Toward Teachable Autotelic Agents

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

Abstract—Autonomous discovery and direct instruction are two distinct sources of learning in children but education sciences demonstrate that mixed approaches such as assisted discovery or guided play result in improved skill acquisition. In the field of artificial intelligence (AI), these extremes, respectively, map to autonomous agents learning from their own signals and interactive learning agents fully taught by their teachers. In between should stand teachable autonomous agents (TAA): agents that learn from both internal and teaching signals to benefit from the higher efficiency of assisted discovery. Designing such agents will enable real-world nonexpert users to orient the learning trajectories of agents toward their expectations. More fundamentally, this may also be a key step to build agents with human-level intelligence. This article presents a roadmap toward the design of teachable autonomous agents from a developmental AI perspective. Building on developmental psychology and education sciences, we start by identifying key features enabling assisted discovery processes in child–tutor interactions. This leads to the production of a checklist of features that future TAAs will need to demonstrate. The checklist allows us to precisely pinpoint the various limitations of current reinforcement learning agents and to identify the promising first steps toward TAAs. It also shows the way forward by highlighting key research directions toward the design or autonomous agents that can be taught by ordinary people via natural pedagogy.

Index Terms—Artificial neural networks, autonomous systems, human-robot interaction, reinforcement learning (RL), social intelligence.

I. 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, if it makes its own decisions. Autonomy matters for artificial intelligence (AI). Indeed, at first glance, intelligence seems to require some autonomy: 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 constraint, e.g., consideration for our needs and expectations, without ethics. This agent would be useless, perhaps even dangerous. Thus, 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 on its own, how can we influence it to make it useful anyways? Charkraboti et al. [1] proposed that autonomous agents should understand and adapt to human behavior much like humans adapt to the behavior of other humans. How to obtain such an adaptation? Part of the answer can be found in the conclusion of the seminal paper of Alan Turing about AI.

“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 it to understand and speak English. That process could follow the normal teaching of a child” [2].

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 in the sense that we cannot fully control their behavior. Nevertheless, we can succeed in influencing their behavior through many ways, including “normal teaching.”

But again, we cannot teach an agent if it is “radically autonomous” in the sense outlined above. So, resolving the contradiction between true autonomy and usefulness requires a weaker notion of autonomy. In the following sections, we propose to equate autonomy with the concept of autotelicity: the ability to set one’s own goals and to learn to achieve them using one’s own learning signals [3], [4]. Autotelic agents are equipped with forms of intrinsic motivations enabling them to represent, generate, and pursue their own goals. Though autotelic agents pursue their own goals, their behavior can still be influenced by external signals. Such agents could be made appropriate and useful through teaching if we can influence its goal representations or goal sampling strategy. Thus, this article proposes to reach useful autonomous agents by developing teachable autotelic agents—agents that choose their own goals but whose choice still benefits from teaching via natural social interactions (see Fig. 1).

According to education sciences, children can learn in three ways: 1) direct instruction, where a tutor explicitly sets the learning goals of children step by step; 2) unassisted discovery—also called free play—where children are left on their
own to discover new things; and 3) guided play or assisted discovery, where the tutor intervenes on the discovery process of children to make it more fruitful [5], [6]. Besides, it has been made clear that when learning in interaction with other agents, children are inferential social learners and that social inference mechanisms play a key role in their extraordinary learning capabilities [7].

Our goal in this article is to provide developmental AI researchers a roadmap toward teachable and autotelic agents which can leverage the three ways of learning outlined above as well as social inference mechanisms. Besides, we target readers from the interactive learning and interactive reinforcement learning communities, showing them the benefits they could draw from considering autotelic agents, and researchers from the reinforcement learning community who tend to ignore the benefits they could draw from including some social learning capabilities into their agents.

Indeed, when transposed to AI research, the first way children learn under the strict guidance of a caregiver or tutor corresponds to interactive learning, and particularly interactive reinforcement learning research [8], whereas the second way, learning on their own, corresponds to research on autotelic reinforcement learning agents [4]. However, the third way, guided play, has no counterpart in current AI research though it is believed to be the most efficient [6]. A major step toward endowing artificial agents with such capabilities consists of combining the properties of interactive and autotelic reinforcement learning agents together with inferential social learning capabilities.

Equipping AI agents with such capabilities will help them acquire the various socio-cultural skills required when immersed in human societies, such as skills specific to regions or social groups, as these can only be acquired through practical interactions with social partners. For these agents, being teachable, autonomous and capable of social inferences means being equipped with the core capabilities to acquire such skills.

Beyond this practical concern, a deeper, more fundamental reason for endowing AI agents with such capabilities stems from the endeavor of building a human-level AI [9]. It might be the case, as put forward by the social situatedness vision of researchers like Vygotsky [10], Bruner [11], [12], [13], Tomasello [14], and others [15], [16] that, in addition to the capability to pursue their own goals, social interactions with caregivers, tutors, and mates are themselves necessary conditions for the emergence of sophisticated forms of intelligence in agents, see [17] for a review. In particular, these social and 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 social interactions in the acquisition of representational capabilities compatible with those of social partners [10], [14].

Thus, what we propose in this article is a roadmap toward a subset of developmental AI interested in the design of agents capable of learning on their own as well as learning from social interactions. This roadmap is organized as follows. In Section II, we extract from the developmental psychology and education sciences literature a teachability checklist tailored for our roadmap and emerging from the natural teaching of children. More precisely, we organize these properties in three categories according to whether they involve the learning child, the tutor, or the whole tutoring process. We argue that these properties are needed to teach autotelic agents. In Section III, we scrutinize current research in interactive reinforcement learning, autotelic reinforcement learning, and inferential social learning under the light of these identified properties to better identify the limits that our proposed roadmap can overcome. We then outline in Section IV the emergence of a new line of research striving to combine the properties of interactive agents and autotelic agents. In particular, we describe the IMAGINE [18], DECSTR [19], and GANGSTR [20] agents as well as the HME protocol [20] as examples of autotelic agent research striving to endow AI agents with both problem-solving capabilities and primitive interactive learning capabilities. The main outcome of this work is Table I, where we recap all the surveyed approaches and state whether they demonstrate the properties expected from teachable autonomous agents. In Section V, we build on the table and highlight key research directions toward the design or autonomous agents that can be taught by ordinary people via natural pedagogy.

