to monitor a model of the knowledge, hypotheses and performance of children as well as a model of the task itself: “*The effective tutor must have at least two theoretical models to which he must attend. One is a theory of the task or problem and how it may be completed. The other is a theory of the performance characteristics of his tutee.*”. These models help the tutor interpreting what children are trying to do, so as to efficiently help them. The authors go further and consider that this interpretation process consists in generating hypotheses about the behavior of children, something at which the believe humans excel.

Finally, and symmetrically to the perspective of this paper about having teachable machine learners, already in 1976 the authors had the visionary idea that tutoring could be performed by a machine: “*if a machine program is to be effective, it too would have to be capable of generating hypotheses in a comparable way.*”. These tutoring machines could also be obtained from learning mechanisms. We now turn to AI work dedicated to the design of these artificial learners.

2 TEACHABLE VERSUS AUTONOMOUS AGENTS

Reinforcement learning (RL) is a process by which an agent learns to solve sequential decision problems from a reward signal Sutton et al. (1998). The learning agent does not know the effect of its actions in its environment, thus it has to explore to discover which action leads to which state and to obtain positive or negative rewards. It then repeats more often the actions that have proven rewarding in some situations and avoids those which have led to punishment. Initially, RL was a model of learning by trial-and-error in animals Thorndike (1911). The RL framework was also shown to convincingly explain conditioning phenomena at the neurosciences level in monkeys Schultz et al. (1997) and from there in many other species including human subjects Daw & Doya (2006). Thus, RL seems to be a natural framework for modelling learning to solve problems in children. However, this framework suffers from several limitations, among which the lack of autonomy, boundedness, sample inefficiency and focus on individual learning. We expand on those limitations below.

Lack of autonomy

The behavior of reinforcement learning agents is fully determined by a reward function. In the standard framework, this reward function is externally provided by a human designer with some specific goal in mind. In the absence of such an externally provided signal, RL agents would learn nothing, thus they are fully dependent on the reward design. As we

outlined in Section 1.1, this is to be contrasted with children who can set their own goals and learn in full autonomy.

Boundedness

Reinforcement learning agents are designed to optimize a fixed reward function. Once this fixed function is given, the behavior of the agent should converge to some optimum determined by this function and stop changing. At first glance, the richer framework of multitask RL Caruana (1997) and its derivatives such as meta-RL Finn (2018) seem to improve over this situation but in the end, as long as the set of task is bounded, the behavior of the agent should also converge to a steady optimum. To overcome these limitations, a new line of AI research on open-ended learning has recently been proposed, where an agent receives a potentially infinite sequence of unknown tasks Doncieux et al. (2018). But the framework is not explicit on where the reward function of each task should come from. From one side, it is unclear how they could be externally provided along time. From the other side, it is hard to imagine an organized space of reward functions that an agent may explore through time to account for all the activities a human adult could learn. As we outlined in Section 1.1, this is to be contrasted with the open-ended learning process observed in children who reward themselves from reaching a potentially infinite set of goals they set to themselves in a rather organized manner, through following their own curriculum.

Sample inefficiency

Reinforcement learning is notoriously slow and sample inefficient Botvinick et al. (2019). The groundbreaking successes that have made the field popular these last years, such as playing Atari games at a human level Mnih et al. (2015), defeating world champions in difficult board games Silver et al. (2018) and more complex strategic games Jaderberg et al. (2019); Vinyals et al. (2019) or even solving difficult robotic dexterous manipulation problems Andrychowicz et al. (2018); Akkaya et al. (2019) have all been obtained with weeks of heavily parallel computation which would correspond to centuries of human training time. As we outlined in Section 1.1, this is to be contrasted with the few shot learning capability of animals Gruber et al. (2019) and humans, e.g. Csibra & Gergely (2009).

Focus on individual learning

The standard RL framework accounts for an agent learning in isolation. As we outlined in Section 1.2, this is to be contrasted with children who benefit a lot from tutoring interactions.

Now that we have established these limitations of the standard RL framework, we describe two lines of artificial agent learning research which separately overcome subsets of these limitations, namely research on interactive reinforcement learners and research on isolated autotelic agents.

2.1 INTERACTIVE REINFORCEMENT LEARNERS

Within AI, the field of interactive learning investigates and models the way a human tutor guides the learning process of an agent by providing teaching signals Breazeal & Thomaz (2008). More specifically, Interactive RL focuses on the case where the agent is an RL agent. From the perspective adopted in this paper, Interactive RL can be seen as solving several of the issues outlined above.

First of all, by definition, it is not restricted to **individual learning**. Second, one perspective on Interactive RL is that it attempts to address the **sample inefficiency** issue. According to this perspective, teaching signals such as instructions Grizou et al. (2014), advice Celemin & Ruiz-del Solar (2015), demonstrations Argall et al. (2009), guidance Najar et al. (2016); Suay & Chernova (2011) and evaluative feedback Knox & Stone (2010); Griffith et al. (2013); Najar et al. (2016) are all provided to the agent in addition to an external reward function to accelerate its learning process.

Another prominent form of interaction in Interactive RL research to accelerate learning is called “Learning from Demonstration” (LfD) Argall et al. (2009). It can be seen as an extremely simplified version of the natural situation where a tutor demonstrates a task to a child. In this approach, an expert first performs highly rewarded trajectories in the environment where the task is defined. Then, data from these trajectories are collected and fed into the replay buffer—a sort of episodic memory—of the learning agent. From this data, the agent learns an efficient policy as if it was its own memories Hester et al. (2018); Večerík et al. (2017). Rather than using RL, an alternative approach called “Behavioral Cloning” (BC) consists in performing regression—another kind of machine learning process—from the same data to directly obtain a policy which behaves like the imitated one Torabi et al. (2018). These methods are not realistic in several ways. First, they assume the experience of the expert is directly transferred into the memory of the learning agent, which cannot happen yet in real life given that both participants have a different viewpoint. Second, they assume that the expert and the learning agent share the same state space, action repertoire and dynamics of interaction with the environment, which is unlikely in real life situations where imitating agents have to solve several correspondence problems to adapt the state representation and actions of the demonstrator to their own context Nehaniv & Dautenhahn (2002). Finally, they assume that the agent can perfectly observe the states and actions of the demonstrator and imitate these actions, which raises several challenges that we cover in Section 4.

Anyways, even if this method was a simplified version of natural learning from demonstration, we already outlined in Section 1.2 that this engineering approach is an unrealistic model of how showing what to do is performed in natural tutoring interactions, as infants do not perform “blind matching”. It is more likely that natural learners recognize the goals of the partner and try on their own to produce the same goals rather than performing blind imitation. This is known as *goal emulation* Tomasello (1998); Ugur et al. (2011) and this can be accounted for in the RL framework through inverse RL processes Abbeel & Ng (2004). Among other things, this approach can help solving the correspondence problem Nehaniv & Dautenhahn (2002).

Beyond the research lines described above which are mostly dedicated to improving the sample efficiency of RL, another line of Interactive RL research addresses the **boundedness** limitation. In this approach, the tutor communicates to the agent some elements to help it infer the reward function itself such as expert trajectories Abbeel & Ng (2004) or preferences about outcomes Christiano et al. (2017) or a combination of both Ibarz et al. (2018); Pinsler et al. (2018). By doing so, agents can recover open-ended learning properties, as a human user may specify a potentially infinite set of increasingly more difficult tasks.

But we immediately see that this approach relies too much on the tutor to drive the learning process and fails to account for the **autonomy** of children. First, it is too much load for the tutor to always monitor the learning progress of an agent and to provide new tasks along a potentially infinite learning trajectory. Second, this approach does not account for the fact that children are also capable of learning new tasks on their own, by contrast with agents studied in the next section. Thus this perspective is more related to the *direct instruction* approach to education than to the assisted discovery processes that we investigate in this paper.

2.2 AUTONOMOUS REINFORCEMENT LEARNERS

The extraordinary transition from the mental life of infants to the sophisticated intelligence of human adults studied in the developmental psychology works outlined in Section 1.1 is mostly modelled in the domain of developmental robotics Weng et al. (2001); Zlatev (2001); Lungarella et al. (2003). A central line of research in the domain is interested in the design of autotelic agents Steels (2004); Schembri et al. (2007). These embodied agents interact with their environment at the sensorimotor level and are provided with the ability to represent and set their own goals and rewarding themselves when they achieve them Oudeyer et al. (2007); Forestier et al. (2017). By definition, they are endowed with a form of **autonomy**.

Fundamentally, these agents are problem solvers. Implementing their learning capabilities using RL is natural, since the RL framework provides the model of choice to account for problem solving capabilities Sutton et al. (1998).

Most of such autotelic agents are equipped with one or several goal spaces and rely on goal-conditioned RL Colas et al. (2020b) and automatic curriculum learning Portelas et al. (2020) to learn to achieve those goals along an open-ended developmental trajectory. This endows them with the capability to decide which goals to target and learn about as a function of their current abilities Florensa et al. (2018); Fournier et al. (2019); Colas et al. (2019); Racaniere et al. (2019). Thus, by contrast to Interactive RL agents, autotelic agents provide a promising solution to the **boundedness** issue: if they explore an unbounded set of goals of increasing complexity, they may end up accounting for the open-ended development of children. As a side note, another process which is considered open-ended is evolution, and some authors have suggested that one may account for open-ended evolution by optimizing novelty in a population Lehman & Stanley (2008). It is of interest to realize that the algorithms behind optimizing novelty in evolution, namely *Novelty Search* Lehman & Stanley (2011) and behind addressing novel goals in autotelic agents learning, namely *Intrinsically Motivated Goal Exploration Processes* Forestier et al. (2017), share similarities, see Sigaud & Stulp (2019).

Nevertheless, as long as they address a finite set of goals, these algorithms may eventually converge to a limited set of behaviors, as we have already outlined above for the case of multitask learning. Addressing continuous goal spaces somewhat alleviates the problem, but the process may still converge to a fixed point if the goal space itself is bounded. These limitations have recently been partially overcome either through learning incrementally new sensorimotor goal representations Laversanne-Finot et al. (2018); Etcheverry et al. (2020), or by leveraging the compositional capability of language to represent goals in potentially open abstract spaces, and imagine new goals from these open spaces Colas et al. (2020a).

Thus, similarly to children, isolated autotelic agents are able to learn to solve problems on their own and to imagine an open-ended list of problems. However, by contrast with children, they are not interactive. As a consequence, they cannot readily benefit from social guidance.

3 FIRST STEPS TOWARDS TEACHABLE AND AUTOTELIC AGENTS

As we have outlined in the previous section, on one hand interactive reinforcement learners are teachable, but they are not autonomous. On the other hand, isolated autotelic agents are autonomous, but they are not teachable. To fully benefit from the efficiency of the assisted discovery process described in Wood et al. (1976), agents should be simultaneously teachable and autonomous. In this section we describe preliminary research in this direction and we outline the limitations of the existing works in the domain.

3.1 LANGUAGE-AUGMENTED AUTOTELIC AGENTS

Language was one of the two main tutoring media in Wood et al. (1976). More generally, communication capabilities may be absolutely necessary for teaching agents in a natural way, if not for endowing them with symbolic intelligence Bruner (2009a); Taniguchi et al. (2018). If we want to teach them, it is thus mandatory to endow autotelic agents with the capability to ground language in their sensorimotor experience Cangelosi et al. (2010). This connection between sensorimotor behavior and language has recently emerged in the RL community under the form of *language-conditioned* agents Chan et al. (2019); Bahdanau et al. (2019); Cideron et al. (2019); Jiang et al. (2019); Luketina et al. (2019); Colas et al. (2020a). However, all these approaches use language as a necessary input to sensorimotor behavior and, for this reason, cannot account for the goal-directed behaviors observed in preverbal infants Mandler (1999).

As far as we know, the first language-augmented autotelic agents have been proposed very recently with the IMAGINE Colas et al. (2020a) and DECSTR Akakzia et al. (2021) agents. Both agents combine the key features of autotelic agents presented in Section 2.2: they

rely on goal-conditioned policies and their choice of goals is driven by intrinsic motivations. But they are additionally endowed with a basic language interaction capability on which we focus below.

The IMAGINE agent

The IMAGINE agent, illustrated in Figure 1, aims to discover and master possible interactions in a *Playground* environment filled with procedurally-generated objects. As it freely explores its world by pursuing its own goals, it receives simple linguistic descriptions of interesting behaviors from a simulated caretaker. From these linguistic social interactions, IMAGINE leverages both the communicative and cognitive functions of language Colas et al. (2020a).

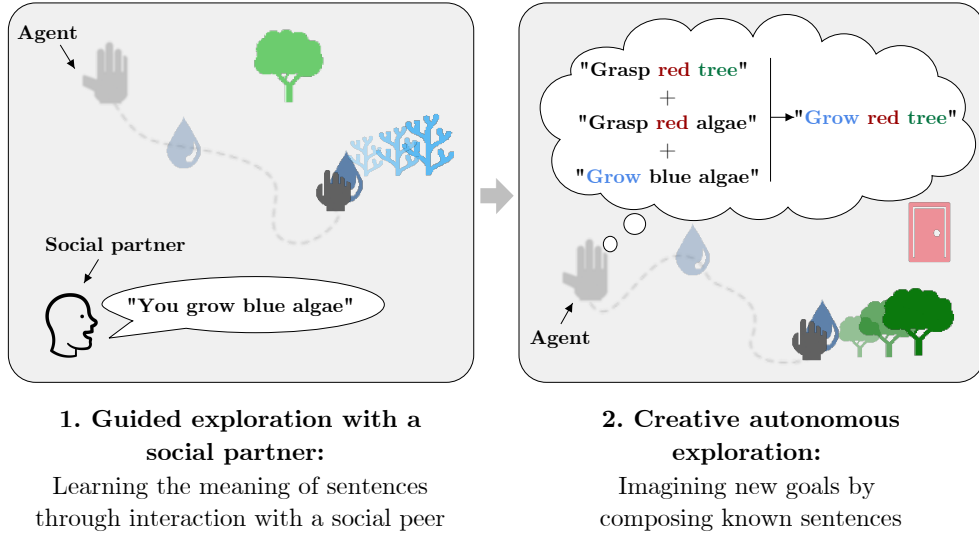


Figure 1: The IMAGINE architecture as a Vygotskian Deep RL system. The agent learns to represent and understand language as a pre-existent social structure through social interactions with a synthetic caretaker (left). This internalization of social language opens the door to a cognitive use of language. The agent can now imagine new goals as systematic recombinations of known sentences. As it pursues its own invented goals, IMAGINE creatively explores its environment (right). IMAGINE also leverages Vygotsky’s notion of *zone of proximal development* Doolittle (1997). In each episode, the caretaker sets the scene to provide optimal challenges: it introduces the necessary objects for the agent to reach its goal (not too hard), but generate them procedurally and add distracting ones (not too easy).

