The SocialAI School:Insights from Developmental Psychology Towards Artificial Socio-Cultural Agents

Grgur Kovač [1] Rémy Portelas [1] Peter Ford Dominey [2] Pierre-Yves Oudeyer [1]

Abstract

Developmental psychologists have long-established the importance of socio-cognitive abilities in human intelligence. These abilities enable us to enter, participate and benefit from human culture. AI research on social interactive agents mostly concerns the *emergence* of culture in a multi-agent setting (often without a strong grounding in developmental psychology). We argue that AI research should be informed by psychology and study socio-cognitive abilities enabling to *enter* a culture too. We discuss the theories of Michael Tomasello and Jerome Bruner to introduce some of their concepts to AI and outline key concepts and socio-cognitive abilities. We present The SocialAI school - a tool including a customizable parameterized suite of procedurally generated environments, which simplifies conducting experiments regarding those concepts. We show examples of such experiments with RL agents and Large Language Models. The main motivation of this work is to engage the AI community around the problem of social intelligence informed by developmental psychology, and to provide a tool to simplify first steps in this direction.

1. Introduction

Our everyday life is immersed in a sociocultural world, which we navigate using a set of sophisticated socio-cognitive abilities. Although at first it might seem this sociocultural world is just another product of our cognition, decades of research in developmental psychology suggest the opposite. Our socio-cultural world, cultural knowledge, and our socio-cognitive abilities are the foundation of our development and both our social and asocial intelligence (Vygotsky & Cole, 1978; Bruner, 1990; Tomasello, 2019).

For Vygotsky, socio-cultural interactions are the main driver for "higher-level" cognition (Vygotsky & Cole, 1978). For him, many high-level cognitive functions first appear at the social level, and *then* develop at the individual level. This leap from interpersonal processes to intrapersonal processes is referred to as *internalization*. Vygotsky's theories influenced multiple works within cognitive science (Clark, 1996; Hutchins, 1996), primatology (Tomasello, 1999) and the developmental robotics branch of AI (Billard & Dautenhahn, 1998; Cangelosi et al., 2010; Mirolli & Parisi, 2011).

Jerome Bruner emphasized the importance of culture in human development too. He presents a pragmatic view of how the practical use of referencing and requesting pushes language development through routinized social interactions (formats), in which those abilities are *necessary* to achieve various ends. He describes these interactions as scaffolded - the caretaker gradually helps less and demands more of the child to achieve those goals (Bruner, 1985).

Finally, Michael Tomasello's work (Tomasello, 1999; 2019; 2020) constitutes a representative and contemporary assessment of the nature and central importance of culture in human development. Tomasello outlined core social abilities and motivations which (when combined with the relevant experience) enable us to participate in the cumulative cultural evolution (a powerful form of cultural transmission enabling the development and perpetuation of complex cultural artifacts and knowledge (Tomasello, 1999)).

Given the key role social cognition plays in human cognition and cultural evolution, it is natural that the field of AI studies those questions as well. A socially competent agent could learn our culture and participate in its cultural evolution, i.e. improve our concepts, theories, inventions, and create new ones. A system capable of out-of-the-box thinking creative solutions and discovering new relevant problems must learn our values and how we see and understand the world (it must learn our culture).

Enriching AI with those skills also has numerous more practical implications. Agents capable of online social learning could efficiently learn novel tasks and tools. A robot that can seamlessly infer the meaning of our gestures and utterances in new contexts could be easily *onboarded* into human teams, without requiring humans to adopt new conventions. Furthermore, robots capable of learning human values and moral norms will be capable of performing tasks in the constraints defined by those values.

AI research on interactive agents is often focused on navigation and object manipulation problems, excised of any social dimension (Mnih et al., 2015; Lillicrap et al., 2016; Andrychowicz et al., 2017). The study of sociality is mostly studied in Multi-Agent settings, where the main focus is often on the *emergence* of culture (often with only a weak grounding in developmental psychology) (Jaques et al., 2019; Baker et al., 2019). While we agree that those directions are interesting and important, in this work we focus on *entering* an already existing complex culture. And we argue that it can be beneficial to be informed by developmental psychology theories.

We do not claim The SocialAI is sufficient to reach that far and complex goal. We only propose that being informed by the concepts discussed in this paper is useful, and we present SocialAI as a tool which could be used to start investigating such questions in more details.

Following the theories of Michael Tomasello and Jerome Bruner, this work identifies a richer set of socio-cognitive skills than those currently considered in most of the DRL literature. More precisely, we focus on three key aspects of social cognition as identified by Tomasello: 1) the ability to infer what others see and to engage in joint attention, 2) the development of referential communication through pointing and the beginnings of conventionalized communication through language, and 3) the use of imitation and role reversal imitation in social learning. We also outline two concepts from Jerome Bruner's work: formats and scaffolding. Formats refer to the way in which social interactions are structured and presented, while scaffolding refers to the temporary support provided by a caretaker to help a learner achieve a task that would be otherwise too difficult.

We present The SocialAI school - a tool simplifying experiments studying the concepts outlined by those theories. It includes a parameterized procedural generation engine for social environments with various kinds of social interactions. This tool enables simple creation, modification, and extension of those environments. In our experiments, we show how various studies can be conducted using the SocialAI school. We present experiments regarding the following questions: generalization of social inferences (the pointing gesture) to new contexts, recreating an experiment from cognitive science (to study role reversal), and the impact of a scaffolded environment on the agent's learning. We conduct those experiments with RL agents, and also present an additional case study with Large Language Models (LLMs). In the appendix, we explore additional questions regarding linguistic inferences, joint attention, imitation, inferring others' field of view, and formats. We hope to encourage future work extending and building on this first set to study various questions regarding social competence (ex. new sociocultural scenarios, architectures, training regimes).