II. NORMAL TEACHING OF CHILD

This section leverages observations from developmental psychology to characterize the aspects of how children learn and how tutors contribute to their learning process that are relevant to our perspective. Its outcome is a list of properties that is in no way meant to be exhaustive, but rather to ground the teachable autotelic agent roadmap that we will describe.
TABLE I
Teachability Checklist. The Table States Whether the Algorithms in the Columns Support the Properties in the Rows. x (RED): NO, o (LIGHT GREEN): PRELIMINARY, • (GREEN): YES, AND N/A (GRAY): NOT APPLICABLE. The Column on Autotelic Agents Considers Isolated Agents, Thus, None of the Properties Related to Social Interactions Can Be Applied

| Properties                  | Agents          | IRL agents | Autotelic Agents | Inferential Social Agents | IMAGINE | DECSTR | HME + GANGSTR |
|-----------------------------|-----------------|------------|------------------|--------------------------|---------|--------|---------------|
| Learner Properties (Autotelism) |                |            |                  |                          |         |        |               |
| Autotelic learning          | N/A             |            | •                | o                        | •       | •      |               |
| Open-ended learning         | x               |            | x                | x                        | x       | x      |               |
| Few shot learning           | x               |            | o                | x                        | o       | o      |               |
| Hierarchical learning       | x               |            | o                | x                        | x       | o      |               |
| Learner Properties (Social awareness) |        |            |                  |                          |         |        |               |
| Sensitivity to social signals | •               |            | N/A             | •                        | •       | •      |               |
| Proficient language learning | o               |            | N/A             | x                        | o       | o      | x             |
| Recognition of pedagogical signals | x         |            | N/A             | x                        | •       | x      | x             |
| Observational learning      | x               |            | N/A             | x                        | x       | x      |               |
| Learner Properties (Social inference) | |            |                  |                          |         |        |               |
| Modelling the tutor         | x               |            | N/A             | •                        | x       | x      | •             |
| Internalization             | x               |            | N/A             | x                        | x       | x      |               |
| Pragmatic learning          | x               |            | N/A             | •                        | x       | x      |               |
| Tutor Properties             |                |            |                  |                          |         |        |               |
| Motivation Regulation       | x               |            | N/A             | •                        | x       | x      | •             |
| ZPD management              | x               |            | N/A             | •                        | o       | o      | •             |
| Pedagogical demonstrations  | x               |            | N/A             | •                        | x       | x      | x             |
| Modelling the learner       | x               |            | N/A             | •                        | x       | x      |               |
| Tutoring Process Properties |                |            |                  |                          |         |        |               |
| Social-based tutoring strategy | •               |            | N/A             | •                        | o       | x      | o             |
| Task-based tutoring strategy | •               |            | N/A             | •                        | x       | x      | x             |
| Social communication-based transparency | •           |            | N/A             | •                        | x       | x      | x             |
| Task-based transparency     | •               |            | N/A             | •                        | •       | x      |               |

in the next sections of this article. First, we extract the properties of learning children (Section II-A), then those of the tutors (Section II-B) and finally, those of tutoring interactions themselves (Section II-C).

A. Properties of Child Learners

In the list of properties below, the first four are related to the task learning capabilities of children and the rest to their social learning capabilities.

Children Are Autotelic Learners: Either through free or assisted play, children engage in sensorimotor interactions with their environment and discover new things which they might later try to reproduce. During free-play, children rely on their intrinsic motivations to spontaneously explore their surroundings and unlock new reachable goals [21], [22], [23]. They automatically take ownership of these discovered goals, try to build their own understanding of them, and attempt to pursue them. This process is very important for subsequent learning under the guidance of a tutor. For instance, the experiments in [24] start with a short period of free-play with the experimental setup. This time is necessary for them to build an understanding of their surrounding before they can be influenced by social signals.

Children Are Open-Ended Learners: An extraordinary property of natural learning in children is that it is open-ended: the child can solve new problems of increasing difficulty up to becoming an adult and keeps learning during their whole life. Given the potentially infinite set of goals that they may pursue, children have to select some goal at all times. For that, they may attribute to potential goals a value of interest that evolves with time, resulting in efficiently organizing their own developmental learning trajectory [25], [26], [27]. In other words, they self-define a learning curriculum that makes them very sample efficient: they avoid spending too much time on goals that are either too easy or too difficult, focusing on goals that present the right level of complexity at the right time.

Children Are Few Shot Learners: Children can transfer what they have learned from solving one task to a novel one. This flexibility allows them to benefit from some prior knowledge when facing a new problem that looks like the one they already know. Consequently, children are few shot learners: they can leverage what they learned in previous tasks to master new tasks in a few trials. 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.

Children Are Hierarchical Learners: It is often the case that our tasks in everyday life have a hierarchical structure. For instance, the block assembly task of [24] involves several repetitions of the same basic block manipulation movements. More generally, the idea that children are hierarchical learners is pervasive in the developmental psychology literature [28]. The elementary skills mastered by children are often stepping...
stones for discovering how to learn other skills of increasing complexity. As Bruner writes:

“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.” [29]

Children Are Social Learners: Even shortly after birth, newborn infants can imitate complex facial expressions, such as happiness, sadness, and surprise [30], [31], [32]. Infants can detect caregiver’s eyes and prefer to look at pictures of direct gaze over averted gaze [33]. The developmental psychology literature reported several evidence of children’s social responsiveness during interaction with caregivers [10], [14], [29]. This literature demonstrates children’s social sensitivity, in particular when actions are directed to them such as for infant-directed speech [34].

Children Are Proficient Language Learners: Only humans master complex compositional and recursive language. Children’s puzzling ability to learn it so rapidly seems to be essential to their development. Indeed, children born deaf with no access to recursive sign language and Romanian children socially abandoned in Ceausescu’s orphanages showed decreased abilities for abstract compositional thinking and mental simulation [35]. The mastery of compositional language seems to further support other cognitive functions such as creativity [36], [37] or analogical reasoning [38].

Children Recognize Pedagogical Signals: Children are able to recognize the pedagogical stance of a tutor from the general context of the initiated interaction and from communicative signals of the action itself [39]. Children are sensitive to ostensive cues, such as gaze, pointing, or gesture modulation (“motionese” [40]), which are generated by the caregivers to signal their pedagogical intent. This specific strategy of caregivers and the children sensitivity to it is called natural pedagogy [39]. The children are able to infer the communicative intention and correctly interpret the instrumental intention to learn from the pedagogical caregiver.

Children Are Observational Learners: In natural interactions, children observe others acting in the environment even when no pedagogical intention is present and still extract a lot of information from these observations, a process referred to as observational learning [41], [42].

Children Construct a Model of the Tutor: During assisted-play, it is crucial that children adjust their understanding of the task at hand using the available social signals. As [24] put it, “children understand goals before being able to produce them.”

Understanding the goal from the behavior of a tutor mostly relies on inferring the tutor’s expectations and reasoning to figure out how to meet them. Thus, this process clearly calls upon a mental model of the tutor’s expectation. Such a mental model derives from a more general model called a Theory of Mind (ToM), as illustrated in Fig. 2 and put forward in recent developmental psychology papers [7], [43]. Having a ToM means being capable of reasoning about other people’s mental states [44]. The most common mental states these theories refer to are beliefs, desires, and intentions.