Let us first discuss the communicative function. If the agent hears “you grasped a red rose”, it turns this description into a potential goal and will try to grasp red roses again. To do so, it needs to understand what that means and to learn to replicate the interaction. The just-received description is an example of aligned data: a trajectory and a corresponding linguistic description. This data can be used to learn a goal-conditioned reward function, i.e. a function that helps the agent recognize when the current state matches the linguistic goal description. Given a few examples, the agent correctly recognizes when goals are achieved and can learn a policy to perform the required interaction via standard RL using self-generated goals and rewards. Here, IMAGINE uses the communicative function of language, it learns to represent the embedding of goals and goal-conditioned reward function from linguistic social interactions. As the social partner decides to describe some interactions and not others, it effectively guides the goal representations of the agent. IMAGINE is thus a *teachable autonomous agent*: its goal representations are influenced by social interactions and, once they are formed, the agent can act autonomously without relying on caretakers.

Now, IMAGINE also uses a cognitive function of language. Once language has been grounded as described above, IMAGINE leverages the *productivity* of language to generate creative goals falling outside of the domain of effects the agent already experienced. Language—and its compositional properties—is here used as a cognitive tool to facilitate the composition and imagination of novel goals. The mechanism is crudely inspired by usage-based linguistic theories Tomasello & Olguin (1993); Tomasello (2000); Goldberg (2003). It detects recurring linguistic patterns, labels words used in similar patterns as *equivalent* and uses language productively by switching equivalent words in the discovered templates. This simple mechanism generates truly creative goals that are both novel and appropriate, two adjectives used to define *creativity*—see discussion in Runco & Jaeger (2012). As the authors show, this simple mechanism powers the creative exploration of the environment and enhances the systematic generalization abilities of the agent. Whereas the communicative function of language mostly help internalize the goal representations of social partners, the cognitive function gives the agents its individuality and open-endedness: the agent can generate novel creative goals that its caretakers did not know about.

Using descriptions rather than instructions in IMAGINE is a deliberate choice given that linguistic guidance through descriptions is a key component of how parents teach language to infants Yoshida & Smith (2003); Tomasello (2009). This is in contrast with the instruction-based approach which is dominant in *language-conditioned* agents research, but rarely seen in real parent-child interactions Bornstein et al. (1992). Thus the IMAGINE agent proposes a model of language acquisition which contrasts with what is done in *instruction-following agents* Hermann et al. (2017); Chan et al. (2019); Bahdanau et al. (2019); Cideron et al. (2019); Jiang et al. (2019), where the social partner provides a linguistic instruction and then a reward upon completion, see Luketina et al. (2019) for a review. Finally, the technical cornerstone of this work is that it learns a goal-conditioned reward function where the goal is learned as a language expression from data coming from the social partner. This removes the need for a lot of feedback and presence for the social partner, which is one of the target functionalities of autonomous teachable agents. More details about this work are presented in Colas et al. (2020a).

The DECSTR agent

In the intrinsically motivated agents outlined in Section 2.2, just as in most goal-conditioned RL algorithms Schaul et al. (2015); Andrychowicz et al. (2017); Colas et al. (2020b), the space of goals is generally defined as a subset of the state space of the agents Nair et al. (2018); Florensa et al. (2018); Pong et al. (2020); Nair et al. (2019). However, this approach falls short of providing the level of abstraction necessary for natural communications with social partners. The IMAGINE agent addresses this concern by directly representing goals in a language embedding space. But, doing so, it cannot account for the fact that infants learn to target sensorimotor goals before they master language.

The DECSTR agent solves the latter issue by introducing an abstract goal representation layer in the architecture where a goal is expressed as general predicates. At the sensorimotor level, this helps targeting more abstract goals, opportunistically making profit of the current situation to address the easiest concrete configurations fulfilling the goals. The DECSTR agent interacts with blocks. So, for instance, having the red block “close to” the blue block can be realized in an infinity of ways and the agent can find the simplest movement to move one of the block so that it comes close to the other depending on the whole scene. At the language level, the abstract goal representation layer also plays a key role in grounding language-based descriptions into sensorimotor experience, as it simplifies the correspondence between natural language instructions such as “*put the red block close to the blue block*” and the goals the agent manipulates in practice.

More precisely, to ground language into its sensorimotor experience, the DECSTR agent relies on a *Language Goal Generator* that takes a language expression as input and samples concrete goals matching the language expression. The way this component endows sensorimotor autotelic agents with language sensitivity by relating language to behavior is depicted in Figure 2. This is precisely through this Language Goal Generator that language is grounded into sensorimotor goals.

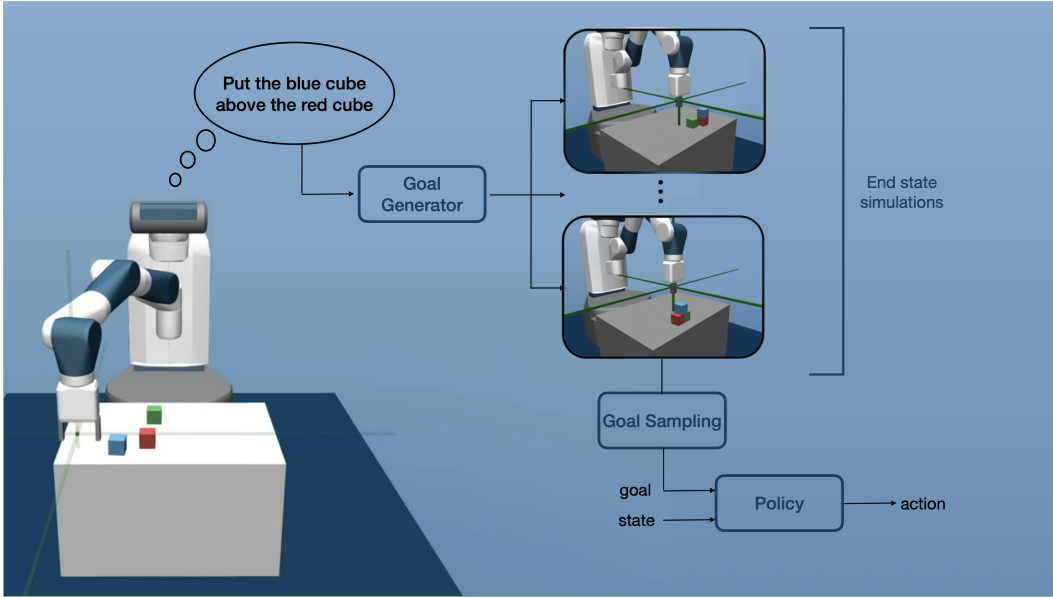


Figure 2: The DECSTR architecture as a Vygotskian Deep RL system. DECSTR learns to ground linguistic descriptions of its trajectories provided by a synthetic caretaker into innate semantic representations. This language grounding occurs in a *language-conditioned goal generator*. Once trained, DECSTR uses language as a cognitive tool to guide the simulation of possible future world configurations matching an input description (e.g. from the tutor or self-generated). As it selects one of them as goal, DECSTR commits to turning the world into this selected future.

The architecture of the DECSTR agent is depicted in Figure 3. It is designed to learn to manipulate blocks in the *Fetch Manipulate Tutoring* environment, a benchmark which was used in several recent works to train hierarchical RL agents Pierrot et al. (2020) and autotelic agents Lanier et al. (2019); Colas et al. (2019); Li et al. (2019).

The DECSTR agent is tested in a three-step process. First, it learns to discover and master all reachable block configurations expressed through a set of binary spatial predicates telling whether blocks are close to or away from each other, and whether one is above the other or not. In a second phase, similarly to the case of the IMAGINE agent, a social partner provides a simplified description of what the DECSTR agent did. It then learns to ground these descriptions into its abstract representations of compatible goal configurations. More precisely, the Language Goal Generator component is trained to generate a diversity of compatible goals for each possible instruction. By doing so, it instantiates the embodied simulation hypothesis by letting the agent project itself towards potential future states and choosing one of them as goal Barsalou (2009), and it accounts for the fact described in Wood et al. (1976) that children can understand a goal before being able to reach it. The capability of generating a diversity of goals also results in an increased diversity of behaviors displayed by the agent and a capability to retry to pursue the same goal in another way Akakzia et al. (2021). In a third phase, the agent is finally instructed to move some block using its learned set of skills, and can do so in a variety of ways.

In the first sensorimotor phase, the DECSTR agent successfully discovers and masters all reachable configurations in its goal space. It first masters the easy goals, such as putting blocks close or away to each other. Putting a block above another is more difficult and takes more time to learn, as it requires the mastery of grasping and releasing the block at the right place. After a while, the DECSTR agent finally learns to build pyramids and towers of three blocks. Once becoming more expert, the DECSTR agent gets opportunistic, making profit of the current configuration to reach its goal with as few block moves as possible. More details about this work are presented in Akakzia et al. (2021).

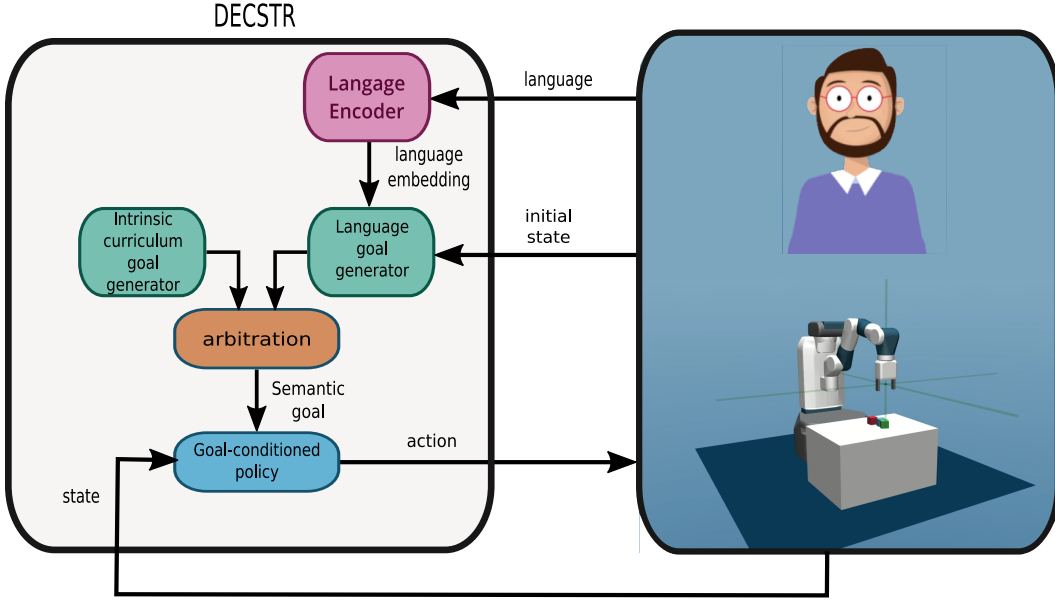


Figure 3: *The DECSTR architecture in the Fetch Manipulate Tutoring environment.* The tutor can interact with the agent by setting the initial scene, providing descriptions or instructions. DECSTR uses aligned descriptions and trajectories to train a language-conditioned goal generator mapping language to a set of matching configurations used as goals. The agent either pursues language-conditioned goals or its own. In either case, it learns to achieve them through intrinsically motivated goal-conditioned reinforcement learning.

3.2 LIMITATIONS OF LANGUAGE-AUGMENTED AUTOTELIC AGENTS

In both IMAGINE and DECSTR studies, the social partner facilitates the sensorimotor self-training process of the agents by setting up the environment in configurations where the goals can be more easily reached so as to maintain them in their *zone of proximal development* Vygotsky (1978). This process has an influence on how fast the agents learn to reach a goal or another. For instance, the discovery of pyramids or stacks of three blocks is much slower in DECSTR when the social partner is removed Akakzia et al. (2021). Thus IMAGINE and DECSTR can already be considered ‘teachable’ in a very elementary way. Besides, both agents are endowed with a basic capability to interpret language. But they differ a lot in terms of their capability to be taught through language.

In the case of the DECSTR agent, since language acquisition and instructions are decoupled from sensorimotor learning, language obviously plays no role in the sensorimotor learning process. Instructions are just used *after* sensorimotor learning to evaluate the responsiveness of the agent to language. In the case of the IMAGINE agent, by contrast, the description of actions plays a direct role in orienting the exploration process of the agent. The IMAGINE agent can thus be said ‘teachable’ through language in a stronger way than DECSTR. However, in all experiments performed in Colas et al. (2020a), the agent is only targeting its own goals and the role of the social partner is limited to describing what the agent does, without providing any direct incentive to target a particular goal rather than another. This is to be contrasted with the studies of Wood et al. (1976) where the tutor is in charge of making sure that children will build a pyramid. Thus the IMAGINE agent has the potential to receive instructions, but this potential has not been demonstrated.

Admittedly, some limitations of these language-augmented autonomous agents could be overcome by combining the language grounding and instruction following capabilities of the DECSTR agent with the capability of the IMAGINE agent to imagine new goals and couple sensorimotor learning to language acquisition. Somehow, achieving this could result in displaying *overlapping waves* of sensorimotor and linguistic development Siegler (1998).

However, such a combination would raise additional difficulties and would still suffer from important limitations, which we investigate in the next section.

4 TOWARDS MORE TEACHABLE AUTOTELIC AGENTS

The skill acquisition process of the DECSTR and IMAGINE agents can be influenced in two ways. First, the caretaker can interact with the agents through language. Second, in DECSTR studies, the caretaker can set the scene at the beginning of each training episode by moving blocks so that some goals get easier to reach than others. From this broad perspective, we recognize the two main ways by which the tutor interacts with children in Wood et al. (1976). However, when looking more closely, there are several discrepancies:

- As we just outlined, the role of verbal communication from the tutor in Wood et al. (1976) is to maintain the motivation of the child towards building the pyramid, whereas neither in DECSTR nor in IMAGINE are the tutor’s utterances used to orient the sensorimotor learning process of these agents towards a specific goal.
- The DECSTR and IMAGINE agents start a new episode with an environment that has already been prepared by the caretaker whereas children of Wood et al. (1976) can observe the tutor intervening and may benefit from these observations.
- In IMAGINE and DECSTR, language is used to describe goals or to ask the agent to reach a goal. By contrast, in Wood et al. (1976), language interactions consist of question-based corrective feedback (“isn’t this pair different from that one?”) or direction maintenance (“do more of this”).

All these differences point to limitations in the capabilities of the DECSTR and IMAGINE agents. In the remainder of this section, driven by these differences, we investigate the properties that are missing to these agents so that they can be taught in a more natural way. For doing so, we start by adopting a general perspective about tutoring as a mutual exchange between the tutor and the learner using two communication channels.