We outline the following main contributions of this work:

- An introduction to Michael Tomasello's and Jerome Bruner's theories on child development and core socio-cognitive abilities
- An outline of a set of core socio-cognitive abilities important for current AI research
- The SocialAI school: a tool including a customizable procedural generation suite of environments aiming to simplify studies of socio-cognitive abilities of AI agents
- Examples of case studies demonstrating how SocialAI can be used to study various questions regarding socio-cognitive abilities in AI

2. Related work

This work aims to connect DRL with developmental robotics (Asada et al., 2009; Cangelosi & Schlesinger, 2014), a robotics field informed by developmental psychology. Developmental robotics has already argued for the importance of social intelligence in artificial agents (Billard & Dautenhahn, 1999; Lindblom & Ziemke, 2003; Mirolli & Parisi, 2011). We aim to expand this to the DL community.

Inside the deep learning community, multiple works studied disembodied social understanding from videos or text. These include classifying videos based on the nature of shown behavior (Shu et al., 2021; 2020), inferring agents' goals, relationships, and predicting future trajectories (Netanyahu et al., 2021). Machine learning models have also been used to model agents' internal states, such as goals, beliefs, and desires (Rabinowitz et al., 2018; Baker et al., 2011). Regarding textual tasks, social reasoning abilities have been recently studied with LLMs. In extensive evaluation, LLMs were shown to struggle on two social benchmarks (Sap et al., 2022): SocialIQA (Sap et al., 2019), and especially on TOMi (Le et al., 2019). In other experiments, LLMs were evaluated on variations of false-belief tasks and exhibited promising performance (Trott et al., 2022; Kosinski, 2023). A gap with human performance was also demonstrated on implicatures - problems which heavily rely on inferring contextual information (Ruis et al., 2022). While our motivation is analogous to these works, we propose to focus on embodied and interactive agents.

DRL works studying embodied social interactions are mostly in multi-agent settings. Various intrinsic rewards were presented to foster cooperation between agents such as influence over the others' actions (Jaques et al., 2019) and fostering joint attention (Lee et al., 2021). While multi-agent reinforcement learning is often used to study the emergence of culture or communication, here we study the process of *entering* an already existing culture.

Closer to our work, an RL agent was shown to adapt (through social learning) to a new environment with an expert (Ndousse et al., 2021). The independent RL agent was trained in a multi-agent environment with various environmental constraints and an auxiliary loss. Similar experiments were also conducted at a larger scale (Bhoopchand et al., 2022). The objective of the SocialAI school is to provide a tool simplifying similar studies, which could explore socio-cognitive abilities outlined by psychology.

3. Cognitive science background

This section introduces Michael Tomasello's and Jerome Bruner's theories and concepts.

3.1. M. Tomasello - The Shared Intentionality Theory

We are born into a culture filled with cultural artifacts, symbols and institutions like language, social norms, tool industries, or even governments (Richerson & Boyd, 2006; Tomasello, 2019). These artifacts are a product of a series of modifications over many generations. Tomasello calls this *cumulative cultural evolution*, and argues that it is behind our most impressive achievements (Tomasello, 1999).

Cumulative cultural evolution is grounded in our socio-cognitive abilities (e.g. social cognition, cultural learning, communication), which enable us to learn, improve, and teach our culture (Tomasello, 2019), i.e. *enter* a culture. Cultural artifacts inherited through this process become the core of our cognition. An example of this is language, which influences our cognition in many ways. For example, it defines how we categorize and construe the world, and enables a powerful form of social learning : instructed learning (Tomasello, 1999). This makes socio-cognitive abilities crucial, as their early development bootstraps both our social and asocial cognition (Herrmann et al., 2007).

Tomasello's *Shared intentionality theory* argues that human socio-cognitive abilities, such as communication and social learning, are transformed by two developmental steps : the emergence of *Joint intentionality* at around 9 months of age (the 9-month revolution), and the emergence of *Collective intentionality* at around 3 years of age (the objective/normative turn) (Tomasello, 2019).

Joint intentionality emerges at around 9 months of age (Tomasello, 2019). It enables children to form a *joint agent* (a dyadic "we") - they understand that they work with a partner towards the same joint goal. Children begin to view dyadic social interactions through a *"dual-level structure"*: a joint agent "we" on one level, and a personal "I" on another, i.e. we both understand that we both have separate roles ("I"), and that we work together towards the same joint goal ("we"). This enables them to take the perspective of others, which can also be done recursively - they are not only both attending to the same goal, they are also both attending to the partner's attention to the goal, and they both know that they both are doing so.

Collective intentionality emerges at around 3 years of age (Tomasello, 2019). It enables children to form a cultural *group-minded "we"*, which in comparison with a dyadic "we" represents an identity for a group. For example, a child might enforce a social norm because "this is how *we*, in this culture, do things". Consequently, children begin to participate in conventions and norms, and to view things from the "objective" perspective.

These two developmental steps transform countless abilities, motivations, and behaviors. For the purpose of this paper, we focus on the following three developmental pathways: social cognition (sec. 3.1.1), communication (sec. 3.1.2), and social learning (sec. 3.1.3), as we consider them the most relevant for AI at the moment.