Children Internalize Social Signals: In Vygotsky’s theory of development, higher level cognitive capacities first appear as interpersonal processes before children internalize them and turn them into intrapersonal psychological tools [10]. Parents first narrate the child’s activities, orient their attention, keep them motivated, or decompose tasks for them. As children grow, they progressively internalize this social narration into private speech (outer speech for oneself) and, eventually, inner speech. Just like we use traditional tools to augment our control on the physical world, psychological tools augment our control on our own thoughts and behaviors. In a wealth of studies, private speech was proved instrumental to children’s ability to reason and solve tasks. They use it for planning [10], [45], and even more so when the task gets harder [46]. Children seem to internalize models of their tutors and self-generate tutor-like guidance and judgments on their own behavior.

Children Are Pragmatic Learners: Several striking experiments have shown that infants use probabilistic inference guided by an intuitive understanding of how other people think, plan, and act [7]. When presented with teaching signals, they consider how the information is generated, by whom, for whom, and why [47]. Children also leverage a causal understanding of how and why those behaviors came to be, that is, a generative model of other minds. Children consider tutor’s mental states (i.e., goals, beliefs, and desires) but also their utilities (i.e., costs and rewards). Given this evidence, they are able to efficiently infer the intent that is communicated to them, and thus rapidly understand the task at hand [7]. All these properties make them pragmatic learners and play a key role in their learning efficiency.

B. Properties of Tutors

Up to now, we have focused on the properties of children as efficient learners. We now investigate the properties of tutors.
We do not aim to design an artificial tutoring agent, but to extract from this perspective the properties that help a teachable agent better respond to natural tutoring signals, coming either from a human or another artificial agent.

**Tutors Regulate Children’s Motivation:** Reference [24] outlines that most tutoring interactions are targeting motivation regulation in children. They intend to keep them engaged in pursuing the instructed goal rather than their own goals. In turn, children must have developed their own interests beforehand, independently from social pressure. To illustrate this, [48] opposes a discovery condition, where free-play precedes social interaction, to a confirmation condition, where it is the opposite: the tutor shows what to do and then leaves children on their own. They show that freely acting on a toy beforehand allows children to construct their own motivations, which then can be further regulated if needed in guided play. Reference [24] also accounts for the discovery condition when children are left to freely engage and get familiarized with the task before the tutoring process kicks off. As the latter starts, the authors identify three processes that regulate the children’s motivations. Through recruitment, the tutor should find a way so that the child engages into the targeted goal rather than its other own goals. Through direction maintenance, the tutor ensures that the child keeps committing to that specific goal, providing incentives to make further progress toward the goal. Finally, frustration control is meant to prevent children from giving up on the instructed goal.

**Tutors Maintain Children in Their Zone of Proximal Development:** In the construction domain, a scaffolding is a temporary physical structure used to support a working crew. The term has been borrowed by education researchers to refer to assistance provided by a tutor to support learning [10], [24], [49], [50]. The two key aspects of scaffolding are 1) helping the learner with yet unmanageable skills while 2) allowing them to do as much as possible without help. First, in order to grasp initially hard tasks, the tutor can, for example, break them into more manageable subparts or engage in a thinking aloud process, providing guidelines on how the task should be performed [51]. Second, by leaving them unassisted as much as possible, the tutor helps learners take responsibility over the task. This makes scaffolding a temporary process, eventually enabling the learner to work independently.

Embedded within the scaffolding process is Vygostsky’s concept of the zone of proximal development (ZPD) [10]. The ZPD is defined as the area between what the learner can accomplish on its own and what can be accomplished with the help of a tutor. Notably, the ZPD is always shifting as the child learns. As a result, tutor interventions must constantly be individualized to address this change and ensure ZPD management until children eventually internalize the information and get exclusively self-regulated.

**Tutors Are Modeling the Tutee:** To efficiently regulate the motivational system of children and provide appropriate instructions bringing them toward their ZPD, 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” [24]. Modeling the learner helps the tutor interpret what children are trying to do, so as to efficiently teach them. The authors go further and consider that this interpretation process consists of generating hypotheses about the behavior of children, something supposedly intuitive for humans.

**Tutors Are Pedagogical:** Reference [24] provides several cues showing that natural tutoring interactions do not rely much on demonstrations to be followed blindly. They observe that, among the 30 children of their study, “there was not a single instance of what might be called blind matching behavior.” Blind matching behavior is what would be observed if children where 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 is not a way to learn how to perform the tutor’s actions, it is more a way to move to the next problem-solving step. Beyond showing what to do next, demonstrations also play a role in learning new skills. If they are not used for blind imitation, how do they help? “Demonstrating or ‘modeling’ 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” [24]. In fact, the tutor is “imitating” in the idealized form an attempted solution tried (or assumed to have been tried) by children in the expectation that they will then “imitate” it back in a more appropriate form. That is, the tutor builds on a model of the learner’s knowledge to communicate on the problem-solving steps that are still inadequate. In other terms, the tutor shows rather than it does, which characterizes pedagogical teaching.

### C. Properties of the Tutoring Process

Finally, we extract properties of the tutoring process as interactions between tutors and their tutees, building on the social learning perspective of [52]. Tutoring can be conceptualized as a mutual exchange process using two main communication channels, the social channel and the task channel, were both partners can play alternatively the role of the emitter or the recipient (see Fig. 2).

**Tutoring Combines the Social and Task Channels:** To be successful, tutoring should exploit and combine both the social and the task channels. On the one hand, the social channel involves instructions, feedback, or gaze (Fig. 3). It allows learners to not only adjust their own understanding of the task through nonmotor signals but also constructively learn to understand the tutor’s signals. On the other hand, the task channel involves watching motor interactions with the environment performed with the intent to teach or not (Fig. 4). Besides, children learn to adjust their beliefs from goal-related signals, thus fostering their proactive communication abilities.

**Partners Are Emitters and Recipients:** Both participants usually alternate the emitter and recipient roles, resulting in a mutual exchange (Fig. 2). For example, when learners execute a task in order to obtain a feedback from the tutor,
they are first the provider and then the recipient. In more
detail, the tutor may emit through the social channel when
providing instructions or nonverbal feedback, and through
the task channel when performing or showing the task. We
call the former social-based tutoring strategy and the latter
task-based tutoring strategy. From the other side, the
learner emits through the social channel when providing feed-
back on their understanding or asking questions. We call
this social communication-based transparency. The learner
emits through the task channel when imitating the tutor or
performing the task to display their current capabilities, we
call it task-based transparency. Note that when both the
tutor and the learner are artificial agents, since the learner
has to perform the task, the presence of task-based trans-
parency corresponds to the fact that the tutor reacts to the
behavior of the learner. Combining both channels results in
various “frames” of exchanges where both participants can
use both channels and both roles. These frames of exchanges
being goal-oriented, they are called “pragmatic frames” in
[53].