4.1 TUTORING AS COMMUNICATION

Interactions between a tutor and a learner, and more generally social learning processes Bandura & McClelland (1977), can be conceptualized as a mutual exchange process using two main communication channels, the social channel and the task channel, see Figure 4.

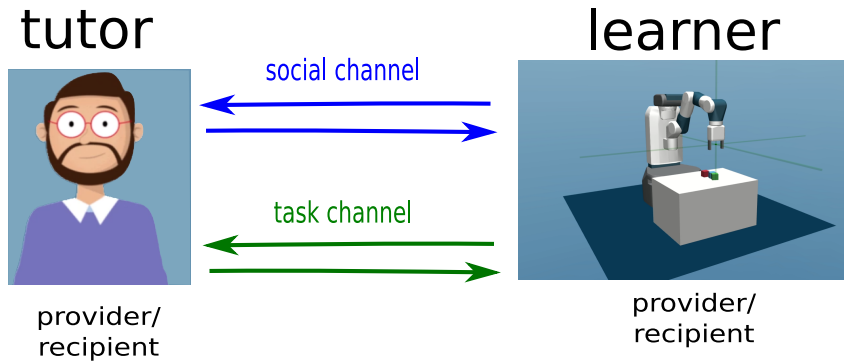


Figure 4: The general mutual exchange perspective using two communication channels. The social channel is used for exchanging social signals such as feedback, requests, commitment signals, etc. The task channel is used for exchanging task-related signals such as demonstrations. Both participants can use both channels either as signal provider or recipient.

From one side, the tutor may use the social channel when providing instructions or non-verbal feedback, and the task channel when providing demonstrations, either on purpose or even without the intention to teach. From the other side, the learner may use the social

channel when providing feedback on her understanding or asking questions, and the task channel when imitating the tutor or performing the task to display her current capabilities. As Figure 4 shows, the tutor and learner can either be the provider or the recipient of the messages depending on the side of the communication. Also, both participants usually alternate these roles, resulting in a mutual exchange. For example, when the learner executes a task in order to obtain a feedback from the tutor, she is first the provider and then the recipient.

From this general perspective, we can now reconsider various forms of tutoring. For instance, when the provider is the tutor, she may use the social channel to provide instructions or feedback, and the task channel to provide demonstrations. In that case, her behaviour is modified with the intention to teach the recipient. She may also provide descriptions of the activity of the agent, as we have seen in Section 3. Reciprocally, the learner can become the provider using both the social and the task channels to inform the tutor about her current understanding and capabilities. The learner can also be the provider when querying more information, implementing a form of more active learning. We argue that, to be successful, tutoring should exploit and combine the social and the task channels, resulting in various “frames” of exchanges where both participants can use both channels and both roles. These frames of exchanges are called “*pragmatic frames*” in Vollmer et al. (2016). Note that Wood et al. (1976) do not study the communication signals from the child to the tutor. In that respect, Wood et al. (1976) do not benefit from the mutual exchange perspective that we adopt here and whose importance has been outlined in several later research works Thomaz & Breazeal (2008a); Vollmer et al. (2016).

So far, the above paragraphs and Figure 4 provide a somewhat static view of the possible exchanges between both participants of the tutoring process. Another dimension illustrated in Figure 5 is more related to the way the mutual exchange process unfolds through time. Generally speaking, when engaged in a complex task, both the tutor and the learner should sustain the interaction using a sequence of the above exchanges Oertel et al. (2020). On these temporal aspects, two dimensions must be distinguished. First, both participants have goals, and the mutual exchange process is about the dynamics of these goals: how they are maintained and how they evolve, how they can become aligned or misaligned between both participants. The second dimension is about the dynamics of elementary exchanges and the dynamics of attention of both participants: the way these exchanges and the attention of participants unfold into patterns whose purpose is to successfully achieve their global goals. As task-related behaviors, the successful interaction patterns can be learned from experience and extracted into a set of efficient protocols that can then be reused to learn various contents through similar interactions. These protocols are exactly the pragmatic frames of Vollmer et al. (2016) and the challenge of automatically extracting such frames from tutoring interactions is still unaddressed.

In the rest of this section, we build on this perspective to identify the key research questions leading to the design of teachable autotelic agents. We first investigate interactions through the task channel (Section 4.2): recognizing pedagogical signals, learning from demonstration, observational learning and inference from indirect goal-related signals. Then, we cover interactions through the social channel (Section 4.3): learning from feedback, instructions, joint attention and engagement. Besides, maintaining this communication also requires the use of mental models from both sides. From one side, the tutor needs models of the goal and the learner’s current understanding of this goal, as put forward in Wood et al. (1976). But from the other side, the learner also needs a model of the tutor’s models to provide adequate feedback signals. These models of the partner are called a Theory of Mind Wellman (1992) and again we outline the specific research challenges raised by the autotelic agent perspective in Section 4.4. While covering these topics, we review pre-existing AI research trying to account for some of the corresponding properties, even when these works do not focus on autotelic agents. A few of the higher-level task-related cognitive capabilities that autotelic agents should display to better account for learning in children are then covered in Section 5.

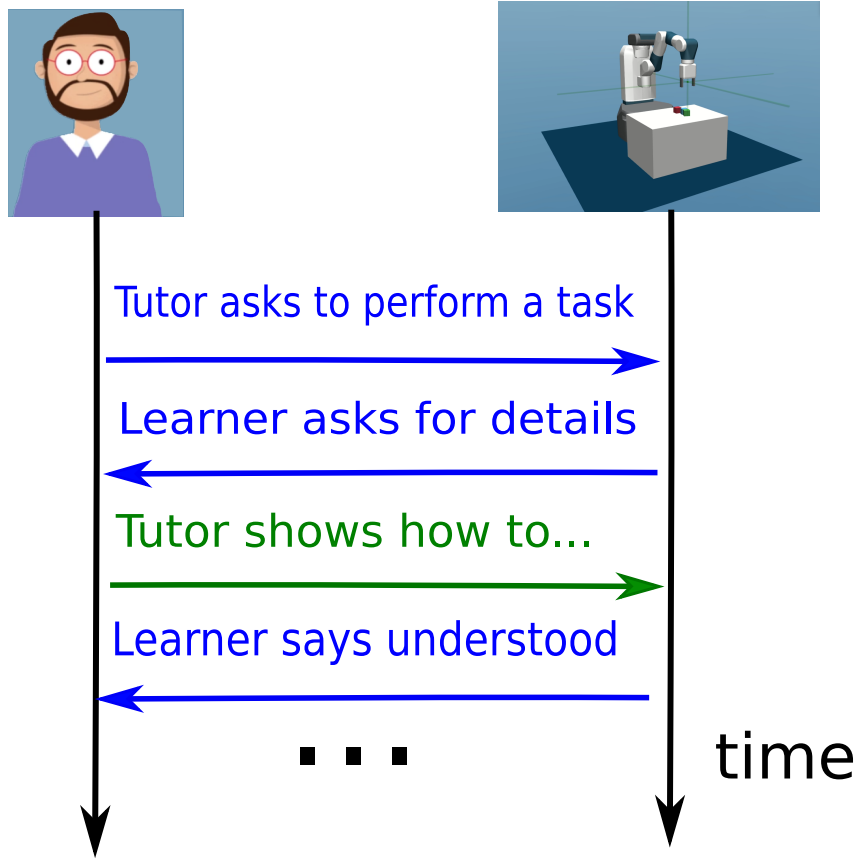


Figure 5: Sequence diagram of a pattern of exchanges. Blue: social channel; green: task channel. From the mutual exchange perspective, the goal of these exchanges is to maintain the dynamics of interaction so that both participants successfully reach their global goals. Several such patterns can be abstracted away into pragmatic frames.

4.2 LEARNING FROM TASK CHANNEL SIGNALS

As depicted in Figure 6, truly teachable autonomous agent that learn from observing the behavior of the tutor should be able to recognize whether the tutor is adopting a pedagogical stance or performing the task on his own, and may even infer information about the task from indirect task channel signals. We cover the corresponding issues in the next paragraphs.

Recognizing Pedagogical Signals

From the framework sketched in Section 4.1, the pedagogical stance of a tutor should be recognized from the general context of the initiated interaction and from the fact that the action itself conveys communicative signals. As Csibra & Gergely (2009) put forward, in *natural pedagogy*, the tutor shows the action with additional signals helping the learner determine that a demonstration is given and identify what matters in this demonstration. Through the task channel, the gestures are “speaking a language” called *motionese* Brand et al. (2002); Nagai & Rohlfing (2007) and the agent has to sort out the communicative part from the operative part of the action to understand the conveyed message. Despite interesting attempts Ho et al. (2017), the capability to recognize these pedagogical signals is overlooked both in Interactive RL agents and in autotelic agents, and the fact that the latter set their own goal does not seem at first glance to make a difference in this respect. This is more in the way these agents will reproduce the observed or demonstrated behaviors that a difference appears.

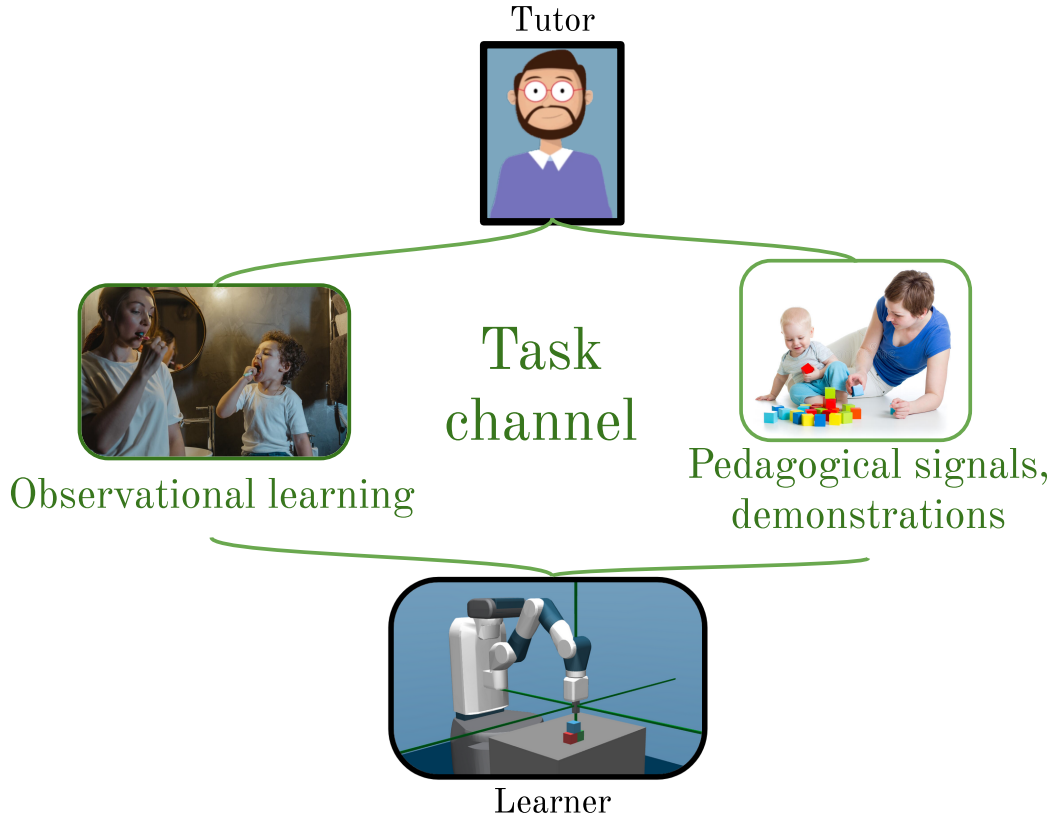


Figure 6: Examples of task channel signals exchanged between the tutor and the learner. Here, these signals consist of the task-related behaviour of the tutor, either as demonstrations or as performance without a pedagogical intention.

Learning from Intentional Demonstrations

A striking capability of children is that, when the tutor adopts a pedagogical stance and performs a unique demonstration, they are capable of generalizing from this unique demonstration to other contexts where what they have been shown is relevant Csibra & Gergely (2009). This outstanding generalization capability is to be contrasted with what children show when simply learning from observing someone else performing the task without a clear pedagogical intent. The authors describe this sensitivity to pedagogy as an innate bias in humans. Clearly, such a bias is missing in current teachable agents.

The key question, then, is to understand the nature of the information conveyed by a demonstration towards children. Since gestures are performed differently between the intentional and the non-intentional demonstration cases, the message cannot be the trajectory itself to be blindly reproduced Wood et al. (1976), but must be inferred from this trajectory. Furthermore, demonstrations are often adapted with respect to the child’s previous mistakes, as if the tutor was trying to convey an *explanation* about how to perform it Csibra & Gergely (2009). A clear understanding of the nature of this explanation or more generally of the conveyed message in a demonstration is currently missing Ho et al. (2016). But, following Wood et al. (1976), understanding the goal of the demonstration is crucial in this process, thus having goals as autotelic agents do is crucial to tackle this question.

Observational Learning

In natural interactions, the child can observe the tutor acting in the environment even when no pedagogical intention is present, and still extracts a lot of information from these obser-

variations, a process referred to as *observational learning* Varni et al. (1979); Meltzoff (1999); Burke et al. (2010). In particular, learning can occur when the child tries to reproduce the observed movements.

To design an agent learning from the reproduction of observed movements, we have to account for four aspects: its attention towards the observed movement, its capability to retain it, its ability to reproduce it and its motivation for doing so Bandura & McClelland (1977).

In the Interactive RL literature, there are attempts to show that combining RL with LfD is enough to account for observational learning, but these attempts are still limited, as they do not consider that the observing agent may have its own goals Borsa et al. (2019).

Closer to our concerns, the autotelic CLIC agent imitates the behavior of other agents acting in the environment without a pedagogical stance Fournier et al. (2019). An interesting feature of CLIC is that it relies on a curriculum learning mechanism to decide *which goal to imitate* from these agents depending on its current capabilities. However, the CLIC approach to imitation still suffers from the same limits of the LfD and BC approaches as standard interactive reinforcement learners that we already outlined in Section 2.1.

Rather than retaining the whole trajectory and adapting it to its own capabilities, the learner may “imitate the goal” of the observed agent. This process called “goal emulation” Tomasello (1998) has already been put forward in interactive agent design Nguyen & Oudeyer (2012). This perspective is backed-up with neuroscience results, where researchers studying how humans learn from the actions of others have found a dedicated and highly sophisticated cognitive system based on *mirror neurons* Rizzolatti et al. (1996). These neurons are believed to help trigger a mental motor simulation of observed actions as if the observer was performing the action. However, though mirror neurons have been attributed many roles in social cognition Gallese et al. (2004), there has been no direct evidence so far of the direct involvement of mirror neurons in imitation Triesch et al. (2007). Thus, it might be the case that these neurons are more involved in action understanding in the context of a Theory of Mind (ToM) of the other agent (see Section 4.4) than responsible for the activation of an imitated sequence of actions Bonaiuto et al. (2007); Bonaiuto & Arbib (2010), which would support the goal emulation perspective. Thus, to endow autotelic agents with observational learning capabilities, it might be necessary to provide them a mental motor simulation system and ToM-related capabilities.