3.1.1. Social cognition

In this section, we discuss the development of the ability to coordinate perspectives and view things from the *objective perspective* (a perspective independent from any individual) (Tomasello, 2019). The starting point is the ability to **infer what another sees or knows**. The earliest example of this is gaze following of six-month-olds (D'Entremont et al., 1997). Here, only one perspective is processed at the time. **Joint attention (JA)** emerges at around 9 months of age. Tomasello defines JA as consisting of two elements: *triangulation* (two participants attending to he same referent) and *recursiveness* (both participants being recursively aware that they are both sharing attention) (Tomasello, 2019). JA is characterized by the dual-level structure of shared attention on one level, and individual perspectives on another. Consequently, children start to align and exchange perspectives. Once children reach a sufficient level of linguistic competence, they start sharing attention to mental content in the form of linguistic discourse (at two to three years of age). The presence of conflicting perspectives in linguistic discourse pushes children to resolve those conflicts, which they do by forming the "objective" perspective, and **coordinating other perspectives** with it. Refer to appendix B.1.1 for details.

3.1.2. Communication

Communication starts with **imperative gestures for self-serving purposes**. An example of such a gesture is the child pulling the adult's hand, requesting them to pick them up. This gesture always has the same imperative meaning, and it never refers to an external object. The 9-month revolution brings forth **referential communication** - children start to communicate triadically to external referents through pointing and pantomiming. The pointing gesture is a powerful way of communicating, as the *same* gesture can be used to express many different meanings in many different scenarios, provided that the observer can correctly infer that meaning. The ability to infer this meaning is based on the emerging abilities of joint intentionality. Those of joint attention and, most notably, "recursive inference" - to interpret a pointing gesture, we make a recursive inference of what "*you* intend for *me* to think". For example, if we are looking for a ball together, and you point to a cupboard behind me. I should infer that you are drawing my attention to the cupboard to communicate that I should look for the ball in the cupboard. The next step is the appearance of **conventionalized linguistic communication**. The underlying principle stays the same: reference to an external entity combined with inferring the meaning through recursive inferences. The difference is that, now, the child also learns the conventional means of referring (for example, words and phrases). Tomasello argues that, at first, children don't understand language as conventional, and they use it as any other tool. The understanding of language as conventional follows the emergence of collective intentionality after the third birthday, and this gives rise to a myriad of different language uses, such as discourse or pedagogy. Refer to section B.1.2 for more details.

3.1.3. Cultural Learning

Human culture is characterized by a powerful form of cultural transmission called cumulative cultural evolution - inventions quickly spread and are improved by following generations (Tomasello, 1999). These inventions spread at such a pace that they are rarely forgotten or lost. This is referred to as the *ratchet* effect (Tomasello et al., 1993) - inventions are iteratively improved without *slippage* back. This effect is enabled by human social learning abilities (ex. imitation, instructed learning), and motivations (to learn from others, but also to affiliate and conform). The earliest form of cultural learning is the **mimicking of facial expressions** (observed even in neonates (Meltzoff & Moore, 1997)). Over the course of the first year, children begin to **imitate other's actions and goals**, and then, they begin doing so in ways which demonstrate their understanding of other's as intentional agents (Meltzoff, 1995). Joint intentionality brings forth a new form of cultural learning called **role reversal imitation**. Children can reverse the roles of a collaborative activity by learning about the partners role only from playing their own. For example, children respond to an adult tickling their arm, by tickling the adult's arm (instead of its own) (Carpenter et al., 2005). This is enabled by the dual-level structure of joint intentionality through which children understand, at the same time, the joint goal of a dyadic interaction on one level, and the individuals' separate roles on another. The next big step in the development of cultural learning is learning from instructions - **instructed learning** (following the emergence of collective intentionality). It is based on the adults' motivation to teach children as well as on the children's ability to understand and learn from linguistic instructions. Children understand knowledge acquired through instructions as objective truth, and generalize it much better than knowledge acquired by other means (Butler & Tomasello, 2016). In this way we acquire the most complex knowledge and skills such as reading or algebra. Refer to section B.1.3 for more details.

3.2. Jerome Bruner

This work is also influenced by Jerone Bruner's theories, especially regarding the concepts of scaffolding (Wood et al., 1976) and formats (Bruner, 1985), which were recently reintroduced to AI as pragmatic frames (Vollmer et al., 2016).

Formats (Pragmatic frames) (Bruner, 1985) simplify learning by providing a stable structure to social interactions. They are regular patterns characterizing the unfolding of possible social interactions (equivalent to an interaction protocol or a grammar of social interactions). Formats consist of a deep structure (the static part) and a surface structure (the varying realizations managed by some rules). An example of a format is the common peek-a-boo game. The deep structure refers to the appearance and the reappearance of an object. The surface structure can be realized in different ways. For example, one might hide an object using a cloth, or hands; one might hide his face or a toy; one might do shorter or longer pauses before making the object reappear. We understand social interactions through such formats, and our social interactions are based on our ability to learn, negotiate, and use them.