III. INTERACTIVE, AUTOTELIC, AND INFERENTIAL SOCIAL AGENTS

After this overview of natural tutoring processes, we turn
to the AI research dedicated to the design of artificial learners and
investigate for three broad classes corresponding to the first
columns of Table I how much the account for the properties
we have listed.

A. Reinforcement Learners

Reinforcement learning (RL) is a process by which an agent
learns to solve sequential decision problems from a reward
signal [54]. The learning agent is initially ignorant of the con-
sequences of its actions and must explore its environment to
discover them. By maximizing future rewards, it progressively
learns to favor more rewarding actions while avoiding costly
ones. RL first appeared as a computational model of learning
by trial and error in rodents [55] but is also useful to explain
conditioning phenomena in monkeys [56] and human decision
making [57]. Thus, RL seems to be a natural framework for
modeling learning to solve problems in children. However, this
framework suffers from several limitations.

First, the behavior of RL agents is fully determined by the
reward. In the standard framework, this reward is externally
provided by a human designer with some specific goal in mind.
In the absence of this external reward, most RL agents would
learn nothing. As outlined in Section II-A, this is to be con-
trasted with children who can set their own goals and learn in
full autonomy.

Second, RL agents optimize a predetermined reward func-
tion. Their behavior should converge to a corresponding
optimum and stop changing, which contrasts with the open-
ended learning capabilities of children. At first glance, the
richer framework of multitask RL [58] and its derivatives
such as meta-RL [59] seem to do better but, as long as the
set of tasks is bounded, the obtained behavior should also
converge to a steady optimum. To overcome these limita-
tions, open-ended learning approaches suggest that RL agents
receive a potentially infinite sequence of unknown tasks [60].
However, this framework does not explicitly answer one of
the most important questions: where should all these tasks
and reward functions come from? Practically, all existing
approaches revert to the bounded learning case: first define
a (bounded) space of reward functions, then sample from it.
Indeed, it is hard to imagine and define a space of reward
functions that would encompass all activities humans are able
to pursue. Even in that case, how should we sample appropri-
ate tasks for an agent along its lifetime? In contrast, humans
seem to generate their own goals, learn from their own reward
signals and organizing their own learning trajectories in a
virtually infinite space of tasks.

Third, RL is notoriously slow and sample inefficient [61].
The groundbreaking successes that have made the field pop-
ular these last years have all been obtained with weeks of
heavily parallel computations that would correspond to cen-
turies of human experience. This is to be contrasted with the
few shot learning capabilities of animals [62] and particularly
humans [39].
Finally, the standard RL framework accounts for an agent learning in isolation. As a consequence, standard RL agents lack all the social properties linked to tutoring interactions that we outlined in Sections II-B and II-C.

Now that we have established the limitations of the standard RL framework, let us describe three AI research lines tackling them: 1) interactive reinforcement learning; 2) autotelic learning; and 3) inferential social learning.

B. Interactive Reinforcement Learners

Agents immersed in the real world cannot be preprogrammed to meet all their users’ expectations. Interactive learning research tackles the fundamental problem by enabling nonexpert users to communicate or teach their preferences and expectations in a natural way, as they would do with children [53]. Thus, the field of interactive learning investigates and models the way a human tutor guides the learning process of an agent by providing teaching signals [63]. More specifically, Interactive RL research focuses on the case where the agent is an RL agent.

From the perspective adopted in this article, Interactive RL can be seen as solving several of the issues outlined above. First, by definition, Interactive RL agents are social learners. Though not all Interactive RL agents are endowed with all these capabilities, one can find works where Interactive RL agents benefit from social-based tutoring strategy [64], [65], [66] or task-based tutoring strategy [67], [68] and works in which the Interactive RL agent displays some task-based transparency [69], [70] and social communication-based transparency [70], [71], [72].

The Interactive RL approach relies too much on the tutor to drive the learning process and fails to account for the autonomy of children. As for their open-ended learning properties, one may consider that a human user may specify a sequence of increasingly more difficult tasks, for instance by specifying preferences about outcomes [73]. But in the absence of autonomous learning, 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. Thus, Interactive RL accounts better map to the direct instruction approach to education than to assisted discovery.

Along the same line, though teaching signals can substantially accelerate learning, Interactive RL research does not generally consider multitask nor hierarchical contexts. Thus, the corresponding agents do not display few shot learning nor hierarchical learning capabilities.

In addition, Interactive RL research does not satisfactorily account for the social interaction properties of the tutoring processes outlined in Section II-C.

First, a well-known weakness of Interactive RL research is that the way naive human users tend to teach an agent is far from meeting the expectations of the RL framework [74], [75].

Second, the way Interactive RL research accounts for task-based tutoring strategy is called Learning from Demonstration (LfD) [68]. 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 of the learning agent—a sort of episodic memory. From this data, the agent learns an efficient policy with RL as if it was its own memories [76], [77]. Rather than using RL, an alternative approach called behavioral cloning (BC) consists of cloning the imitated policy by applying standard—another kind of machine learning process—from the same data to directly obtain a policy which behaves like the imitated one [78]. These methods assume the experience of the expert to be directly transferred into the memory of the learning agent, which cannot happen yet in real life given that both participants have different viewpoints. Besides, they assume that the tutor and the learner share the same state space, action repertoire, and dynamics of interaction with the environment. This is unlikely in real-life situations as made obvious by works on the so-called “correspondence problem” [79]. Finally, they assume that the agent can perfectly observe the states and actions of the demonstrator and imitate these actions, in sharp contrast with what we outlined in Section II-B. It is more likely that natural learners recognize the goals of the partner and try on their own to reach these goals. This is known as goal emulation [80], [81] and can be accounted for in the RL framework through inverse RL processes [67]. Among other things, this approach can help solve the correspondence problem.

Beyond this, in Interactive RL, teaching signals lack a communicative intent. As outlined in Section II, infants and more generally humans are sensitive to pedagogical teaching, where the teaching signals are specifically emitted to optimize learning efficiency. This is overlooked in the Interactive RL literature, where a tutor will provide the same demonstrations to all learners without taking their current knowledge in consideration, thus failing to account for the ZPD management, pedagogical teaching, and modeling the learner properties. Reciprocally, these methods expect all learners to learn equally well from a given set of demonstrations, thus Interactive RL agents cannot be seen as pragmatic learners.

Finally, natural teaching methods build on linguistic description of behaviors, instructions, explanations, and both verbal and nonverbal feedback. As for the linguistic signals, Interactive RL agents rely on a limited set of predefined tokens, which cannot be confused with a form of language proficiency. For educating agents with the richer signals used in natural teaching, these agents must be equipped with richer capabilities. We claim that having the capability to represent and pursue goals, to autonomously imagine and select goals, to infer the goals of others, and to interact with them about these goals are some of the required capabilities, as these goals can play a pivotal role at the interface between user expectations and autonomous behavior learning. These concerns play a key role in the emergence of the line of research that we describe in Section IV.