Inference from Goal-Related Signals

When setting up the environment or handing a block to a child, the tutor’s intent is certainly not limited to facilitating the child’s job in reaching a goal. Most probably, the tutor also communicates information about the goal itself, even when the way to reach this goal is not displayed. In that case, the learner has to infer information about the goal and the way to reach it, a capability that is crucial to the social nature of humans Tomasello et al. (2005).

Indeed, children tend to consider the objects the tutor attends to as relevant to a potential goal. More generally, they can infer the intent of the tutor and consider her intervention as an implicit specification of the goal they are expected to pursue. Such implicit specification can be further completed or reinforced by additional social signals such as gaze orientation, posture, gesture exaggeration or even verbal signals.

Driven by the Interactive RL framework where agents learn from an external reward function, some works study how an agent can infer information about a reward function from observed premises in the tutoring context Reddy et al. (2020); Bobu et al. (2020); Jeon et al. (2020).

Moving to autotelic agents, similar processes could be transposed to infer a goal rather than a reward function. A few recent works start addressing this issue by captioning the goal of a demonstration through natural language and a goal generator Zhou & Small (2020); Nguyen et al. (2021). We expect follow-up of such works to contribute to answering the key question of the nature of the information conveyed by the tutor through the task channel MacGlashan & Littman (2015); Ho et al. (2016; 2017).

4.3 LEARNING FROM SOCIAL CHANNEL SIGNALS

As shown in Figure 7, feedback signals and instructions are exchanged through the social channel. Together with demonstrations, these signals are the most covered teaching mechanisms in the Interactive RL literature where agents do not have explicit goals but feedback and instructions are just about elementary actions. Here, we investigate the differences it makes to consider autotelic agents to deal with these signals.

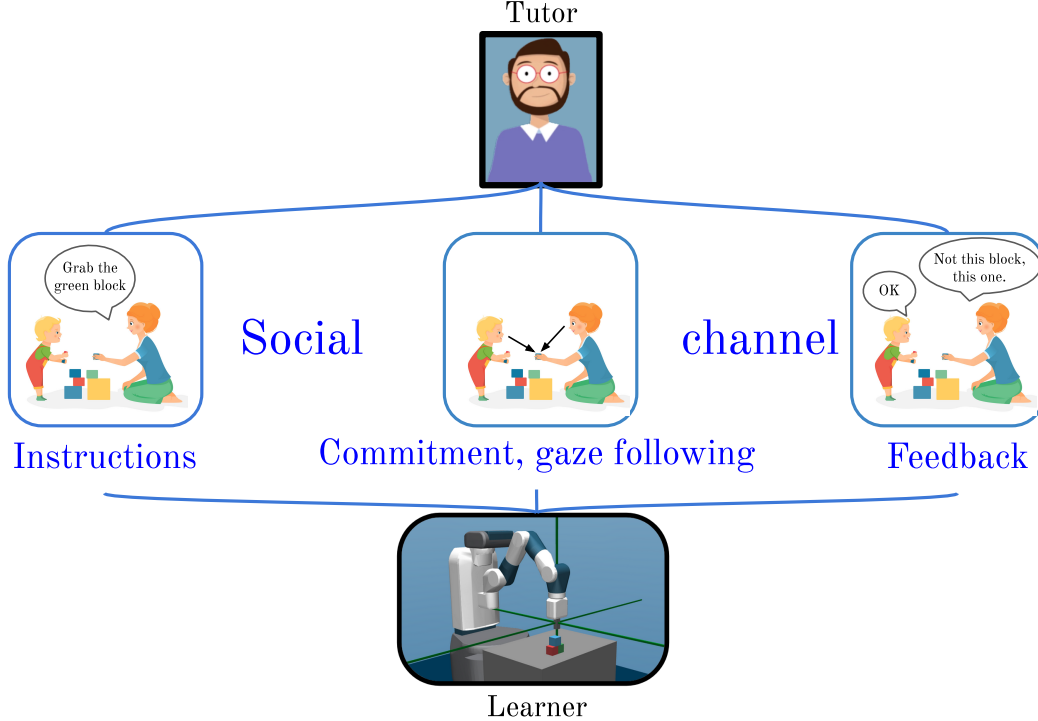


Figure 7: Examples of social channel signals exchanged between the tutor and the learner. The signals consists of verbal feedback such as instructions or feedback from the learner or the tutor. This also includes non-verbal interaction signals used to maintain the learner engaged into the learning activity, through gaze following for instance.

Learning from Tutor Feedback

Though the protocol of their study does not explicitly build on such aspects, Wood et al. (1976) mention maintaining a positive atmosphere as a prerequisite for tutoring to be effective. Besides, the verbal feedback asking children to compare their constructions to positive examples can be seen as a form of negative feedback whereas asking them to do “more of the same” can be seen as positive feedback.

Reinforcement learning agents are explicitly designed to learn from a feedback signal which can be either a reward or a punishment. Using feedback in Interactive RL is thus natural, and papers doing so abound, see e.g. Lin et al. (2020) for a short review. But in isolated RL agents, the signal is just a scalar value whose semantics (whether it is a reward or a punishment) is immediately given by its sign. In the Interactive RL context, the feedback signal can be more arbitrary, but the semantics is generally still considered given Cederborg & Oudeyer (2014); Najar et al. (2020). Exceptions include Grizou et al. (2013), and Najar et al. (2016), where an agent learns to interpret feedback signals based on their correspondence to rewards received from the environment. However, autotelic agents do not receive any reward from the environment. So how can they interpret arbitrary feedback? It might be the case that humans and other animals are biased to interpreting some signals as re-

ward or punishment without having to learn it. But without doubts, we can learn how to interpret less immediate feedback signals such as the verbal ones in the work of Wood et al. (1976). Do we do this by leveraging correlations to signals whose semantics is more innate? How to account for these learning processes and transfer them into autotelic agents is still an open question.

Besides, in Interactive RL, feedback is generally about the just performed action or sequence of actions in the context of a single task. By contrast, according to Wood et al. (1976), tutoring interactions are mostly dedicated to maintaining children on the right goal. For instance, feedback can be used to re-engage or maintain the commitment of the child towards the tutor’s goals. Thus, in contrast with Interactive RL agents, research on autotelic agents could consider feedback about goal selection itself. We are not aware of such research, though many phenomena could be studied, such as the fact that a disengaged child will tend to ignore feedback about goals it is not attending to.

Finally, to benefit from feedback during sensorimotor learning, the architecture of a teachable autonomous agent should also contain an arbitration mechanism to combine external and intrinsic reward signals. Our own evaluation about the way we achieved a goal and the evaluation of a tutor may differ; if one is positive and the other negative, how this feedback combination should influence behavior is unclear, resulting in interesting research questions.

Learning from Instructions

In the standard Interactive RL literature, instructions generally tell the agent which elementary action or eventually which sequence of such actions to take in the current context to achieve a given task. By contrast, when considering autotelic agents who set their own goals among many, instructions and feedback can be not only about the choice of immediate actions, but also about the choice of more distant goals. Existing autotelic agents already possess the necessary mechanisms to learn from instructions. For instance, the DECSTR agent can translate an instruction into a set of potential goals. By learning on its own how to reach these goals, it learns how to follow instructions. As we pointed in Section 3.1, the IMAGINE agent rather learns from descriptions but can then follow instructions characterizing behaviors close to the described ones.

As is the case with feedback, an agent learning from both intrinsic motivations and instructions would need an arbitration mechanism to choose between sticking to its own goal or attending the ones of the tutor.

To conclude this part, in principle, autotelic agents should be able to deal with richer forms of tutoring signals such as feedback and instructions. But they should also be endowed with mechanisms to arbitrate between, from one side, their own goals and intrinsic learning signals and, from the other side, those coming from the tutor. These arbitration mechanisms should, in turn, endow these agents with a form of ‘personality’ where different agents with different parameters would be more or less easily taught, exactly as in the normal teaching of children.

Maintaining Commitment Through Gaze Following

The ability to engage in reciprocal exchanges is required for efficient interactions. Interestingly, a lack of engagement or at least difficulties to engage with partners have been identified as important predictors of several developmental disorders Dawson et al. (2004); Ouss et al. (2014).

As outlined in Section 4.1, both partners in an interaction use social signals to start and maintain engagement or to disengage. Among these forms of communication, joint attention through gaze following and feedback from both sides play a crucial role. We now investigate these engagement processes.

Let us first consider the ubiquitous role of gaze following. From one side, gaze following is a required component of engagement, as it is necessary for the learner to attend the goal the tutor is referring to. But achieving a task and observing the tutor performing a demonstration also mobilize gaze. In that respect, an arbitration mechanism is necessary

to decide where to attribute gaze resources. From another side, gaze following can support learning under guidance of a tutor without the need for any additional external reward. For instance, in the work of Fournier et al. (2017), an agent learns to achieve the task intended by a tutor by leveraging a simple gaze following mechanism: it tends to engage in actions towards objects it is looking at, and tends to look at the same objects as the tutor. Thus, by watching the right objects at the right time, the tutor can drive the agent towards the intended behavior. In a sense, this work is a precursor of using engagement in autotelic agents, but the agent was lacking a goal representation capability.

Thus, again, we are not aware of any autotelic agent using gaze following to learn the right behaviors under the guidance of a tutor, though these joint attention processes might be at the heart of the mutual exchange processes described in Section 4.1.

Maintaining Commitment by Emitting Feedback

Emitting feedback is also essential from the learner to maintain commitment. It helps the tutor acquiring a model of the learner’s current understanding of the task. For instance, adding a transparency component in a learning agent behavior helps the tutor providing more effective reward signals, resulting in improved training efficiency Thomaz & Breazeal (2008b). On the same line, by providing feedback to improve transparency, a robot learner influences the tutor’s movement demonstrations in the process of action learning and improves the overall training efficiency Vollmer et al. (2014).

Adding transparency mechanisms to autotelic agents would be quite straightforward and would significantly contribute to making them more easily teachable. But to decide what to emit, one should endow these agents with a capability to learn a model of the mind of their tutor, which we study in the next section.

4.4 THEORY OF MIND

Reading Wood et al. (1976) makes it clear that the tutor uses a model of the child’s understanding to manage the tutoring interaction. Reciprocally, a teachable autonomous agent should be capable of understanding a goal from the behavior of the tutor, which relies mostly on inferring the tutor’s expectation and then reasoning to figure out how to meet this expectation. Thus, this process should clearly call upon a mental model of the tutor’s expectation. Such a mental model is part of a larger model called a *Theory of Mind* (ToM), as illustrated in Figure 8 and put forward in recent developmental psychology papers Vélez & Gweon (2021); Gweon (2021). Having a theory of mind means being capable of reasoning about other people’s mental states Jara-Ettinger (2019). The most common mental states these theories refer to are beliefs, desires and intentions.

There has been recent attempts to account for the acquisition of a ToM through inverse RL Jara-Ettinger (2019) and in the domain of multi-agent RL Nguyen et al. (2020a), but these works generally suffer from the same limitations as Interactive RL approaches: they do not consider explicit goals. In robotics, more convincing studies about having a ToM are prominent in the domain of human-robot collaboration, as having a model of the teammate seems mandatory for collaboration to be efficient Chakraborti et al. (2017). In particular, these works can be organized along increasing capabilities from the agents as *behavior-based*, *goal-based*, *proactive* and *social agents*, the latter being the only ones to have a full ToM. Thus, one can see that all these types of agents but the first are autotelic.

However, none of the above works consider the specific case of a tutoring interaction. The most interesting exception we are aware of is Rabinowitz et al. (2018), where the authors design a machine learning algorithm endowing an agent with the capability to learn a model of other agents. The algorithm belongs to a sub-field of RL called meta-reinforcement learning where the goal is to “learn to learn”, that is to let the algorithm design its own learning dynamics by using gradients over learning capabilities. Their specific instance learns to distinguish between agents displaying different behaviors, to relate sub-optimal behaviors to limitations in the perception of the agents, or even to interpret sub-optimal behaviors in terms of false beliefs from these agents. Though the authors do not explicitly mention this, the last point is of utmost interest when one considers an artificial tutor-

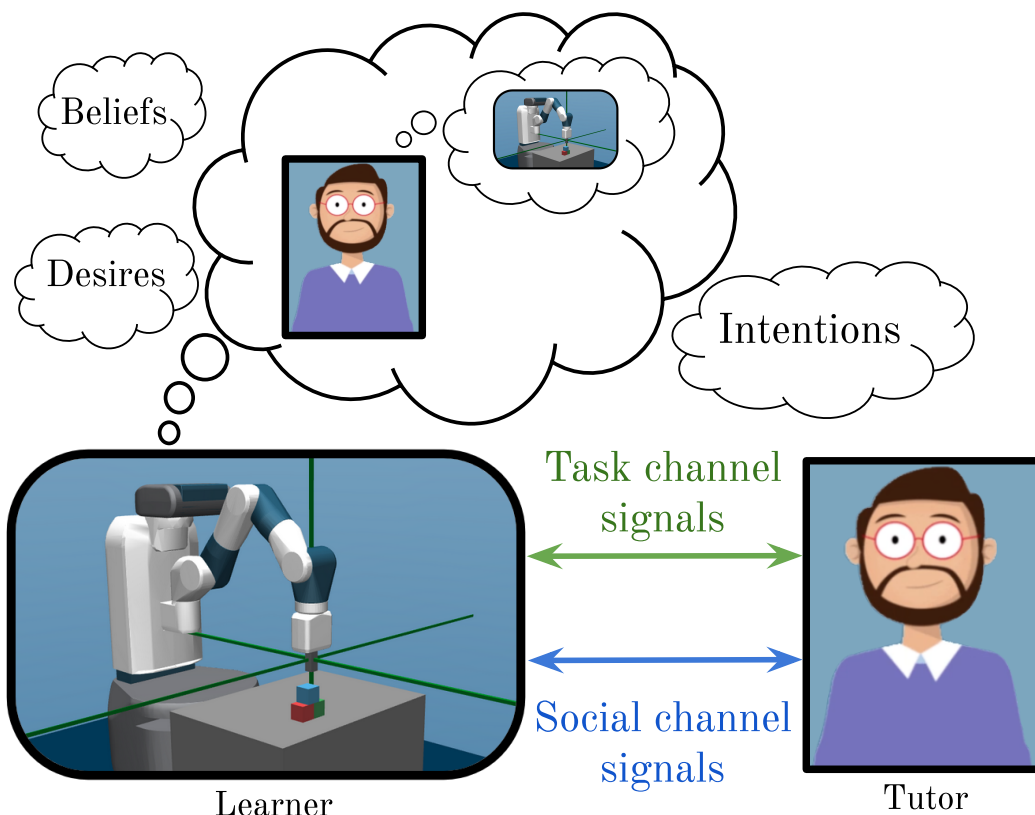


Figure 8: In our context, the Theory of Mind states that the learner builds and maintains a model of the tutor’s beliefs about what the agent knows and is capable of doing, its intentions and desires. The tutor also has a model of the learner, and maintains it to provide adapted teaching feedback along the learning phase.

learner interaction. From the side of the tutor, identifying a false belief raises the question of finding the most adequate way to help the learner correct this belief. But, reciprocally, the learner may also learn a theory of the beliefs of the tutor about itself, and find the most adequate communicative signals to help the tutor correct these false beliefs or more generally maintain a better model of the current knowledge of the agent.