Another relevant concept is scaffolding (Wood et al., 1976) (similar to Vygotsky's zone of proximal development (Vygotsky & Cole, 1978)). Scaffolding is a process through which an adult bootstraps the child's learning. The adult controls aspects of a task which are currently too hard for the child (scaffolds the interaction). The scaffold is gradually reduced as the child is ready to take on more aspects of the task, until they can solve the task alone (without scaffolding). An example is a child constructing a pyramid with the help of an adult (Wood et al., 1976). At first, the child is not even focusing on the task, and the adult tries to get its attention to the task by connecting blocks and building the pyramid in front of them. Once the child is able to focus on the task, the adult starts passing the blocks to the child to connect. In the next phase, the child is grabbing blocks by itself, and the adult is helping through verbal suggestions. Then, only verbal confirmations are needed to guide the child. Finally, the child can construct the pyramid by itself. In summary, the adult observes the child and gradually transfers parts of the task (removes the scaffold) to the child. Through this process, the caretaker enables the child to master a task they would not be able to master alone.

Theory of Mind (ToM) can be defined as the understanding that others have intentions, desires, beliefs, perceptions, and emotions different from one's own and that such intentions, desires, and so forth affect people's actions and behaviors (APA, 2023). Theories discussed here (especially Tomasello's) do not use this term much, rather they separate it into more fine-grained ones. For example, the ability to "imagine what another knows" require the processing of only one perspective, this is only the first step which later develops to the ability to coordinate multiple perspectives, and finally to form the objective perspective and address false-beliefs. Another example is the concept of recursiveness in JA: it is not sufficient to infer what another knows, but also that another knows about what you know, and so on. The socio-cognitive abilities discussed here are different aspects of ToM. Some of these those are a clear subset of the usual definition (ex. imagine what another knows), and some go beyond (coordinate multiple perspectives).

4. The SocialAI school

The SocialAI school is a tool for building interactive environments to study various questions regarding social competence, such as "What do concepts, such as social abilities and motivations, outlined by developmental psychology mean in the scope of AI?", "How can we evaluate their presence in different agents?", "What are their simplest forms and how can agents acquire them?"

To construct SocialAI, we rely on a set of key experiments and studies from developmental psychology, which were used to outline the most important abilities, motivations and developmental steps in humans. From the work of Tomasello, we focus on developments before and around the age of 9 months (we believe it is important to address those before more complex ones relating to development of 3-year-olds, see section 3.1). We study the following developmental pathways: Social cognition (inferring other's perception and joint attention), Communication (referential communication through the pointing gesture and the beginning of conventionalized communication through simple language), and Cultural Learning (imitation and role reversal imitation). From the work of Bruner, we study the concepts of Formats and Scaffolding (see section 3.2).

SocialAI, which is built on top of Minigrid (Chevalier-Boisvert et al., 2018), includes a *customizable parameterized* suite of *procedurally* generated environments. We implement this procedural generation with a tree-based structure (the parametric tree). This makes it simple to add and modify new environments, and control their sampling. All the current environments are single-agent and contain a scripted peer. The agent has to interact with the peer to reach an apple. This setup enables a controlled and minimal representation of social interactions. SocialAI also includes RL-based and LLM-based learners and evaluation protocols.

In our experiments, we present case studies to as examples of how SocialAI could be easily used and modified to ask various questions and conduct diverse experiments. To facilitate future research, SocialAI was made to be very easy to use and modify. It will be completely open sourced, and we hope that it will be useful to the community to study the questions of social intelligence in AI.

We do not claim that the SocialAI school is sufficient to construct a socially competent agent as this is a very far-reaching and complex goal. However, we believe that in aiming for this goal, concepts from developmental psychology can serve as signposts for AI - give directions and enable us to define short term goals. Given that the outlined skills are at the very core of human social and cognitive competences, artificial agents aimed at participating in and learning from social interactions with humans are likely to require the same core competences. We present the SocialAI school merely as a first step towards this goal. Refer to section C in the appendix for technical details on the environment, agents, and the procedural generation mechanisms.

5. Experiments

We show how the SocialAI school can be used to conduct experiments inspired by theories and studies described in section 3. In all our case studies, except the study with language models (sec. 5.4), we use a PPO (Schulman et al., 2017) agents with different exploration bonuses (refer to appendix D for technical details). We study the pointing gesture, role reversal imitation and scaffolding, and present a study with large language models. In appendix G we present additional experiments regarding linguistic cues, joint attention, in-episode imitation learning, the ability to infer the peer's field of view, and formats.

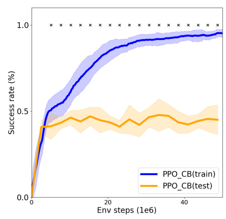


Figure 1. **The Pointing experiments.** The figure compares the success rate (mean +/- std over 8 seeds) on the training environments with the evaluation on the testing environment. The cross marks depict statistical significance ($p = 0.05$). The agent is able to infer the meaning of a pointing gesture on the training environments, but it is not able to generalize to a new social context.

5.1. Understanding the pointing gesture

In this experiment, we study the ability of an RL agent to understand the pointing gesture. This experiment is motivated by a study of childrens' ability to understand pointing gestures (Behne et al., 2005) (see section B.1.2 in the appendix). We study can an RL agent infer the meaning of a pointing gesture, and generalize this ability to new situations (infer the new meaning of a pointing gesture in a new context). It is interesting to study this kind of generalization as the power of inferring pointing gestures is based precisely on being able to infer it's meaning to *new* referents based on *new* social contexts.