C. Autotelic Reinforcement Learners

The extraordinary transition from the mental life of human infants to the sophisticated intelligence of adults is mostly modeled in the domain of developmental robotics and AI [16], [82], [83]. A central line of research in this domain is
interested in the design of autotelic agents [3], [84]. 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 [4], [85], [86]. By definition, they are autotelic learners.

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 [54]. Most of these agents are equipped with one or several goal spaces and rely on goal-conditioned RL [4] and automatic curriculum learning [87] 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 capabilities [88], [89], [90], [91], [92]. Thus, from the one side, these agents depart from the behaviorist roots of the RL framework, making it possible to represent some fundamental cognitive processes related to goals. Besides, in contrast to Interactive RL agents, autotelic agents are open-ended and generally few shot learners. Note however that very few works focus on the importance of hierarchical learning in such agents [28], [93]. In contrast, there is a growing tendency to combine autotelic architectures with language learning capabilities. Many papers combining goal-conditioned RL and language understanding have recently flourished under the banner of instruction following agents [94], but the corresponding agents are not truly autotelic in the sense that they do not set their own goals. Autotelic agents endowed with language learning capabilities are covered in Section IV.

Thus, autotelic reinforcement learners are endowed with a lot of the properties that are missing to their interactive counterparts. But symmetrically, they generally model isolated agents, thus they miss all the interactive and social learning properties listed in Section II.

### D. Inferential Social Learners

While Interactive RL research provides efficient ways for AI agents to learn from demonstrations, feedback, or instructions, they lack methods for implementing the richest social interactions between a tutor and a tutee presented in Section II. A key to model these richer interactions consists of simultaneously addressing tutoring and learning processes.

A new family of agents that we call inferential social learners have started to implement these more elaborated types of interactions where the tutor and the learner build a model of each other and infer the other’s beliefs, desires, and intentions using these models. These models are inspired by the inferential social learning framework of [7], which considers that learners recover the meaning of underlying others’ actions by inverting intuitive causal model of the way others think, plan, and act. Such inference mechanisms are crucial for social learning, notably to improve the efficiency of the tutoring process.

By drawing inspiration from language-based communication, inferential social learning goes beyond literal interpretation of actions and considers pragmatic inference [95]. The current approach of inferential social learning mechanisms exploits probabilistic inference over structured representations of the world, the other and even the representations of the other. Goodman and Frank [96] described the rational speech act (RSA) for pragmatic reasoning in language understanding. Their approach results in probabilistic models of pragmatic speakers and listeners able to capture the meanings of complex phenomena of linguistic interaction. Taking a more developmental perspective, some other works try to account for the way parents employ “motherese” [34], [97] to talk to their children [98], so as to endow social agents with the roots of language proficiency.

Transposing these linguistic considerations to the domain of interactions with objects, some works study communicative demonstrations [99]. Such demonstrations are not just directed toward the manipulated objects, but also accompanied with nonverbal cues such as eye gaze and or exaggerations of the demonstrations in the space-time dimensions used to convey the pedagogical intent. Again, similarly to the language case, some studies have focused on “motionese,” the nonverbal equivalent of motherese [40], [100]. The corresponding agents can partly be seen as autotelic, as they have some desires and intentions, but as the corresponding works focus on the interaction itself, generally they do not come with sophisticated agents capable of open-ended learning, few shot learning or hierarchical learning.

The mechanism behind these studies always requires to have models of the other’s beliefs, desires, and intentions and to reason about them to choose the most appropriate way to communicate, for instance selectively choose when to provide feedback and corrections [101].

In the human–robot interaction community, similar works about agents not just intending to perform the ordinary action but also to convey something about it are using the concept of legibility [102], [103]. The key idea is to design transparency through motion models by which a robot communicates its intent to a human observer [70]. The main assumption is that humans will be able to infer the robot’s intention from its motion by inverting a generative model. Thus, such works account for both forms of transparency.

Ho et al. [99] proposed a pedagogical model based on Bayesian Inference to generate communicative demonstrations. They argue that this model should help the teacher select examples to communicate a concept to the learner. They further performed experiments involving real human instructors and showed that the results were aligned with their proposed model. A more recent work proposed a demonstration-based ZPD teaching strategy where demonstrations are not perfect, but are adapted to the current capabilities of the agent [104].

Driven by the Interactive RL framework where agents learn from an external reward function, some works also study how an agent can infer information about a reward function from observed premises in the tutoring context [105], [106], [107]. 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
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 [99], [101], [110].

As we noted above, social inferences require that both learners and teachers reason about the beliefs and intent of each other. Thus, the corresponding agents need to be endowed with a Theory of Mind (ToM). There have been recent attempts to account for the acquisition of a ToM through inverse RL [44] and in the domain of multiagent RL [111], but these works generally suffer from the same limitations as Interactive RL approaches: they do not consider explicit goals nor social inference processes. In robotics, efficient interactive reinforcement learners and inferential social learners reason about the beliefs and intent of each other. Thus, the corresponding agents need to be teachable at all. This is only when combining all approaches the IMAGINE agent learns to represent and understand language as a preexistent social structure through social interactions with a synthetic tutor (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 ZPD [112]. In each episode, the tutor sets the scene to provide optimal challenges: it introduces the necessary objects for the agent to reach its goal (not too hard) but generates them procedurally and adds distracting ones (not too easy).

**IV. First Steps Toward Teachable Autotelic Agents**

As outlined in the previous section, on the one hand, interactive reinforcement learners and inferential social learners are teachable, but they are not autonomous. On the other hand, autotelic agents are autonomous, but they are not teachable. To fully benefit from the efficiency of assisted discovery, agents should be simultaneously autonomous and teachable. In this section, we first describe preliminary research aiming at making autotelic agents more teachable and modeling efficient tutoring processes.

In this section, we describe three autotelic agents endowed with some teachability properties, as well as a tutoring protocol, as depicted in Fig. 5. The IMAGINE agent learns to achieve goals described through language by a social partner and can build on the compositional nature of language to imagine new goals and pursue them. The DECSTR agent represents its sensorimotor interactions with objects using abstract spatial predicates and learns to achieve goals represented in terms of these predicates. A social partner describes the performed actions and the corresponding linguistic input can be turned into instructions. The GANGSTR agent extends DECSTR with better sensorimotor learning capabilities and is trained with a tutoring protocol called “Help Me Explore” (HME) which significantly improves its capability to be taught.

**A. IMAGINE Agent**

The IMAGINE agent, illustrated in Fig. 6, is an autotelic agent, thus it benefits from all the properties of autotelic agents, apart from hierarchical learning. The aim of IMAGINE is 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 tutor. It then leverages both the communicative and cognitive functions of language to benefit from these signals [18].