To conclude this part, endowing artificial agents with the capability to learn a ToM in a tutor-learner interaction raises a lot of interesting research questions which are widely open, and which constitute avenues for further investigations towards teachable and autonomous agents.

5 TOWARDS MORE AUTONOMOUS TEACHABLE AGENTS

Though we want to keep this paper focused on the interactive properties that an agent must have so that it can be taught like a child, we have to admit that to account for a lot of teaching phenomena such as the ones described by Wood et al. (1976), these agents must also be endowed with task-related cognitive capabilities that are still missing. These capabilities are not specific to teachability, but progress on modelling them will help design better models of teachable agents. In this part we quickly list some of these missing capabilities.

5.1 MORE NATURAL LANGUAGE ACQUISITION AND LANGUAGE GROUNDING

The study of Wood et al. (1976) being with children older than 3, it does not consider the language acquisition and language grounding processes. In the language-augmented agents described in Section 3.1, a simplified template-based language is learned using standard machine learning tools without any effort to mimic the language acquisition processes of infants. However, the developmental processes and constraints involved in language acquisition may play a crucial role in the more general acquisition of interaction capabilities. Thus, more realistic models of language acquisition and learning of semantic predicates in infants Goldstein et al. (2003); Kuhl (2004); Mandler (2012) may be required to better account for these capabilities. Besides, communication itself could be extended to richer natural language interactions Lynch & Sermanet (2020); Brown et al. (2020) so as to increase the teachability of autotelic agents.

5.2 MENTAL COMPARISON OF COMPLEX CONFIGURATIONS

In the work of Wood et al. (1976), some language interactions consists of question-based corrective feedback under the form “isn’t this pair different from that one?”. For children to use such a feedback, they need to be able to mentally compare their construction with another, which requires the capability to build mental models of the two constructions and perform a mental comparison. The use of such mental models are at the heart of cognitive sciences Johnson-Laird (1983). However, despite fantastic progress on representation learning capabilities, machine learning research is still far for endowing learning agents with the capabilities of using mental models displayed by humans Marcus (2018); Lake et al. (2017). In particular, being capable of relating several models through abstract features, which is a key capability in humans Gentner (1983); Gentner & Hoyos (2017); Gentner (2016), is only an emerging question in machine learning Hill et al. (2019).

5.3 HIERARCHY OF GOALS AND WORKING MEMORY

Children of Wood et al. (1976) are addressing a global goal configuration – the wooden pyramid – but at any moment they can be addressing an intermediate goal such as taking two blocks apart. This suggests that children can build and maintain in mind a hierarchy of goals, a sophisticated cognitive process which has been related to executive control in the Prefrontal Cortex (PFC) Koechlin & Summerfield (2007) and may account for working memory properties of human subjects Baddeley (2000). In particular, *private speech* may play a key role in maintaining goals and plans in working memory, as the phonological loop may benefit from better short term memory properties Baddeley (1992; 2003).

In DECSTR, there is a unique goal-conditioned policy which targets a unique global configuration at a time. The way this global configuration unfolds into a sequence of elementary actions is directly encoded into the weights of the neural policy and hard to reverse engineer. Having an agent capable of maintaining a hierarchy of goals and to reach them sequentially based on a hierarchy of policies would dramatically improve the learning capabilities and the interpretability of the knowledge acquired by the agent, as is the case for instance of the AlphaNPI-X agent Pierrot et al. (2020).

Besides, we have outlined in Sections 4.4 and 5.2 the need for endowing teachable autonomous agents with specific model building capabilities such as building mental models of observed objects and social partners. A teachable agent would thus certainly benefit from the integration of more general model building capabilities Gentner & Stevens (2014); Lake et al. (2017). Finally, combining hierarchical skill learning and hierarchical model building is certainly the way to go to endow autonomous agents with more natural few-shot generalization capabilities Eppe et al. (2020).

5.4 CAUSALITY AND COUNTERFACTUAL THINKING

A key aspect of human intelligence lies in the ability of understanding the causality of the events that happen in the world, and also mentally generate alternatives in order to reason about the right actions to choose, i.e. counterfactual thinking (“what would have happened

if I had chosen to move this block rather than this one?”). Reasoning about the world by considering causal events and imagining alternatives that could have happened is a powerful tool that could help agents to actively understand their experience with the world. It should help them to more efficiently evaluate the feasibility of teaching instructions and, to better understand feedback and to imagine goals more easily using their understanding of the causal nature of the events that happen in their environments.

There is a current effort in developing machine learning algorithms directly incorporating those abilities. Fundamental research questions about causality and counterfactual thinking have been asked, notably by Pearl (2009); Pearl & Mackenzie (2018), and further research efforts on how to apply those concepts to agent-environment scenarios are being explored Forney et al. (2017); Zhu et al. (2020); Lee & Bareinboim (2020).

This naturally leads to the questions of interpretability and explainability, also investigated by Pearl (2009); Pearl & Mackenzie (2018). Causality and counterfactual thinking are essentially tools that allow humans to explain how and why events occur. These cognitive abilities can efficiently help the agent to learn how to act in the world using limited experience, which relate to the few-shot learning ability of infants. This is already investigated in the case of RL, where research has attempted to provide explainability of RL agents using frameworks based on causality Madumal et al. (2020b;a).

5.5 LEARNING EFFICIENCY OF AUTOTELIC AGENTS

IMAGINE and DECSTR, and more generally language-conditioned and autotelic agents, employ RL as a learning mechanism for deciding which actions to perform given current and possibly past states. Thus, they rely on this learning paradigm that currently suffers from sample inefficiency Buckman et al. (2018).

As mentioned in Section 2.1, when richer teaching signals are present, learning speed can be increased, because agents can benefit from more information at each time step, compared to the standard RL scenario where they only get the reward signal. However this is not the only way to increase learning speed in the case of RL. In parallel, active research Xu et al. (2020); Schmitt et al. (2020); Schrittwieser et al. (2020); Dorner (2021) aims at tackling this issue without resorting to rich teaching signals, but rather by implementing novel algorithmic modules or considering various approaches such as off-policy learning strategies or model-based RL, which combines RL with self-supervised learning. These fundamental developments in the field of RL should improve the sample efficiency of teachable and autonomous agents towards the goal of few-shot learning.

5.6 PLANNING AND REASONING

Autonomous agents should be endowed with higher-level cognitive capabilities such as reasoning about their own goals Eppe et al. (2019) using language in the reasoning and planning processes Nguyen et al. (2020b) or leveraging natural communication with the tutor about these goals. Leveraging social interactions to acquire their own semantic representations is certainly a crucial feature for autonomous agents if they are to account for the planning and reasoning capabilities of children.

5.7 LOCAL CONCLUSION

Definitely, the short list of topics covered in this section does not exhaust the cognitive capabilities that teachable autonomous agents should display to account for tutoring interactions in children. But we consider that the state of the art in machine learning is close to reaching these capabilities, hence we may soon see such agents. In turn, the tutoring of these more capable autotelic agents will raise specific research questions such as defining learning curricula taking into account the hierarchical relations between goals of the agents, arbitrating between learning a mental model of the task or of the tutor, or leveraging reasoning capabilities to better communicate about learning difficulties.

6 DISCUSSION

Many efforts have been made to endow artificial agents with the capacity to learn from humans, in a natural and unconstrained manner. However, for now, we are still far from achieving “normal teaching of a child”, in reference to Turing’s view. In Section 4, we have given an overview of some of the points that need to be addressed to increase the capability of autonomous agents to be taught, and we have reviewed some of the existing work in the corresponding directions. Then, in Section 5, we have pointed out fundamental research questions that need to be addressed, or existing research that requires to be improved upon for obtaining more autonomous learning agents. Now we first describe the next steps that one can expect in the immediate future to push further the frontiers of research in the domain from the integration of the corresponding works. Then, in a second part, we review more fundamental questions that remain open and may lead to deeper scientific revolutions in the design of the next generation of teachable autonomous agents.

6.1 NEXT STEPS

In Section 3.2, we already outlined the additional properties that one may expect from the integration of the capabilities of the IMAGINE and DECSTR agents into a single more advanced agent. Such an agent would simultaneously learn how to act from its sensorimotor interactions with its environment and ground these sensorimotor capabilities into linguistic representations that would make it sensitive to the instructions of a tutor. Ultimately, this agent could also be enriched with a capability to communicate about its own goals, giving rise to some transparency in the tutoring interaction loop.

To benefit from the tutor’s instructions, the agent would also need to quickly guess their meaning, which is more difficult in an incremental learning context than in the batch learning approach used in DECSTR studies. For that, the enriched agent could leverage additional gaze following and joint attention mechanisms which have already been used in artificial agents Fournier et al. (2017).

Designing such richer teachable autotelic agents is just a matter of integrative effort. Similarly, we see no fundamental issue behind extending these richer agents with capabilities to learn and manipulate hierarchical skills as AlphaNPI-X does Pierrot et al. (2020).

Another direction in which some steps can easily be made is the integration of a basic theory of mind, leveraging the meta-reinforcement learning methods described in Rabinowitz et al. (2018) or the Bayesian inference processes proposed in Vélez & Gweon (2021).

We have also pointed to the lack of studies where feedback from the tutor is about the choice of a goal from the learner, rather than about achievement of the current goal. Designing such studies and augmenting teachable autonomous agents with the capability to make profit of such feedback does not seem to raise any critical difficulty.

Integrating all these features into a single teachable autonomous agent would significantly increase their flexibility, making it possible to leverage a more significant part of the vast repertoire of interaction protocols or “*pragmatic frames*” used in human tutoring Vollmer et al. (2016) and, more generally, opening the possibility to more natural, non template-based interaction Zhou & Small (2020).

6.2 OPEN QUESTIONS AND PROSPECTS FOR DESIGNING AGENTS

In contrast with the above steps which should be made soon given the current pace of learning agents research, throughout this work we have met a few more fundamental questions which remain largely unaddressed. We give a short list below, without pretending to be exhaustive:

- How can an autonomous agent learn to determine the positive or negative valence of a sophisticated feedback signal such as attitudes or verbal signals?

-
- How can the observed movement of the tutor be turned into a goal to reach? How can we consider this movement as an explanation that helps the agent understanding the goal?
 - More generally, how can we endow an autonomous agent with the capability to learn to reach goals just from the observation of intended or non intended demonstrations?
 - How can an agent infer information about the goal by reasoning about the way the environment is setup or by observing the behavior of the tutor?
 - How can we endow agents with the capability to generalize immediately what they have learned from an intended demonstration to other contexts?

Finally, while the DECSTR and IMAGINE agents address several aspects of the *language grounding* issue, there is still some remaining work before these agents can solve the harder *symbol grounding* problem Harnad (1990). In short, language in DECSTR and IMAGINE is more indexical than symbolic because the language tokens do not form a system Nieder (2009). The acquisition of symbolic behaviour is now an emerging topic Santoro et al. (2021) which can be of fundamental importance for teachable autonomous agents as, from one side, considering a social partner is necessary to establish conventional meaning and, from the other side, such agents may need the flexibility of symbolic behaviour to appropriately learn from natural tutors.

6.3 PROSPECTS ON DEVELOPMENTAL PSYCHOLOGY AND EDUCATION

In the introduction, we have already mentioned that the design of teachable autonomous agents would a lot improve the applicability of robots in interaction with non-expert users, and could provide a set of fundamental capabilities towards stronger AI systems. Beyond this, we believe these researches can have a strong impact on developmental psychology and practical educational concerns, such as teaching to children suffering from autistic spectrum disorders.

As was already the case for research on Reinforcement Learning (RL) for computational neurosciences Daw & Doya (2006), research on teachable autonomous agents from an AI perspective could be turned into research on computational modelling of developmental psychology and educational sciences phenomena. As put forward in Vollmer & Schillingmann (2018), using teachable agents or robots would help a lot performing studies about tutoring processes with strictly controlled experimental conditions from the side of the learner, which is not easy with children. However, for such studies to provide useful models of natural teaching between a tutor and a child, the robot or agent has to be a convincing model of a human learner. Currently, the existing agents and robots are far from meeting these expectations, which mostly explains why there has been few studies focusing on human tutoring processes in the context where a human subject is teaching a robot or an agent. Among other things, current agents lack the autonomy and curiosity displayed by children. Thus, more autonomous and better teachable agents could contribute to improving our understanding and models of human tutoring processes, from the teacher and the learner side. In turn, this would help a lot designing better teaching agents, finally resulting in very fruitful synergies at the crossroad between developmental psychology, developmental robotics and social robotics.

Finally, a teachable autonomous agent could be of interest in the context of practical education when embodied into a robot. Children are often fond of robots and a robotic teacher is a tool of interest for many reasons Vollmer & Schillingmann (2018). But a robotic learner can also be of interest for practical education in the context of *learning by teaching*, which as been shown to provide the best level of retention in children Leelawong & Biswas (2008); Biswas et al. (2005); Zhu et al. (2018); Gargot et al. (2021). Indeed, similarly to robotic teachers, a teachable autotelic robotic agent would provide an always available and motivating partner, even though engagement may not always result in better learning performance Nasir et al. (2021). But, in the case of learning by teaching, it would also provide a valuable learner for any person engaged in teaching activities, such as preschool children Council et al. (2001) or children with an autism spectrum disorder Vismara & Rogers (2010). Indeed, although results are mixed Tapus et al. (2012), research indicates that robots generate

a high degree of motivation even in these children Boucenna et al. (2014). Given that their behavior can easily be customized for specific needs, these robots could thus be valuable interaction partners either as teachers or as learners, in the context of learning by teaching.