The environment consists of two objects (ex. boxes) and the peer that points to the correct object. The agent then has to interact with that object (ex. open the box) to get access to an apple. The agent is trained on five problems each with different objects (Boxes, Switches, Levers, Marble, Generators), and on the *asocial* version of the Doors problem (only one door and no peer). Training on the asocial version enables the agent to learn how to use a door, which is a prerequisite for generalization of the pointing gesture to an environment with two doors. The agent is evaluated on the Doors problem in the social setting (two doors and a peer pointing to the correct one) The agent needs to combine the knowledge of how to use a door (learned on the asocial version of that problem), with inferring the meaning of the pointing gesture (learned on the other five problems), and generalize that to a new scenario where the peer points to a door. Refer to section F.1 in the appendix for details.

Figure 1 shows the success rate of the agent on the training environments ("PPO_CB_train") and its evaluation on the evaluation environment (PPO_CB(test)). We can see that while the agent easily solves the training environments (with the success rate of 95.2%), it fails to generalize. It reaches the success rate of 45.2%, which corresponds to randomly guessing the correct object. These results demonstrate that the agent can learn to infer the meaning of a pointing gesture in a familiar context, but cannot generalize to new social contexts. These results motivate future research on how an agent can be endowed with abilities for such combinatorial generalization, a potential solution could leverage LLMs.

Appendix G.1 presents experiments in which the peer, instead of pointing, provides linguistic cues for the color or the proximity of the correct object. As in the pointing experiments, we observe that while PPO agents master the training environments, they fail to generalize to a new context.

5.2. Role reversal imitation

In this experiment, we study the role-reversal capabilities of an RL agent - to what extent can it learn about the partner's role from playing its own. In doing so, we also show how a cognitive science experiment can be recreated in the scope of AI. In Fletcher (2012) apes and children were trained on one role (role B), and then tested on how long it took them to master the opposite role (role A). Results showed that children, but not apes, master role A faster than the control group (not pretrained). These results imply that children learn about the opposite role just from playing their own, i.e. they see the interaction from a bird's eye perspective. We study the following two questions: 1) How much do RL agents learn about the partner's role during a collaborative activity? 2) Does increasing diversity in the training (training on more tasks in both roles) enable the agent to learn more about the partner's role?

We conduct this study on the MarblePass task. This task consists of two roles: one participant pushes the marble to the right side of the environment (role A), from where the other can push it to the a generator, which generates apples (role B). We aim to assess how much the agent learns about the opposite role (role A), from training in its own (role B). Following Fletcher (2012) we measure the sample efficiency of fine-tuning agents to the test role. Unlike in Fletcher (2012) it is not sufficient to compare an agent pretrained on the training role with an unpretrained agent. Even if the agent pretrained on the training role learns nothing about the testing role, it would still learn about environment dynamics and one would expect it to learn faster than the unpretrained agent. For this reason, we compare with an agent pretrained on the asocial version of the training role. In this version, the agent obtains reward in the same way as in the social version, but no peer is needed - the agent and the marble are placed on the right side of the environment and the agent has to push the marble towards the generator. Therefore, this agent learns all about the relevant environment dynamics, but not about the specific collaborative activity. This agent represents the control group in Fletcher (2012).

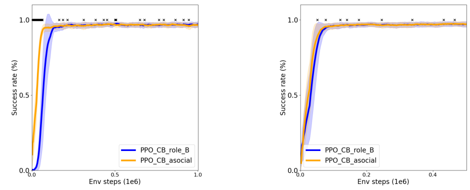
We conduct two experiments: *single* and *group*. In *single* experiments, the agents are trained only on one task : role B and the asocial version of the MarblePass problem. In *group* experiments, both agents are also trained both roles of all additional six collaborative problems (a total of 13 environments). In other words, we compare the agents pretrained in the four following ways: 1) experimental (*single*): pretrained only on role B of the MarblePass problem, 2) control (*single*): pretrained only on the asocial version of the MarblePass problem, 3) experimental (*group*): pretrained on role B of the MarblePass problem, and on both roles of all other problems, 4) control (*group*): pretrained on the asocial version of the MarblePass problem, and on both roles of all other problems. Refer to appendix for additional details.

How much do RL agents learn about the partner's role during a collaborative activity? Figure 2(a) shows the success rate of fine-tuning to role A of the MarblePass task.

It compares the experimental and the control conditions of the *single* experiments. It is interesting to note that the agent pretrained on the asocial version ("asocial") masters role A of the task slightly faster than the agent pretrained on role B of the task ("role_B"). This implies that, not only, the agent does not learn anything useful about the peer's role, but pretraining on role B actually makes it harder for the agent to learn about role A. We believe that this is because, during training in role B, the agent learns to first wait for the peer, while in the asocial version it pushes the marble right away. As, in role A, the agent pushes the marble right away too, we believe this makes it slightly easier for the asocially pretrained agent to adapt to the new role. In other words, from an egocentric view the asocial version is closer (than role B) to role A. This shows that the RL agent, rather than understanding the interaction from a bird's-eye perspective, finds the simplest way to solve the task.

Does training on additional problems enable the agent to learn more about the partner's role? Figure 2(b) shows the success rate of fine-tuning to role A of the MarblePass task. It compares the experimental and the control conditions of the *group* experiments. Here we can see that there is no significant difference in sample efficiency. We can make two observations from this. First, as the socially pretrained agent was less sample efficient in the *single* experiments, we can conclude that pretraining on many tasks reduces overfitting on role B. And second, as this agent is not more sample efficient than the asocially pretrained baseline, we can conclude that this agent does not learn anything usefull about the peer's role too.

These results imply an interesting avenue of research into how agent's attention can be directed to the partner's role and the birds-eye-view of the activity.