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 some 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 chooses to describe some interactions and not others, it effectively guides goal representation learning in the agent. Thus, the goal representations of the IMAGINE agent are influenced by social interactions.
and, once they are formed, the agent can pursue them without relying on tutors.

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 it 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 from usage-based linguistic theories [113], [114], [115]. 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 as discussed in [116]. As the authors show, this simple mechanism empowers the creative exploration of the environment and enhances the systematic generalization abilities of the agent. Whereas the communicative function of language mostly helps internalize the goal representations of social partners, the cognitive function provides the agent with individuality and open-endedness: it can generate novel creative goals that its tutor did not mention.

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 [14], [117]. This contrasts with the instruction-based approach dominant in language-conditioned agents research [94] but rarely seen in real parent-child interactions [118]. 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 from the social partner, which is one of the target functionalities of teachable autonomous agents.

Despite these advanced uses of language, IMAGINE learns simplified pattern-based sentences and, thus, cannot be said fully language proficient. Besides, the description of actions plays a direct role in orienting the exploration process of the agent. However, in all experiments performed in [18], the agent is only targeting its own goals and the social partner only describes 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 [24] where the tutor is in charge of making sure that children will build a pyramid. Thus, though the IMAGINE agent has the potential to receive instructions, this potential has not been demonstrated and IMAGINE cannot be said sensitive to task-based tutoring strategy nor to motivation regulation. Beyond that, it does not benefit from any of the inferential social mechanisms described in Section III-D, and it does not display social communication-based transparency.

B. DECSTR Agent

In the intrinsically motivated agents outlined in Section III-C, just as in most goal-conditioned RL algorithms [4], [119], [120], the space of goals is generally defined as a subset of the state space of the agents [88], [121], [122]. 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 with general predicates. At the sensorimotor level, this helps targeting more abstract goals, opportunistically making profit of the current situation to obtain the easiest realization of the goals. The DECSTR agent interacts with blocks in the Fetch Manipulate Tutoring environment, a benchmark which was used in several recent works to train hierarchical RL agents [123] and autotelic agents [90], [124], [125]. To take an example from this domain, 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 blocks 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 linguistic 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. The capability of generating a diversity of goals for a same description 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 [19].

In more detail, the architecture of the DECSTR agent is depicted in Fig. 7. To ground language into its sensorimotor experience, it relies on a language goal generator (LGG) that takes a language expression as input and samples concrete goals matching the linguistic expression. Fig. 8 depicts the way this component endows sensorimotor autotelic agents with language sensitivity by relating language to behavior. This is precisely through this LGG that language is grounded into sensorimotor goals.

After a while, the DECSTR agent consistently learns to build pyramids and towers of three blocks. Once becoming more expert, it 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 [19].

The DECSTR agent benefits from nearly the same properties as IMAGINE. The main difference is that, in DECSTR, since language acquisition and instructions are decoupled from sensorimotor learning, language obviously plays no role in the sensorimotor learning phase thus there is no room for social-based tutoring strategy with DECSTR.

C. GANGSTR + HME Agent

The last work we investigate combines an autotelic agent and an interaction protocol.

The GANGSTR Agent: Similar to DECSTR, GANGSTR is a predicate-based autotelic agent designed for object manipulation. It faces five blocks within the Fetch Manipulate Tutoring environment and strives to discover a large spectrum of diverse
The HME Protocol: HME is a simple interaction protocol where a tutor helps an autotelic agent explore its goal space. The role of the tutor is to propose communicated goals that help the agent quickly discover its entire goal space. The autotelic agent then remembers the communicated goals and trains to pursue them as well as its own goals. When GANGSTR uses the HME protocol, both communicated goals and autotelic goals are subsets of the same predicate-based goal space as in DECSTR.

To select a goal to communicate, the tutor has a model of the agent’s current exploration limits. In GANGSTR, this model is a copy of the graph of the semantic goals discovered by the learner. Using this model, the tutor first communicates a goal at the frontier of the learner’s knowledge. If the learner succeeds in reaching this goal, the tutor uses its own, full-grown graph of goals to select a goal just beyond that frontier that the learner may reach with just one block manipulation from the current configuration. The learner then memorizes this (frontier, beyond) pair of goals and may train reaching them on its own until it sufficiently masters it. In practice, the agent first discovers easy goals. Once these goals are discovered, the agent can either follow 1) autotelic episodes by uniformly sampling a goal to pursue among the set of autonomously discovered goals or 2) social episodes where the tutor communicates a pair of goals as explained above.

The HME protocol accounts for several features of human tutoring processes described in Section II-B. First, memorizing the (frontier, beyond) pair of goals is a form of internalization mechanism where the learner can train on its own on goals communicated by the tutor. Moreover, by proposing (frontier, beyond) pair of goals, the tutor maintains the learner into its ZPD, where goals at the frontier act as stepping stones toward further discoveries. Even more strikingly, an experimental study where the rate of social episodes is varied from 0% (pure autotelic learning) to 100% (pure instruction following) shows that a moderate rate of social episodes works best. This is clearly reminiscent of the claim in educational sciences that guided play and assisted discovery work better than purely autonomous learning or direct instructions. More details about GANGSTR and the HME protocol are presented in [20].

Properties of GANGSTR + HME: The properties of GANGSTR + HME are close to those of the DECSTR agent. We already outlined that it benefits from hierarchical learning but not from language proficiency. Additionally, in contrast with DECSTR, GANGSTR incrementally learns to reach goals in interaction with its tutor, who can thus play a much more direct role in orienting the exploration process of the agent. Thus, GANGSTR + HME can thus be said sensitive to motivation regulation and to task-based tutoring strategy.

As for missing properties, GANGSTR + HME, when the tutor suggests a goal, the agent immediately pursues it. A more realistic agent should be endowed with mechanisms to arbitrate between their own goals and 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 teachable,
exactly as children. But it would also require a form of a social communication-based transparency which is currently missing in the autotelic agents we have listed.

V. DISCUSSION AND FUTURE DIRECTIONS

For all the properties we have listed in Section II, Table I recaps whether all the agents that we have presented so far display or not these properties.

One can see that IMAGINE, DECSTR, and HME + GANGSTR all display some of the properties one may expect from teachable autotelic agents. They are even complementary with respect to several of these properties. For instance, IMAGINE and DECSTR are endowed with a basic capability to interpret language, which is not the case of HME + GANGSTR. But the latter implements reciprocal modeling of the tutor and the learner, in contrast with the former methods. Thus, a good deal of the limitations of the above agents could be overcome by combining their capabilities.

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 [53] and, more generally, opening the possibility to more natural, non-template-based interaction [108]. It would also open the way toward new research questions, such as modeling feedback about goal selection rather than just about action, or the need for an arbitration mechanism between intrinsic motivations and social feedback. Besides, such an integration may require some developmental dynamics, resulting in displaying overlapping waves of sensorimotor and linguistic development [126].