7 CONCLUSION

In this paper we have advocated for a line of research building on truly autonomous agents, who decide what to do driven by their own goals, but endowing them with an additional capability to be taught so that they choose their goals in accordance with what their users are expecting from them.

By investigating the way children are taught in a classical developmental psychology work, we claimed that considering autotelic agents endowed with a capability to set their own goals was a better starting point towards teachable agents than sticking to standard RL. We have then described some of the ongoing and immediate future work along this line of research, and listed some of the central issues which must be overcome to get closer to the way children are taught.

Though we have shown that a lot remains to be done, we believe that combining this better starting point with the fast progress currently observed in the design of autonomous learning agents can soon result in the availability of good enough teachable autonomous agents to use them for quantitative analyses in developmental psychology studies and for a better design of educational programs. We also believe that this starting point will be a key move towards artificial agents which are better inserted in the society, with improved capabilities to communicate with and to adapt to their human users, which is one of the central concerns of AI research.

ACKNOWLEDGMENTS

The authors would like to thank Katarina Begus for advices about this paper. This project has partly received funding from the European Union’s Horizon 2020 research and innovation programmes under grant agreement No 761758 (HumanE AI NET) and No 765955 (ANIMATAS).

REFERENCES

- Abbeel, Pieter and Ng, Andrew Y. Apprenticeship learning via inverse reinforcement learning. In Brodley, Carla E. (ed.), *Machine Learning, Proceedings of the Twenty-first International Conference (ICML 2004), Banff, Alberta, Canada, July 4-8, 2004*, volume 69 of *ACM International Conference Proceeding Series*. ACM, 2004. doi: 10.1145/1015330.1015430. URL <https://doi.org/10.1145/1015330.1015430>.
- Akakzia, Ahmed, Colas, Cédric, Oudeyer, Pierre-Yves, Chetouani, Mohamed, and Sigaud, Olivier. Grounding language to autonomously-acquired skills via goal generation. In *ICLR 2021-Ninth International Conference on Learning Representation*, pp. 1–21, 2021.
- Akkaya, Ilge, Andrychowicz, Marcin, Chociej, Maciek, Litwin, Mateusz, McGrew, Bob, Petron, Arthur, Paino, Alex, Plappert, Matthias, Powell, Glenn, Ribas, Raphael, et al. Solving rubik’s cube with a robot hand. *arXiv preprint arXiv:1910.07113*, 2019.
- Andrychowicz, Marcin, Crow, Dwight, Ray, Alex, Schneider, Jonas, Fong, Rachel, Welinder, Peter, McGrew, Bob, Tobin, Josh, Abbeel, Pieter, and Zaremba, Wojciech. Hind-sight experience replay. In Guyon, Isabelle, von Luxburg, Ulrike, Bengio, Samy, Wallach, Hanna M., Fergus, Rob, Vishwanathan, S. V. N., and Garnett, Roman (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 5048–5058, 2017. URL <https://proceedings.neurips.cc/paper/2017/hash/453fadbdb8a1a3af50a9df4df899537b5-Abstract.html>.

-
- Andrychowicz, Marcin, Baker, Bowen, Chociej, Maciek, Jozefowicz, Rafal, McGrew, Bob, Pachocki, Jakub, Petron, Arthur, Plappert, Matthias, Powell, Glenn, Ray, Alex, et al. Learning dexterous in-hand manipulation. *arXiv preprint arXiv:1808.00177*, 2018.
- Argall, Brenda D., Chernova, Sonia, Veloso, Manuela, and Browning, B. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57:469–483, 2009.
- Baddeley, Alan. Working memory. *Science*, 255(5044):556–559, 1992.
- Baddeley, Alan. The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, 4(11):417–423, 2000.
- Baddeley, Alan. Working memory and language: An overview. *Journal of communication disorders*, 36(3):189–208, 2003.
- Bahdanau, Dzmitry, Hill, Felix, Leike, Jan, Hughes, Edward, Hosseini, Seyed Arian, Kohli, Pushmeet, and Grefenstette, Edward. Learning to understand goal specifications by modelling reward. 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=H1xsSjC9Ym>.
- Bandura, Albert and McClelland, David C. *Social learning theory*, volume 1. Englewood cliffs Prentice Hall, 1977.
- Barsalou, Lawrence W. Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521):1281–1289, 2009.
- Berlyne, Daniel E. Curiosity and exploration. *Science*, 153(3731):25–33, 1966.
- Biswas, Gautam, Leelawong, Krittaya, Schwartz, Daniel, Vye, Nancy, and at Vanderbilt, The Teachable Agents Group. Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19(3-4):363–392, 2005.
- Bobu, Andreea, Wiggert, Marius, Tomlin, Claire, and Dragan, Anca D. Feature expansive reward learning: Rethinking human input. *arXiv preprint arXiv:2006.13208*, 2020.
- Bonaiuto, James and Arbib, Michael A. Extending the mirror neuron system model, ii: what did i just do? a new role for mirror neurons. *Biological cybernetics*, 102(4):341–359, 2010.
- Bonaiuto, James, Rosta, Edina, and Arbib, Michael. Extending the mirror neuron system model, i. *Biological cybernetics*, 96(1):9–38, 2007.
- Bornstein, Marc H., Tamis-LeMonda, Catherine S., Tal, Joseph, Ludemann, Pamela, Toda, Sueko, Rahn, Charles W., Pêcheux, Marie-Germaine, Azuma, Hiroshi, and Vardi, Danya. Maternal responsiveness to infants in three societies: The United States, France, and Japan. *Child development*, 63(4):808–821, 1992.
- Borsa, Diana, Heess, Nicolas, Piot, Bilal, Liu, Siqi, Hasenclever, Leonard, Munos, Remi, and Pietquin, Olivier. Observational learning by reinforcement learning. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1117–1124. International Foundation for Autonomous Agents and Multiagent Systems, 2019.
- Botvinick, Matthew, Ritter, Sam, Wang, Jane X, Kurth-Nelson, Zeb, Blundell, Charles, and Hassabis, Demis. Reinforcement learning, fast and slow. *Trends in cognitive sciences*, 23(5):408–422, 2019.
- Boucenna, Sofiane, Narzisi, Antonio, Tilmont, Elodie, Muratori, Filippo, Pioggia, Giovanni, Cohen, David, and Chetouani, Mohamed. Interactive technologies for autistic children: A review. *Cognitive Computation*, 6(4):722–740, 2014.
- Brand, Rebecca J, Baldwin, Dare A, and Ashburn, Leslie A. Evidence for ‘motionese’: modifications in mothers’ infant-directed action. *Developmental science*, 5(1):72–83, 2002.

-
- Breazeal, Cynthia and Thomaz, Andrea L. Learning from human teachers with socially guided exploration. In *2008 IEEE International Conference on Robotics and Automation*, pp. 3539–3544. IEEE, 2008.
- Brown, Tom B., Mann, Benjamin, Ryder, Nick, Subbiah, Melanie, Kaplan, Jared, Dhariwal, Prafulla, Neelakantan, Arvind, Shyam, Pranav, Sastry, Girish, Askell, Amanda, Agarwal, Sandhini, Herbert-Voss, Ariel, Krueger, Gretchen, Henighan, Tom, Child, Rewon, Ramesh, Aditya, Ziegler, Daniel M., Wu, Jeffrey, Winter, Clemens, Hesse, Christopher, Chen, Mark, Sigler, Eric, Litwin, Mateusz, Gray, Scott, Chess, Benjamin, Clark, Jack, Berner, Christopher, McCandlish, Sam, Radford, Alec, Sutskever, Ilya, and Amodei, Dario. Language models are few-shot learners. In Larochelle, Hugo, Ranzato, Marc'Aurelio, Hadsell, Raia, Balcan, Maria-Florina, and Lin, Hsuan-Tien (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL <https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfc4967418bfb8ac142f64a-Abstract.html>.
- Bruner, Jerome. The narrative construction of reality. *Critical inquiry*, 18(1):1–21, 1991.
- Bruner, Jerome. Culture, mind, and education. *Contemporary theories of learning*, pp. 159–168, 2009a.
- Bruner, Jerome S. Organization of early skilled action. *Child development*, pp. 1–11, 1973.
- Bruner, Jerome S. *The process of education*. Harvard University Press, 2009b.
- Bruner, Jerome Seymour. *Acts of meaning*, volume 3. Harvard University Press, 1990.
- Buckman, Jacob, Hafner, Danijar, Tucker, George, Brevdo, Eugene, and Lee, Honglak. Sample-efficient reinforcement learning with stochastic ensemble value expansion. In Bengio, Samy, Wallach, Hanna M., Larochelle, Hugo, Grauman, Kristen, Cesa-Bianchi, Nicolò, and Garnett, Roman (eds.), *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pp. 8234–8244, 2018. URL <https://proceedings.neurips.cc/paper/2018/hash/f02208a057804ee16ac72ff4d3cec53b-Abstract.html>.
- Burke, Christopher J, Tobler, Philippe N, Baddeley, Michelle, and Schultz, Wolfram. Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences*, 107(32):14431–14436, 2010.
- Cangelosi, Angelo, Metta, Giorgio, Sagerer, Gerhard, Nolfi, Stefano, Nehaniv, Chrystopher, Fischer, Kerstin, Tani, Jun, Belpaeme, Tony, Sandini, Giulio, Nori, Francesco, et al. Integration of action and language knowledge: A roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development*, 2(3):167–195, 2010.
- Caruana, Richard. Multitask learning. *Machine Learning*, 28(1):41–75, 1997.
- Cederborg, Thomas and Oudeyer, Pierre-Yves. A social learning formalism for learners trying to figure out what a teacher wants them to do. *Paladyn: Journal of Behavioral Robotics*, 5:64–99, 2014.
- Celemin, Carlos and Ruiz-del Solar, Javier. Coach: learning continuous actions from corrective advice communicated by humans. In *2015 International Conference on Advanced Robotics (ICAR)*, pp. 581–586. IEEE, 2015.
- Chakraborti, Tathagata, Kambhampati, Subbarao, Scheutz, Matthias, and Zhang, Yu. Ai challenges in human-robot cognitive teaming. *arXiv preprint arXiv:1707.04775*, 2017.
- Chan, Harris, Wu, Yuhuai, Kiros, Jamie, Fidler, Sanja, and Ba, Jimmy. Actrce: Augmenting experience via teacher’s advice for multi-goal reinforcement learning. *arXiv preprint arXiv:1902.04546*, 2019.

-
- Chen, Annie S, Nam, HyunJi, Nair, Suraj, and Finn, Chelsea. Batch exploration with examples for scalable robotic reinforcement learning. *arXiv preprint arXiv:2010.11917*, 2020.
- Christiano, Paul F., Leike, Jan, Brown, Tom B., Martic, Miljan, Legg, Shane, and Amodei, Dario. Deep reinforcement learning from human preferences. In Guyon, Isabelle, von Luxburg, Ulrike, Bengio, Samy, Wallach, Hanna M., Fergus, Rob, Vishwanathan, S. V. N., and Garnett, Roman (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 4299–4307, 2017. URL <https://proceedings.neurips.cc/paper/2017/hash/d5e2c0adad503c91f91df240d0cd4e49-Abstract.html>.
- Chu, Junyi and Schulz, Laura E. Play, curiosity, and cognition. *Annual Review of Developmental Psychology*, 2:317–343, 2020.
- Cideron, Geoffrey, Seurin, Mathieu, Strub, Florian, and Pietquin, Olivier. Self-educated language agent with hindsight experience replay for instruction following. *arXiv preprint arXiv:1910.09451*, 2019.
- Colas, Cédric, Oudeyer, Pierre-Yves, Sigaud, Olivier, Fournier, Pierre, and Chetouani, Mohamed. CURIOS: Intrinsically motivated multi-task, multi-goal reinforcement learning. In *International Conference on Machine Learning (ICML)*, pp. 1331–1340, 2019.
- Colas, Cédric, Karch, Tristan, Lair, Nicolas, Dussoux, Jean-Michel, Moulin-Frier, Clément, Dominey, Peter F., and Oudeyer, Pierre-Yves. Language as a cognitive tool to imagine goals in curiosity driven exploration. In Larochelle, Hugo, Ranzato, Marc’Aurelio, Hadsell, Raia, Balcan, Maria-Florina, and Lin, Hsuan-Tien (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020a. URL <https://proceedings.neurips.cc/paper/2020/hash/274e6fcf4a583de4a81c6376f17673e7-Abstract.html>.
- Colas, Cédric, Karch, Tristan, Sigaud, Olivier, and Oudeyer, Pierre-Yves. Intrinsically motivated goal-conditioned reinforcement learning: a short survey. *arXiv preprint arXiv:2012.09830*, 2020b.
- Council, National Research et al. *Educating children with autism*. National Academies Press, 2001.
- Csibra, Gergely and Gergely, György. Natural pedagogy. *Trends in cognitive sciences*, 13(4):148–153, 2009.
- Dautenhahn, Kerstin. Getting to know each other—artificial social intelligence for autonomous robots. *Robotics and autonomous systems*, 16(2-4):333–356, 1995.
- Daw, Nathaniel D and Doya, Kenji. The computational neurobiology of learning and reward. *Current opinion in neurobiology*, 16(2):199–204, 2006.
- Dawson, Geraldine, Toth, Karen, Abbott, Robert, Osterling, Julie, Munson, Jeff, Estes, Annette, and Liaw, Jane. Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental psychology*, 40(2):271, 2004.
- Doncieux, Stephane, Filliat, David, Díaz-Rodríguez, Natalia, Hospedales, Timothy, Duro, Richard, Coninx, Alexandre, Roijers, Diederik M., Girard, Benoît, Perrin, Nicolas, and Sigaud, Olivier. Open-ended learning: a conceptual framework based on representational redescription. *Frontiers in Robotics and AI*, 12, 2018. doi: 10.3389/fnbot.2018.00059.
- Doolittle, Peter E. Vygotsky’s zone of proximal development as a theoretical foundation for cooperative learning. *Journal on Excellence in College Teaching*, 8(1):83–103, 1997.
- Dorner, Florian E. Measuring progress in deep reinforcement learning sample efficiency, 2021.