(a) *Single* experiment: learning role A given pretraining on role B (1 environment).

(b) *Group* experiment: learning role A given pretraining on role B and 6 other two-roles tasks (13 environments).

Figure 2. **Role reversal imitation experiments**. We study to what extent is an RL agent able to transfer knowledge from one role of a collaborative activity to another. Figure shows the success rate of fine-tuning to role A (mean $\pm$ std over 8 seeds), the cross marks depict statistical significance ($p = 0.05$). We compare a PPO agent pretrained on role B ("role_B") to that pretrained on the asocial version of the environment ("asocial"), which learns only about the environment dynamics. Agents pretrained on role B do not master role A faster than asocially pretrained agents, implying that the RL agents do exhibit role reversal capabilities.

5.3. Scaffolding

In this section, we study the concept of scaffolding (see sec. 3.2 for details): Can a scaffolded environment help an agent learn more complex interaction sequences (formats)? The environment is similar to the one in section 5.1. However, we evaluate on all six problems (instead of one) in the social version, and the peer does not point until the agent performs a more complex introductory sequence (establishes eye contact and utters "Help, please").

We compare two types of scaffolding: "scaf_4" and "scaf_8". The agent is trained in two phases. In the first phase, the agent is trained on environments with varying complexity (defined by scaffolding type). After reaching a set success rate, the training goes to the second phase in which the agent is trained only on the six testing environments. The agent denoted by "scaf_4" is trained on four different introductory sequences (requiring or not requiring eye contact and the utterance). This agent is trained on 18 different environments (six problems, four sequences). The "scaf_8" agent is also trained with those four different options. In addition, the peer can help in two different ways: pointing to the object or interacting with it and leaving the apple for the agent to eat (36 environments). The easiest environments on which the "scaf_8" agent is trained do not require an introduction and the peer leaves the apple for the agent (the agent just goes to and eats the apple). The hardest ones require the introduction with both the utterance and eye contact and the peer points to the object. The agents are evaluated on those.

Figure 3 compares the success rate of the agents trained with the two scaffolding types ("scaf_4" and "scaf_8") to that of an agent trained only on the six testing environments ("no_scaf"). We can see that only the scaffolded agents solve the testing environments, and that the agent with a more detailed scaffolding ("scaf_8") solves the environment faster. These results show that scaffolding enables the agents to learn more complex formats, and that a more thorough scaffolding further improves the efficiency. In future work, more advanced scaffolding could be explored, ex. based on learning progress (Oudeyer & Kaplan, 2007) or other surrogate objectives (Portelas et al., 2020).

5.4. Large language models as interactive agents

Large language models (LLMs) are staring to be used in various tasks (Brown et al., 2020; Touvron et al., 2023), including to control interactive agents (Yao et al., 2022; Carta et al., 2023). In order to study LLMs as interactive agents, SocialAI school enables parsing of visual grid observations

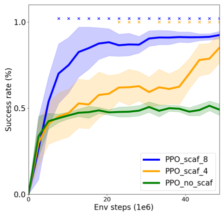


Figure 3. **The scaffolding experiment.** The comparison of agents trained on multiple environments of varying difficulty to that trained on an unscaffolded environment. The figure show success rates on the testing environments (mean ± std over 8 seeds) and the cross marks depict statistical significance ($p = 0.05$) with respect to the "no_scaf" baseline. Only the scaffolded agents ("scaf_4" and "scaf_8") solve the environment, and the scaffolding with eight difficulty levels is more sample efficient.

to text (this parsing can be easily modified and experimented with). This process is depicted in figure 24 in the Appendix.

We use two environments: AsocialBox and ColorBoxes. The AsocialBox environment contains a box, which the agent has to open to get the apple. The ColorBoxes enviornment contains two boxes and a peer. The peer says the color of the correct box, which the agent then has to open.

An LLM acts by continuing a prompt containing in-context examples, observations and actions from the current episode (last 3 steps for ColorBoxes and the full episode for AsocialBox as this gave the best performance), and the action query ("Act :"). We manually create expert trajectories for in-context examples - 7 episodes (545 words) for the AsocialBox environment, and 9 (687 words) for ColorBoxes (the full in context examples are given in appendix G.6). The model generates the textual continuation of this prompt (3 tokens for GPT models, and 3 words for bloom). If one of the available actions ("turn left", "turn right", "move forward", "toggle") is a substring of the generated text, that action is executed, otherwise, the "no_op" action is executed (the agent does not act this step). The executed action and the new observation are then added to the prompt.

We compare three LLMs: bloom-560m (Scao et al., 2022), and two GPT models (Brown et al., 2020), ada ("text-ada-001") and davinci-3 ("text-davinci-003"). We compare them to a baseline that samples a random action. We evaluate on a fixed test set of 20 environments, with 10 timesteps.

Table1 shows that, on the AsocialBox environment, the GPT models achieve a success rate of 90%, despite only observing seven expert trajectories. On ColorBoxes, davinci-3 outperforms all other models with a success rate of 35%. As this is belowe 50%, this implies that none of the models are able to use cues from the peer. All models outperform the random values and both GPT models outperform bloom-560m. The environments used in this case study are much simpler than those in other case studies (only one problem, and no introductory sequence). Nonetheless, it impressive that such a performance is achieved from only a few expert trajectories: seven for AsocialBox and nine for ColorBoxes.