However, Table I shows that such a combination would still lack most of the properties of inferential social agents. Thus, an important direction for future research will consist in combining the properties of the existing teachable autotelic agents and endow them with further social inference capabilities. To do so, the model of the learner’s skills used in the HME protocol could serve to leverage the Bayesian inference processes proposed in [43]. Besides, all tutoring processes that IMAGINE, DECSTR, and HME + GANGSTR agents can benefit from consist in descriptions and instructions about goals. These agents could be extended with capabilities to learn from pedagogical demonstrations.

Looking more closely at Table I, one can see that an agent combining the properties of IMAGINE, DECSTR, HME + GANGSTR and inferential social learning agents would still lack one property, which is language proficiency. In language-augmented agents, such as IMAGINE and DECSTR, the template-based language learning mechanism does not mimic the language acquisition processes of infants. Though, 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 [127], [128], [129] may be required to better account for these capabilities. Besides, communication itself could be extended to richer natural language interactions [130] so as to increase the teachability of these agents. The very fast progress in language learning with large transformer models [131], [132] creates an opportunity in this direction.

Finally, in contrast with the above questions which should be answered soon given the current pace of learning agents research, there are a few questions which remain largely undressed. For instance, how can an autonomous agent learn to determine the positive or negative valence of sophisticated feedback signals such as attitudes or linguistic nuances? Or how can we endow agents with the capability to generalize immediately what they have learned from an intended demonstration to other contexts? More fundamentally, how can we extend the language grounding capabilities of current teachable autotelic agents to solve the harder symbol grounding problem [133]? In short, language in DECSTR and IMAGINE is more indexical than symbolic because the language tokens do not form a system [134]. The acquisition of symbolic behavior is an emerging topic [135] which can be of fundamental importance for future agents as, from the 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 behavior to appropriately learn from natural tutors.

VI. CONCLUSION

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 this article, by investigating the way children are taught, we claimed that autotelic agents were a better starting point for such a research than standard RL agents. The resulting agents would pursue their own goals, but should be endowed with an additional capability to be taught so that they choose their goals in accordance with the expectation of their users. We have then described some of the ongoing and immediate future work along this line of research and revealed some of the issues which must be overcome to get closer to the way children are taught.

In the immediate future, the existing teachable autotelic agents could be combined and integrated with inferential social learning capabilities and more natural language learning capabilities. Given the fast progress currently observed in the design of autotelic learning agents, we expect to soon see good enough teachable autonomous agents to use them for quantitative analyses in developmental psychology studies and for a better design of education programs. We also believe that this starting point is a key move toward better insertion of AI agents 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.

ACKNOWLEDGMENT

The authors would like to thank Katarina Begus for advice about an early version of this article.
REFERENCES

[1] T. Chakraborti, S. Kambhampati, M. Scheutz, and Y. Zhang, “AI challenges in human–robot cognitive teaming,” 2017, arXiv:1707.04775.

[2] J. A. Winn, “Promises and challenges of scaffolded instruction,” Learn. Disability Quart., vol. 17, 1 no. 1, pp. 89–104, 1994.

[3] A. Bandura and D. C. McClelland, Animal Intelligence. New York, NY, USA: MacMillan, 1977.

[4] D. Klahr, “Guided play: Principles and practices,” Front. Psychol., vol. 9, p. 1152, Jul. 2018.

[5] J. A. Winn, “Promises and challenges of scaffolded instruction,” Learn. Disability Quart., vol. 17, 1 no. 1, pp. 89–104, 1994.

[6] J. S. Bruner, “Organization of early skilled action,” Child Dev., vol. 44, no. 1, pp. 1–11, 1973.

[7] A. Jara-Ettinger, “Theory of mind as inverse reinforcement learning,” Trends Cogn. Sci., vol. 13, no. 4, pp. 148–153, 2009.

[8] J. Lin, Z. Ma, R. Gomez, K. Namakura, B. He, and G. Li, “A review on interactive reinforcement learning from human social feedback,” IEEE Access, vol. 12, pp. 120757–120765, 2020.

[9] A. N. Meltzoff and K. J. Spelke, “Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli,” Dev. Psychol., vol. 24, no. 4, p. 470, 1988.

[10] J. S. Bruner, Constructing a Language. Cambridge, MA, USA: Harvard Univ. Press, 1990.

[11] D. S. Weisberg, K. Hirsh-Pasek, R. M. Golinkoff, A. K. Kittredge, and P.-Y. Oudeyer, “Intrinsically motivated goal-conditioned reinforcement learning: A short survey,” 2020, arXiv:2012.09830.

[12] A. M. Turing, “Computing machinery and intelligence,” Mind, vol. 59, no. 236, pp. 433–460, 1950.

[13] M. Tomasello, “Getting to know each other—Artificial social intelligence for autonomous robots,” Robot. Auton. Syst., vol. 16, nos. 2–4, pp. 333–356, 1995.

[14] J. S. Bruner, Actions of Meaning. vol. 3, Cambridge, MA, USA: Harvard Univ. Press, 1990.

[15] L. S. Vygotsky, Tool and symbol in child development, in Mind in Society, Cambridge, MA, USA: Harvard Univ. Press, 1978, pp. 19–30.

[16] J. L. Steels, “The autotelic principle,” in Embodied Artificial Intelligence, F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, Eds. Berlin, Germany: Springer, pp. 231–242, doi: 10.1007/978-3-540-27833-7_17.

[17] C. Colas, T. Karch, O. Sigaud, and P.-Y. Oudeyer, “Intrinsically motivated goal-conditioned reinforcement learning: A short survey,” 2020, arXiv:2012.09830.

[18] T. Chakraborti, S. Kambhampati, M. Scheutz, and Y. Zhang, “AI challenges in human–robot cognitive teaming,” 2017, arXiv:1707.04775.

[19] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, “Building machines that learn and think like people,” 2021, arXiv:2101.00872.

[20] T. Farroni, G. Csibra, and M. H. Johnson, “Eye contact slowed down the prefrontal cortex maturation simultaneously in two or more children—the Romulus and Remus hypothesis,” Res. Ideas Outcomes, vol. 5, Jul. 2019, Art. no. e38546.

[21] T. Hoffmann, “Creativity and construction grammar: Cognitive and psychological issues,” Zeitschrift Anglistik Amerikanistik, vol. 66, no. 3, pp. 259–276, 2018.

[22] L. Steels, “The autotelic principle,” in Embodied Artificial Intelligence, F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, Eds. Berlin, Germany: Springer, pp. 231–242, doi: 10.1007/978-3-540-27833-7_17.

[23] J. S. Bruner, Constructing a Language. Cambridge, MA, USA: Harvard Univ. Press, 1990.

[24] T. Chakraborti, S. Kambhampati, M. Scheutz, and Y. Zhang, “AI challenges in human–robot cognitive teaming,” 2017, arXiv:1707.04775.