-
- Eppe, Manfred, Nguyen, Phuong DH, and Wermter, Stefan. From semantics to execution: Integrating action planning with reinforcement learning for robotic causal problem-solving. *Frontiers in Robotics and AI*, 6:123, 2019.
- Eppe, Manfred, Gumbsch, Christian, Kerzel, Matthias, Nguyen, Phuong DH, Butz, Martin V, and Wermter, Stefan. Hierarchical principles of embodied reinforcement learning: A review. *arXiv preprint arXiv:2012.10147*, 2020.
- Etcheverry, Mayalen, Moulin-Frier, Clément, and Oudeyer, Pierre-Yves. Hierarchically organized latent modules for exploratory search in morphogenetic systems. In Larochelle, Hugo, Ranzato, Marc’Aurelio, Hadsell, Raia, Balcan, Maria-Florina, and Lin, Hsuan-Tien (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL <https://proceedings.neurips.cc/paper/2020/hash/33a5435d4f945aa6154b31a73bab3b73-Abstract.html>.
- Finn, Chelsea. *Learning to Learn with Gradients*. PhD thesis, UC Berkeley, 2018.
- Florensa, Carlos, Held, David, Geng, Xinyang, and Abbeel, Pieter. Automatic goal generation for reinforcement learning agents. In Dy, Jennifer G. and Krause, Andreas (eds.), *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1514–1523. PMLR, 2018. URL <http://proceedings.mlr.press/v80/florensa18a.html>.
- Forestier, Sébastien, Mollard, Yoan, and Oudeyer, Pierre-Yves. Intrinsically motivated goal exploration processes with automatic curriculum learning. *arXiv preprint arXiv:1708.02190*, 2017.
- Forney, Andrew, Pearl, Judea, and Bareinboim, Elias. Counterfactual data-fusion for online reinforcement learners. In Precup, Doina and Teh, Yee Whye (eds.), *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1156–1164. PMLR, 2017. URL <http://proceedings.mlr.press/v70/forney17a.html>.
- Fournier, Pierre, Sigaud, Olivier, and Chetouani, Mohamed. Combining artificial curiosity and tutor guidance for environment exploration. In *Workshop on Behavior Adaptation, Interaction and Learning for Assistive Robotics at IEEE RO-MAN*, pp. 1–8, Lisbon, Portugal, 2017.
- Fournier, Pierre, Colas, Cédric, Chetouani, Mohamed, and Sigaud, Olivier. Clic: Curriculum learning and imitation for object control in non-rewarding environments. *IEEE Transactions on Cognitive and Developmental Systems*, 2019.
- Gallese, Vittorio, Keysers, Christian, and Rizzolatti, Giacomo. A unifying view of the basis of social cognition. *Trends in cognitive sciences*, 8(9):396–403, 2004.
- Gargot, Thomas, Asselborn, Thibault, Zammouri, Ingrid, Brunelle, Julie, Johal, Wafa, Dillenbourg, Pierre, Archambault, Dominique, Chetouani, Mohamed, Cohen, David, and Anzalone, Salvatore M. "it is not the robot who learns, it is me" treating severe dysgraphia using children-robot interaction. *Frontiers in Psychiatry*, 12:5, 2021.
- Gentner, Dedre. Structure-mapping: A theoretical framework for analogy. *Cognitive science*, 7(2):155–170, 1983.
- Gentner, Dedre. Language as cognitive tool kit: How language supports relational thought. *American psychologist*, 71(8):650, 2016.
- Gentner, Dedre and Hoyos, Christian. Analogy and abstraction. *Topics in cognitive science*, 9(3):672–693, 2017.
- Gentner, Dedre and Stevens, Albert L. *Mental models*. Psychology Press, 2014.

-
- Goldberg, Adele E. Constructions: A New Theoretical Approach to Language. *Trends in cognitive sciences*, 7(5):219–224, 2003. Publisher: Elsevier.
- Goldstein, Michael H, King, Andrew P, and West, Meredith J. Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, 100(13):8030–8035, 2003.
- Gopnik, Alison, Meltzoff, Andrew N, and Kuhl, Patricia K. *The scientist in the crib: Minds, brains, and how children learn*. William Morrow & Co, 1999.
- Griffith, Shane, Subramanian, Kaushik, Scholz, Jonathan, Jr., Charles L. Isbell, and Thomaz, Andrea Lockerd. Policy shaping: Integrating human feedback with reinforcement learning. In Burges, Christopher J. C., Bottou, Léon, Ghahramani, Zoubin, and Weinberger, Kilian Q. (eds.), *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pp. 2625–2633, 2013. URL <https://proceedings.neurips.cc/paper/2013/hash/e034fb6b66aacc1d48f445ddfb08da98-Abstract.html>.
- Grizou, Jonathan, Lopes, Manuel, and Oudeyer, Pierre-Yves. Robot learning simultaneously a task and how to interpret human instructions. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, pp. 1–8. IEEE, 2013.
- Grizou, Jonathan, Iturrate, Iñaki, Montesano, Luis, Oudeyer, Pierre-Yves, and Lopes, Manuel. Interactive learning from unlabeled instructions. In Zhang, Nevin L. and Tian, Jin (eds.), *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence, UAI 2014, Quebec City, Quebec, Canada, July 23-27, 2014*, pp. 290–299. AUAI Press, 2014. URL https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=2464&proceeding_id=30.
- Gruber, Romana, Schiestl, Martina, Boeckle, Markus, Frohnwieser, Anna, Miller, Rachael, Gray, Russell D, Clayton, Nicola S, and Taylor, Alex H. New caledonian crows use mental representations to solve metatool problems. *Current Biology*, 29(4):686–692, 2019.
- Gweon, Hyowon. Inferential social learning: how humans learn from others and help others learn. *Preprint at, https://doi.org/10.31234*, 2021.
- Harnad, Steve. The symbol grounding problem. *Physica D*, 42:335–346, 1990.
- Hermann, Karl Moritz, Hill, Felix, Green, Simon, Wang, Fumin, Faulkner, Ryan, Soyer, Hubert, Szepesvari, David, Czarnecki, Wojciech Marian, Jaderberg, Max, Teplyashin, Denis, et al. Grounded language learning in a simulated 3d world. *arXiv preprint arXiv:1706.06551*, 2017.
- Hester, Todd, Vecerik, Matej, Pietquin, Olivier, Lanctot, Marc, Schaul, Tom, Piot, Bilal, Horgan, Dan, Quan, John, Sendonaris, Andrew, Osband, Ian, Dulac-Arnold, Gabriel, Agapiou, John P., Leibo, Joel Z., and Gruslys, Audrunas. Deep q-learning from demonstrations. In McIlraith, Sheila A. and Weinberger, Kilian Q. (eds.), *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 3223–3230. AAAI Press, 2018. URL <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16976>.
- Hill, Felix, Santoro, Adam, Barrett, David G. T., Morcos, Ari S., and Lillicrap, Timothy P. Learning to make analogies by contrasting abstract relational structure. 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=SylLYsCcFm>.

-
- Ho, Mark K., Littman, Michael L., MacGlashan, James, Cushman, Fiery, and Austerweil, Joseph L. Showing versus doing: Teaching by demonstration. In Lee, Daniel D., Sugiyama, Masashi, von Luxburg, Ulrike, Guyon, Isabelle, and Garnett, Roman (eds.), *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pp. 3027–3035, 2016. URL <https://proceedings.neurips.cc/paper/2016/hash/b5488aeff42889188d03c9895255cecc-Abstract.html>.
- Ho, Mark K, MacGlashan, James, Littman, Michael L, and Cushman, Fiery. Social is special: A normative framework for teaching with and learning from evaluative feedback. *Cognition*, 167:91–106, 2017.
- Ibarz, Borja, Leike, Jan, Pohlen, Tobias, Irving, Geoffrey, Legg, Shane, and Amodei, Dario. Reward learning from human preferences and demonstrations in atari. In Bengio, Samy, Wallach, Hanna M., Larochelle, Hugo, Grauman, Kristen, Cesa-Bianchi, Nicolò, and Garnett, Roman (eds.), *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pp. 8022–8034, 2018. URL <https://proceedings.neurips.cc/paper/2018/hash/8cbe9ce23f42628c98f80fa0fac8b19a-Abstract.html>.
- Jaderberg, Max, Czarnecki, Wojciech M, Dunning, Iain, Marris, Luke, Lever, Guy, Castaneda, Antonio Garcia, Beattie, Charles, Rabinowitz, Neil C, Morcos, Ari S, Ruderman, Avraham, et al. Human-level performance in 3d multiplayer games with population-based reinforcement learning. *Science*, 364(6443):859–865, 2019.
- Jara-Ettinger, Julian. Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, 29:105–110, 2019.
- Jeon, Hong Jun, Milli, Smitha, and Dragan, Anca D. Reward-rational (implicit) choice: A unifying formalism for reward learning. In Larochelle, Hugo, Ranzato, Marc’Aurelio, Hadsell, Raia, Balcan, Maria-Florina, and Lin, Hsuan-Tien (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL <https://proceedings.neurips.cc/paper/2020/hash/2f10c1578a0706e06b6d7db6f0b4a6af-Abstract.html>.
- Jiang, Yiding, Gu, Shixiang, Murphy, Kevin, and Finn, Chelsea. Language as an abstraction for hierarchical deep reinforcement learning. In Wallach, Hanna M., Larochelle, Hugo, Beygelzimer, Alina, d’Alché-Buc, Florence, Fox, Emily B., and Garnett, Roman (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 9414–9426, 2019. URL <https://proceedings.neurips.cc/paper/2019/hash/0af787945872196b42c9f73ead2565c8-Abstract.html>.
- Johnson-Laird, Philip Nicholas. *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press, 1983.
- Knox, W. Bradley and Stone, Peter. 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*, pp. 5–12. International Foundation for Autonomous Agents and Multiagent Systems, 2010.
- Koechlin, Etienne and Summerfield, Christopher. An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences*, 11(6):229–235, 2007.
- Kuhl, Patricia K. Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, 5(11):831–843, 2004.
- Lake, Brenden M, Ullman, Tomer D., Tenenbaum, Joshua B., and Gershman, Samuel J. Building machines that learn and think like people. *Behavioral and brain sciences*, 40, 2017.

-
- Lanier, John B., McAleer, Stephen, and Baldi, Pierre. Curiosity-driven multi-criteria hindsight experience replay. *CoRR*, abs/1906.03710, 2019. URL <http://arxiv.org/abs/1906.03710>.
- Laversanne-Finot, Adrien, Péré, Alexandre, and Oudeyer, Pierre-Yves. Curiosity driven exploration of learned disentangled goal spaces. *arXiv preprint arXiv:1807.01521*, 2018.
- Lee, Sanghack and Bareinboim, Elias. Causal effect identifiability under partial-observability. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 5692–5701. PMLR, 2020. URL <http://proceedings.mlr.press/v119/lee20a.html>.
- Leelawong, Krittaya and Biswas, Gautam. Designing learning by teaching agents: The betty’s brain system. *International Journal of Artificial Intelligence in Education*, 18(3): 181–208, 2008.
- Lehman, Joel and Stanley, Kenneth O. Exploiting open-endedness to solve problems through the search for novelty. *Artificial Life*, 11:329, 2008.
- Lehman, Joel and Stanley, Kenneth O. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary computation*, 19(2):189–223, 2011.
- Li, Richard, Jabri, Allan, Darrell, Trevor, and Agrawal, Pulkit. Towards practical multi-object manipulation using relational reinforcement learning. *arXiv preprint arXiv:1912.11032*, 2019.
- Lin, Jinying, Ma, Zhen, Gomez, Randy, Nakamura, Keisuke, He, Bo, and Li, Guangliang. A review on interactive reinforcement learning from human social feedback. *IEEE Access*, 8:120757–120765, 2020.
- Lindblom, Jessica and Ziemke, Tom. Social situatedness of natural and artificial intelligence: Vygotsky and beyond. *Adaptive Behavior*, 11(2):79–96, 2003.
- Luketina, Jelena, Nardelli, Nantas, Farquhar, Gregory, Foerster, Jakob N., Andreas, Jacob, Grefenstette, Edward, Whiteson, Shimon, and Rocktäschel, Tim. A survey of reinforcement learning informed by natural language. In Kraus, Sarit (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. doi: 10.24963/ijcai.2019/880. URL <https://doi.org/10.24963/ijcai.2019/880>.
- Lungarella, Max, Metta, Giorgio, Pfeifer, Rolf, and Sandini, Giulio. Developmental robotics: a survey. *Connection Science*, 15(4):151–190, 2003.
- Lynch, Corey and Sermanet, Pierre. Grounding language in play. *arXiv preprint arXiv:2005.07648*, 2020.
- MacGlashan, James and Littman, Michael L. Between imitation and intention learning. In Yang, Qiang and Wooldridge, Michael J. (eds.), *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pp. 3692–3698. AAAI Press, 2015. URL <http://ijcai.org/Abstract/15/519>.
- Madumal, Prashan, Miller, Tim, Sonenberg, Liz, and Vetere, Frank. Distal explanations for model-free explainable reinforcement learning. *arXiv preprint arXiv:2001.10284*, 2020a.
- Madumal, Prashan, Miller, Tim, Sonenberg, Liz, and Vetere, Frank. Explainable reinforcement learning through a causal lens. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 2493–2500. AAAI Press, 2020b. URL <https://aaai.org/ojs/index.php/AAAI/article/view/5631>.

-
- Mandler, Jean M. Preverbal representation and language. *Language and space*, pp. 365, 1999.
- Mandler, Jean M. On the spatial foundations of the conceptual system and its enrichment. *Cognitive science*, 36(3):421–451, 2012.
- Marcus, Gary. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*, 2018.
- Meltzoff, Andrew N. Born to learn: What infants learn from watching us. *The role of early experience in infant development*, pp. 1–10, 1999.
- Mnih, Volodymyr, Kavukcuoglu, Koray, Silver, David, Rusu, Andrei A, Veness, Joel, Belle-mare, Marc G., Graves, Alex, Riedmiller, Martin, Fidjeland, Andreas K., Ostrovski, Georg, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.
- Nagai, Yukie and Rohlfing, Katharina J. Can motionese tell infants and robots” what to imitate”. In *Proceedings of the 4th International Symposium on Imitation in Animals and Artifacts*, pp. 299–306. Citeseer, 2007.
- Nair, Ashvin, Pong, Vitchyr, Dalal, Murtaza, Bahl, Shikhar, Lin, Steven, and Levine, Sergey. Visual reinforcement learning with imagined goals. In Bengio, Samy, Wallach, Hanna M., Larochelle, Hugo, Grauman, Kristen, Cesa-Bianchi, Nicolò, and Garnett, Roman (eds.), *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pp. 9209–9220, 2018. URL <https://proceedings.neurips.cc/paper/2018/hash/7ec69dd44416c46745f6edd947b470cd-Abstract.html>.
- Nair, Ashvin, Bahl, Shikhar, Khazatsky, Alexander, Pong, Vitchyr, Berseth, Glen, and Levine, Sergey. Contextual imagined goals for self-supervised robotic learning. *arXiv preprint arXiv:1910.11670*, 2019.
- Najar, Anis, Sigaud, Olivier, and Chetouani, Mohamed. Training a robot with evaluative feedback and unlabeled guidance signals. In *25th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 261–266. IEEE, 2016.
- Najar, Anis, Sigaud, Olivier, and Chetouani, Mohamed. Interactively shaping robot behaviour with unlabeled human instructions. *Autonomous Agents and Multi-Agent Systems*, 34:1–35, 2020.
- Nasir, Jauwairia, Bruno, Barbara, Chetouani, Mohamed, and Dillenbourg, Pierre. What if social robots look for productive engagement? *International Journal of Social Robotics*, pp. 1–17, 2021.
- Nehaniv, Chrystopher L. and Dautenhahn, Kerstin. The correspondence problem. *Imitation in animals and artifacts*, 41, 2002.
- Nguyen, Dung, Venkatesh, Svetha, Nguyen, Phuoc, and Tran, Truyen. Theory of mind with guilt aversion facilitates cooperative reinforcement learning. In *Asian Conference on Machine Learning*, pp. 33–48. PMLR, 2020a.
- Nguyen, Khanh, Misra, Dipendra, Schapire, Robert, Dudík, Miro, and Shafra, Patrick. Interactive learning from activity description. *arXiv preprint arXiv:2102.07024*, 2021.
- Nguyen, Sao Mai and Oudeyer, Pierre-Yves. Active choice of teachers, learning strategies and goals for a socially guided intrinsic motivation learner. *Paladyn Journal of Behavioral Robotics*, 3(3):136–146, 2012.
- Nguyen, Son Tung, Oguz, Ozgur S, Driess, Danny, and Toussaint, Marc. From images to task planning: How NLP can help physical reasoning, 2020b.
- Nieder, Andreas. Prefrontal cortex and the evolution of symbolic reference. *Current opinion in neurobiology*, 19(1):99–108, 2009.