Table 1. Comparison of large language models on two SocialAI environments (success rate on 20 environments). The best model ("davinci-3") reaches the success rates of 90% and 35%. While this performance is impressive given that the models observed only seven (for AsocialBox) and nine (for ColorBoxes) expert trajectories, it leaves much room for improvement.

	ada	davinci-3	bloom-560m	rand
AsocialBox	**90%**	**90%**	75%	5%
ColorBoxes	10%	**35%**	10%	0%

We are optimistic that LLM-based agents could solve these and much more complex tasks with various improvements such as planning (Huang et al., 2022), chain-of-thought (Wei et al., 2022; Zhang et al., 2023), or fine-tuning (Carta et al., 2023). As the main motivation of this case study was to show that it is easy to study LLMs with the SocialAI school, we leave those experiments for future work.

6. Conclusion

We highlight the importance of socio-cognitive abilities in AI research and present an introduction to Michael Tomasello's and Jerome Bruner's theories of socio-cognitive development. Following these theories, we outlined a set of key socio-cognitive abilities and concepts for AI: social cognition (inferring other's perception and joint attention), communication (referential and early conventionalized communication), cultural learning (imitation and role reversal imitation), scaffolding, and formats.

We present the SocialAI school - a tool simplifying the research of core socio-cognitive abilities. We show how the SocialAI school can be used to easily create environments studying various questions inspired by developmental psychology. With RL agents, we conduct experiments regarding the pointing gesture, scaffolding, and role reversal (by recreating an experiment from developmental psychology). We demonstrate that, by using SocialAI to parse environments into text, Large Language Models be easily studied too. In the appendix, we present additional studies concerning linguistic communication, joint attention, imitation learning, inferring others' field of view, and formats. Our experiments demonstrated the diversity of studies that can be conducted with the SocialAI school, highlighted the limitations of standard RL agents, and showed that while large language models learn with high sample efficiency, additional methods such as fine-tuning or chain-of-thought might be needed.

References

Abramson, J., Ahuja, A., Brussee, A., Carnevale, F., Cassin, M., Fischer, F., Georgiev, P., Goldin, A., Harley, T., Hill, F., Humphreys, P. C., Hung, A., Landon, J., Lillicrap, T. P., Merzic, H., Muldal, A., Santoro, A., Scully, G., von Glehn, T., Wayne, G., Wong, N., Yan, C., and Zhu, R. Creating multimodal interactive agents with imitation and self-supervised learning. *CoRR*, abs/2112.03763, 2021. URL https://arxiv.org/abs/2112.03763.

Andreas, J. Language models as agent models. In *Conference on Empirical Methods in Natural Language Processing*, 2022.

Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Abbeel, O. P., and Zaremba, W. Hindsight experience replay. In *NeurIPS*, 2017.

APA, A. P. A. Theory of mind, 2023. URL https://dictionary.apa.org/theory-of-mind. Accessed: 2023-06-22.

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., and Yoshida, C. Cognitive developmental robotics: A survey. *IEEE Transactions on Autonomous Mental Development*, 1(1):12–34, 2009. doi: 10.1109/TAMD.2009.2021702.

Baker, B., Kanitscheider, I., Markov, T., Wu, Y., Powell, G., McGrew, B., and Mordatch, I. Emergent tool use from multi-agent autocurricula. *ArXiv*, abs/1909.07528, 2019.

Baker, C. L., Saxe, R., and Tenenbaum, J. B. Bayesian theory of mind: Modeling joint belief-desire attribution. In Carlson, L. A., Hölscher, C., and Shipley, T. F. (eds.), *Proceedings of the 33th Annual Meeting of the Cognitive Science Society, CogSci 2011, Boston, Massachusetts, USA, July 20-23, 2011*. cognitivesciencesociety.org, 2011. URL https://mindmodeling.org/cogsci2011/papers/0583/index.html.

Behne, T., Carpenter, M., and Tomasello, M. One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, 8(6):492–499, 2005. doi: https://doi.org/10.1111/j.1467-7687.2005.00440.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-7687.2005.00440.x.

Bellemare, M., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., and Munos, R. Unifying count-based exploration and intrinsic motivation. *Advances in neural information processing systems*, 29, 2016.

Bhoopchand, A., Brownfield, B., Collister, A., Lago, A. D., Edwards, A. D., Everett, R., Frechette, A., Oliveira, Y. G., Hughes, E., Mathewson, K. W., Mendolicchio, P., Pawar, J., Pislar, M., Platonov, A. K., Senter, E., Singh, S., Zacherl, A., and Zhang, L. M. Learning robust real-time cultural transmission without human data. *ArXiv*, abs/2203.00715, 2022.

Billard, A. and Dautenhahn, K. Grounding communication in autonomous robots: An experimental study. *Robotics and Autonomous Systems*, 24(1):71 – 79, 1998. ISSN 0921-8890. doi: https://doi.org/10.1016/S0921-8890(98)00023-2. URL http://www.sciencedirect.com/science/article/pii/S0921889098000232. Scientific Methods in Mobile Robotics.

Billard, A. and Dautenhahn, K. Experiments in learning by imitation - grounding and use of communication in robotic agents. *Adaptive Behavior*, 7(3-4):415–438, 1999. doi: 10.1177/105971239900700311. URL https://doi.org/10.1177/105971239900700311.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901, 2020.

Bruner, J. Child's talk: Learning to use language. *Child Language Teaching and Therapy*, 1(1):111–114, 1985. doi: 10.1177/026565908500100113. URL https://doi.org/10.1177/026565908500100113.