[25] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, “Building machines that learn and think like people,” 2021, arXiv:2101.00872.

[26] J. Lin, Z. Ma, R. Gomez, K. Namakura, B. He, and G. Li, “A review on interactive reinforcement learning from human social feedback,” IEEE Access, vol. 12, pp. 120757–120765, 2020.

[27] A. M. Turing, “Computing machinery and intelligence,” Mind, vol. 59, no. 236, pp. 433–460, 1950.

[28] L. S. Vygotsky, Tool and symbol in child development, in Mind in Society, Cambridge, MA, USA: Harvard Univ. Press, 1978, pp. 19–30.

[29] J. L. Steels, “The autotelic principle,” in Embodied Artificial Intelligence, F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, Eds. Berlin, Germany: Springer, pp. 231–242, doi: 10.1007/978-3-540-27833-7_17.

[30] C. Colas, T. Karch, O. Sigaud, and P.-Y. Oudeyer, “Intrinsically motivated goal-conditioned reinforcement learning: A short survey,” 2020, arXiv:2012.09830.

[31] T. Chakraborti, S. Kambhampati, M. Scheutz, and Y. Zhang, “AI challenges in human–robot cognitive teaming,” 2017, arXiv:1707.04775.

[32] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, “Building machines that learn and think like people,” 2021, arXiv:2101.00872.

[33] L. Steels, “The autotelic principle,” in Embodied Artificial Intelligence, F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, Eds. Berlin, Germany: Springer, pp. 231–242, doi: 10.1007/978-3-540-27833-7_17.

[34] J. S. Bruner, Constructing a Language. Cambridge, MA, USA: Harvard Univ. Press, 1990.

[35] T. Chakraborti, S. Kambhampati, M. Scheutz, and Y. Zhang, “AI challenges in human–robot cognitive teaming,” 2017, arXiv:1707.04775.

[36] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, “Building machines that learn and think like people,” 2021, arXiv:2101.00872.

[37] L. Steels, “The autotelic principle,” in Embodied Artificial Intelligence, F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, Eds. Berlin, Germany: Springer, pp. 231–242, doi: 10.1007/978-3-540-27833-7_17.
[110] J. MacGlashan and M. L. Littman, “Between imitation and intention learning,” in Proc. 24th Int. Joint Conf. Artif. Intell., 2015, pp. 3692–3698.

[111] D. Nguyen, S. Venkatesh, P. Nguyen, and T. Tran, “Theory of mind with guilt aversion facilitates cooperative reinforcement learning,” in Proc. Asian Conf. Mach. Learn., 2020, pp. 33–48.

[112] P. E. Doolittle, “Vygotsky’s zone of proximal development as a theoretical foundation for cooperative learning.” J. Excellence College Teach., vol. 8, no. 1, pp. 83–103, 1997.

[113] M. Tomasello and R. Olguin, “Twenty-three-month-old children have a grammatical category of noun,” Cogn. Develop., vol. 8, no. 4, pp. 451–464, 1993.

[114] M. Tomasello, “The item-based nature of children’s early syntactic development,” Trends Cogn. Sci., vol. 4, no. 4, pp. 156–163, 2000.

[115] A. E. Goldberg, “Constructions: A new theoretical approach to language,” Trends Cogn. Sci., vol. 7, no. 5, pp. 219–224, 2003.

[116] M. A. Runco and G. J. Jaeger, “The standard definition of creativity,” Creativity Res. J., vol. 24, no. 1, pp. 92–96, 2012.

[117] H. Yoshida and L. B. Smith, “Sound symbolism and early word learning in two languages,” in Proc. Annu. Meeting Cogn. Sci. Soc., vol. 25, 2003, pp. 1287–1292.

[118] M. H. Bornstein et al., “Maternal responsiveness to infants in three societies: The United States, France, and Japan,” Child Develop., vol. 63, no. 4, pp. 808–821, 1992.

[119] T. Schaul, D. Horgan, K. Gregor, and D. Silver, “Universal value function approximators,” in Proc. 32nd Int. Conf. Mach. Learn., vol. 37, 2015, pp. 1312–1320.

[120] M. Andrychowicz et al., “Hindsight experience replay,” in Proc. Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 5048–5058.

[121] V. Pong, M. Dalal, S. Lin, A. Nair, S. Bahl, and S. Levine, “Skew-fit: State-covering self-supervised reinforcement learning,” in Proc. 37th Int. Conf. Mach. Learn., vol. 119, 2020, pp. 7783–7792.

[122] A. Nair, S. Bahl, A. Khazatsky, V. Pong, G. Berseth, and S. Levine, “Contextual imagined goals for self-supervised robotic learning,” 2019, arXiv:1910.11670.

[123] T. Pierrot et al., “Learning compositional neural programs for continuous control,” 2020, arXiv:2007.13363.

[124] J. B. Lanier, S. McAleer, and P. Baldi, “Curiosity-driven multi-criteria hindsight experience replay,” 2019, arXiv:1906.03710.

[125] R. Li, A. Jabri, T. Darrell, and P. Agrawal, “Towards practical multi-object manipulation using relational reinforcement learning,” 2019, arXiv:1912.11032.

[126] R. S. Siegler, Emerging Minds: The Process of Change in Children’s Thinking. New York, NY, USA: Oxford Univ. Press, 1998.

[127] M. H. Goldstein, A. P. King, and M. J. West, “Social interaction shapes babbling: Testing parallels between birdsong and speech,” Proc. Nat. Acad. Sci., vol. 100, no. 13, pp. 8030–8035, 2003.

[128] P. K. Kuhl, “Early language acquisition: Cracking the speech code,” Nat. Rev. Neurosci., vol. 5, no. 11, pp. 831–843, 2004.

[129] J. M. Mandler, “On the spatial foundations of the conceptual system and its enrichment,” Cogn. Sci., vol. 36, no. 3, pp. 421–451, 2012.

[130] C. Lynch and P. Sermanet, “Language conditioned imitation learning over unstructured data,” 2020, arXiv:2005.07648.

[131] L. Floridi and M. Chiriatti, “GPT-3: Its nature, scope, limits, and consequences,” Minds Mach., vol. 30, no. 4, pp. 681–694, 2020.

[132] J. W. Rae et al., “Scaling language models: Methods, analysis & insights from training gopher,” 2021, arXiv:2112.11446.

[133] S. Harnad, “The symbol grounding problem,” Physica D, vol. 42, pp. 335-346, Jun. 1990.

[134] A. Nieder, “Prefrontal cortex and the evolution of symbolic reference,” Current Opin. Neurobiol., vol. 19, no. 1, pp. 99–108, 2009.

[135] A. Santoro, A. Lampinen, K. Mathewson, T. Lillicrap, and D. Raposo, “Symbolic behaviour in artificial intelligence,” 2021, arXiv:2102.03406.