-
- Oertel, Catharine, Castellano, Ginevra, Chetouani, Mohamed, Nasir, Jauwairia, Obaid, Mohammad, Pelachaud, Catherine, and Peters, Christopher. Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI*, 2020.
- Oudeyer, Pierre-Yves, Kaplan, Frédéric, and Hafner, Verena V. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286, 2007.
- Ouss, Lisa, Saint-Georges, Catherine, Robel, Laurence, Bodeau, Nicolas, Laznik, Marie-Christine, Crespín, Graciela C, Chetouani, Mohamed, Bursztejn, Claude, Golse, Bernard, Nabbout, Rima, Desguerre, Isabelle, and Cohen, David. Infant’s engagement and emotion as predictors of autism or intellectual disability in west syndrome. *European child & adolescent psychiatry*, 23(3):143–149, 2014. ISSN 1018-8827. doi: 10.1007/s00787-013-0430-x.
- Pearl, Judea. *Causality*. Cambridge university press, 2009.
- Pearl, Judea and Mackenzie, Dana. *The book of why: the new science of cause and effect*. Basic books, 2018.
- Piaget, Jean. *The development of thought: Equilibration of cognitive structures*. Viking, 1977. (Trans A. Rosin).
- Pierrot, Thomas, Perrin, Nicolas, Behbahani, Feryal, Laterre, Alexandre, Sigaud, Olivier, Beguir, Karim, and de Freitas, Nando. Learning compositional neural programs for continuous control. *arXiv preprint arXiv:2007.13363*, 2020.
- Pinsler, Robert, Akrou, Riad, Osa, Takayuki, Peters, Jan, and Neumann, Gerhard. Sample and feedback efficient hierarchical reinforcement learning from human preferences. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 596–601. IEEE, 2018.
- Pong, Vitchyr, Dalal, Murtaza, Lin, Steven, Nair, Ashvin, Bahl, Shikhar, and Levine, Sergey. Skew-fit: State-covering self-supervised reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7783–7792. PMLR, 2020. URL <http://proceedings.mlr.press/v119/pong20a.html>.
- Portelas, Rémy, Colas, Cédric, Weng, Lilian, Hofmann, Katja, and Oudeyer, Pierre-Yves. Automatic curriculum learning for deep RL: A short survey. In Bessiere, Christian (ed.), *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pp. 4819–4825. ijcai.org, 2020. doi: 10.24963/ijcai.2020/671. URL <https://doi.org/10.24963/ijcai.2020/671>.
- Rabinowitz, Neil C., Perbet, Frank, Song, H. Francis, Zhang, Chiyuan, Eslami, S. M. Ali, and Botvinick, Matthew. Machine theory of mind. In Dy, Jennifer G. and Krause, Andreas (eds.), *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4215–4224. PMLR, 2018. URL <http://proceedings.mlr.press/v80/rabinowitz18a.html>.
- Racaniere, Sebastien, Lampinen, Andrew K, Santoro, Adam, Reichert, David P, Firoiu, Vlad, and Lillicrap, Timothy P. Automated curricula through setter-solver interactions. *arXiv preprint arXiv:1909.12892*, 2019.
- Reddy, Siddharth, Dragan, Anca D., Levine, Sergey, Legg, Shane, and Leike, Jan. Learning human objectives by evaluating hypothetical behavior. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8020–8029. PMLR, 2020. URL <http://proceedings.mlr.press/v119/reddy20a.html>.
- Rizzolatti, Giacomo, Fadiga, Luciano, Gallese, Vittorio, and Fogassi, Leonardo. Premotor cortex and the recognition of motor actions. *Cognitive brain research*, 3(2):131–141, 1996.

-
- Runco, Mark A. and Jaeger, Garrett J. The Standard Definition of Creativity. *Creativity Research Journal*, 24(1):92–96, 2012. ISSN 1040-0419, 1532-6934. doi: 10.1080/10400419.2012.650092. URL <http://www.tandfonline.com/doi/abs/10.1080/10400419.2012.650092>.
- Santoro, Adam, Lampinen, Andrew, Mathewson, Kory, Lillicrap, Timothy, and Raposo, David. Symbolic behaviour in artificial intelligence. *arXiv preprint arXiv:2102.03406*, 2021.
- Schaul, Tom, Horgan, Daniel, Gregor, Karol, and Silver, David. Universal value function approximators. In Bach, Francis R. and Blei, David M. (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 1312–1320. JMLR.org, 2015. URL <http://proceedings.mlr.press/v37/schaul15.html>.
- Schembri, Massimiliano, Mirolli, Marco, and Baldassarre, Gianluca. Evolving internal reinforcers for an intrinsically motivated reinforcement-learning robot. In *2007 IEEE 6th International Conference on Development and Learning*, pp. 282–287. IEEE, 2007.
- Schmitt, Simon, Hessel, Matteo, and Simonyan, Karen. Off-policy actor-critic with shared experience replay. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8545–8554. PMLR, 2020. URL <http://proceedings.mlr.press/v119/schmitt20a.html>.
- Schrittwieser, Julian, Antonoglou, Ioannis, Hubert, Thomas, Simonyan, Karen, Sifre, Laurent, Schmitt, Simon, Guez, Arthur, Lockhart, Edward, Hassabis, Demis, Graepel, Thore, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.
- Schultz, W., Dayan, P., and Montague, P. R. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997.
- Siegler, Robert S. *Emerging minds: The process of change in children’s thinking*. Oxford University Press, 1998.
- Sigaud, Olivier and Stulp, Freek. Policy search in continuous action domains: an overview. *Neural Networks*, 113:28–40, 2019.
- Silver, David, Hubert, Thomas, Schrittwieser, Julian, Antonoglou, Ioannis, Lai, Matthew, Guez, Arthur, Lanctot, Marc, Sifre, Laurent, Kumaran, Dhharshan, Graepel, Thore, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.
- Smith, Linda and Gasser, Michael. The development of embodied cognition: Six lessons from babies. *Artificial life*, 11(1-2):13–29, 2005.
- Steels, Luc. The autotelic principle. In *Embodied artificial intelligence*, pp. 231–242. Springer, 2004.
- Suay, Halit Bener and Chernova, Sonia. Effect of human guidance and state space size on interactive reinforcement learning. In *2011 Ro-Man*, pp. 1–6. IEEE, 2011.
- Sutton, Richard S et al. *Introduction to reinforcement learning*. MIT Press, 1998.
- Taniguchi, Tadahiro, Ugur, Emre, Hoffmann, Matej, Jamone, Lorenzo, Nagai, Takayuki, Rosman, Benjamin, Matsuka, Toshihiko, Iwahashi, Naoto, Oztog, Erhan, Piater, Justus, et al. Symbol emergence in cognitive developmental systems: a survey. *IEEE Transactions on Cognitive and Developmental Systems*, 11(4):494–516, 2018.
- Tapus, Adriana, Peca, Andreea, Aly, Amir, Pop, Cristina, Jisa, Lavinia, Pintea, Sebastian, Rusu, Alina S, and David, Daniel O. Children with autism social engagement in interaction with nao, an imitative robot: A series of single case experiments. *Interaction studies*, 13(3):315–347, 2012.

-
- Thelen, Esther and Smith, Linda B. *A dynamic systems approach to the development of cognition and action*. MIT press, 1996.
- Thomaz, Andrea L and Breazeal, Cynthia. Experiments in socially guided exploration: Lessons learned in building robots that learn with and without human teachers. *Connection Science*, 20(2-3):91–110, 2008a.
- Thomaz, Andrea L and Breazeal, Cynthia. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence*, 172(6-7):716–737, 2008b.
- Thorndike, E. L. *Animal Intelligence*. MacMillan Company, New York, 1911.
- Tomasello, Michael. Emulation learning and cultural learning. *Behavioral and Brain Sciences*, 21(5):703–704, 1998.
- Tomasello, Michael. The Item-Based Nature of Children’s Early Syntactic Development. *Trends in cognitive sciences*, 4(4):156–163, 2000. Publisher: Elsevier.
- Tomasello, Michael. *Constructing a language*. Harvard university press, 2009.
- Tomasello, Michael and Olguin, Raquel. Twenty-Three-Month-Old Children Have a Grammatical Category of Noun. *Cognitive development*, 8(4):451–464, 1993. Publisher: Elsevier.
- Tomasello, Michael, Carpenter, Malinda, Call, Josep, Behne, Tanya, and Moll, Henrike. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, 28(5):675–691, 2005.
- Torabi, Faraz, Warnell, Garrett, and Stone, Peter. Behavioral cloning from observation. In Lang, Jérôme (ed.), *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pp. 4950–4957. ijcai.org, 2018. doi: 10.24963/ijcai.2018/687. URL <https://doi.org/10.24963/ijcai.2018/687>.
- Triesch, Jochen, Jasso, Hector, and Deák, Gedeon O. Emergence of mirror neurons in a model of gaze following. *Adaptive Behavior*, 15(2):149–165, 2007.
- Turing, Alan M. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.
- Ugur, Emre, Oztop, Erhan, and Sahin, Erol. Goal emulation and planning in perceptual space using learned affordances. *Robotics and Autonomous Systems*, 59(7-8):580–595, 2011.
- Varni, James W, Lovaas, O Ivar, Koegel, Robert L, and Everett, Nancy L. An analysis of observational learning in autistic and normal children. *Journal of Abnormal Child Psychology*, 7(1):31–43, 1979.
- Večerík, Matej, Hester, Todd, Scholz, Jonathan, Wang, Fumin, Pietquin, Olivier, Piot, Bilal, Heess, Nicolas, Rothörl, Thomas, Lampe, Thomas, and Riedmiller, Martin. Leveraging demonstrations for deep reinforcement learning on robotics problems with sparse rewards. *arXiv preprint arXiv:1707.08817*, 2017.
- Vélez, Natalia and Gweon, Hyowon. Learning from other minds: An optimistic critique of reinforcement learning models of social learning. *Current Opinion in Behavioral Sciences*, 38:110–115, 2021.
- Vinyals, Oriol, Babuschkin, Igor, Czarnecki, Wojciech M, Mathieu, Michaël, Dudzik, Andrew, Chung, Junyoung, Choi, David H, Powell, Richard, Ewalds, Timo, Georgiev, Petko, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.
- Vismara, Laurie A and Rogers, Sally J. Behavioral treatments in autism spectrum disorder: what do we know? *Annual review of clinical psychology*, 6:447–468, 2010.

-
- Vollmer, Anna-Lisa and Schillingmann, Lars. On studying human teaching behavior with robots: a review. *Review of Philosophy and Psychology*, 9(4):863–903, 2018.
- Vollmer, Anna-Lisa, Mühlig, Manuel, Steil, Jochen J, Pitsch, Karola, Fritsch, Jannik, Rohlfing, Katharina J, and Wrede, Britta. Robots show us how to teach them: Feedback from robots shapes tutoring behavior during action learning. *PloS one*, 9(3):e91349, 2014.
- Vollmer, Anna-Lisa, Wrede, Britta, Rohlfing, Katharina J., and Oudeyer, Pierre-Yves. Pragmatic frames for teaching and learning in human–robot interaction: Review and challenges. *Frontiers in neurorobotics*, 10:1–10, 2016.
- Vygotsky, L. S. Tool and Symbol in Child Development. In *Mind in Society*, chapter Tool and Symbol in Child Development, pp. 19–30. Harvard University Press, 1978. ISBN 0674576292. doi: 10.2307/j.ctvjf9vz4.6.
- Wellman, Henry M. *The child’s theory of mind*. The MIT Press, 1992.
- Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. Autonomous mental development by robots and animals. *Science*, 291(5504):599–600, 2001.
- Wood, David, Bruner, Jerome Seymour, and Ross, Gail. The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, 17(2):89–100, 1976.
- Xu, Zhongwen, van Hasselt, Hado, Hessel, Matteo, Oh, Junhyuk, Singh, Satinder, and Silver, David. Meta-gradient reinforcement learning with an objective discovered online, 2020.
- Yoshida, Hanako and Smith, Linda B. Sound Symbolism and Early Word Learning in Two Languages. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 25, pp. 1287 – 1292, 2003.
- Yu, Yue, Shafto, Patrick, Bonawitz, Elizabeth, Yang, Scott C-H, Golinkoff, Roberta M, Corriveau, Kathleen H, Hirsh-Pasek, Kathy, and Xu, Fei. The theoretical and methodological opportunities afforded by guided play with young children. *Frontiers in psychology*, 9: 1152, 2018.
- Zhou, Li and Small, Kevin. Inverse reinforcement learning with natural language goals. *arXiv preprint arXiv:2008.06924*, 2020.
- Zhu, Shengyu, Ng, Ignavier, and Chen, Zhitang. Causal discovery with reinforcement learning. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL <https://openreview.net/forum?id=S1g2skStPB>.
- Zhu, Xiaojin, Singla, Adish, Zilles, Sandra, and Rafferty, Anna N. An overview of machine teaching. *arXiv preprint arXiv:1801.05927*, 2018.
- Zlatev, Jordan. The epigenesis of meaning in human beings, and possibly in robots. *Minds and Machines*, 11(2):155–195, 2001.