Bruner, J. *Acts of meaning*. Harvard university press, 1990.

Burda, Y., Edwards, H., Storkey, A. J., and Klimov, O. Exploration by random network distillation. *CoRR*, abs/1810.12894, 2018. URL http://arxiv.org/abs/1810.12894.

Butler, L. P. and Tomasello, M. Two- and 3-year-olds integrate linguistic and pedagogical cues in guiding inductive generalization and exploration. *Journal of Experimental Child Psychology*, 145:64–78, 2016. ISSN 0022-0965. doi: https://doi.org/10.1016/j.jecp.2015.12.001. URL https://www.sciencedirect.com/science/article/pii/S0022096515002945.

Cangelosi, A. and Schlesinger, M. *Developmental Robotics: From Babies to Robots*. The MIT Press, 2014. ISBN 0262028018.

Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K., Tani, J., Belpaeme, T., Sandini, G., Nori, F., et al. Integration of action and language knowledge: A roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development*, 2(3):167–195, 2010.

Carpenter, M., Akhtar, N., and Tomasello, M. Fourteen- to 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, 21: 315–330, 1998a. ISSN 0163-6383.

Carpenter, M., Nagell, K., and Tomasello, M. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63 4:i–vi, 1–143, 1998b.

Carpenter, M., Tomasello, M., and Striano, T. Role reversal imitation and language in typically developing infants and children with autism. *Infancy*, 8(3): 253–278, 2005. doi: 10.1207/s15327078in0803_4. URL https://www.tandfonline.com/doi/abs/10.1207/s15327078in0803_4.

Carta, T., Romac, C., Wolf, T., Lamprier, S., Sigaud, O., and Oudeyer, P.-Y. Grounding large language models in interactive environments with online reinforcement learning. 2023.

Carvalho, E. Social affordance. In Vonk, J. and Shackelford, T. (eds.), *Encyclopedia of Animal Cognition and Behavior*, pp. 1–4. 2020.

Chen, L., Lu, K., Rajeswaran, A., Lee, K., Grover, A., Laskin, M., Abbeel, P., Srinivas, A., and Mordatch, I. Decision transformer: Reinforcement learning via sequence modeling. *Advances in neural information processing systems*, 34:15084–15097, 2021.

Chevalier-Boisvert, M., Willems, L., and Pal, S. Minimalistic gridworld environment for openai gym. https://github.com/maximecb/gym-minigrid, 2018.

Chevalier-Boisvert, M., Bahdanau, D., Lahlou, S., Willems, L., Saharia, C., Nguyen, T. H., and Bengio, Y. Babyai: A platform to study the sample efficiency of grounded language learning. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL https://openreview.net/forum?id=rJeXCo0cYX.

Chung, J., Gülçehre, Ç., Cho, K., and Bengio, Y. Gated feedback recurrent neural networks. In Bach, F. R. and Blei, D. 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. 2067–2075. JMLR.org, 2015. URL http://proceedings.mlr.press/v37/chung15.html.

Clark, A. *Being There: Putting Brain, Body, and World Together Again*. MIT Press, Cambridge, MA, USA, 1st edition, 1996. ISBN 0262032406.

De Waal, F. *Are we smart enough to know how smart animals are?* WW Norton & Company, 2016.

D'Entremont, B., Hains, S., and Muir, D. A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development*, 20(4):569–572, 1997. ISSN 0163-6383. doi: https://doi.org/10.1016/S0163-6383(97)90048-5. URL https://www.sciencedirect.com/science/article/pii/S0163638397900485.

Erikson, E. H. *Childhood and society*. WW Norton & Company, 1993.

Fletcher, G. E., Warneken, F., and Tomasello, M. Differences in cognitive processes underlying the collaborative activities of children and chimpanzees. *Cognitive Development*, 27(2):136–153, 2012. ISSN 0885-2014. doi: https://doi.org/10.1016/j.cogdev.2012.02.003. URL https://www.sciencedirect.com/science/article/pii/S0885201412000093.

Gergely, G., Bekkering, H., and Király, I. Rational imitation in preverbal infants. *Nature*, 415(6873): 755–755, February 2002. doi: 10.1038/415755a. URL https://ideas.repec.org/a/nat/nature/v415y2002i6873d10.1038_415755a.html.

Gopnik, A. and Meltzoff, A. N. *Words, thoughts, and theories*. Mit Press, 1997.

Hare, B., Call, J., and Tomasello, M. Do chimpanzees know what conspecifics know? *Animal behaviour*, 61(1): 139–151, 2001.

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., and Tomasello, M. Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, 317 (5843):1360–1366, 2007. doi: 10.1126/science.1146282. URL https://www.science.org/doi/abs/10.1126/science.1146282.

Hester, T., Vecerik, M., Pietquin, O., Lanctot, M., Schaul, T., Piot, B., Horgan, D., Quan, J., Sendonaris, A., Osband, I., et al. Deep q-learning from demonstrations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Heyes, C. Précis of cognitive gadgets: The cultural evolution of thinking. *Behavioral and Brain Sciences*, 42:e169, 2019. doi: 10.1017/S0140525X18002145.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735. URL https://doi.org/10.1162/neco.1997.9.8.1735.

Huang, W., Abbeel, P., Pathak, D., and Mordatch, I. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. *CoRR*, abs/2201.07207, 2022. URL https://arxiv.org/abs/2201.07207.

Hui, D. Y.-T., Chevalier-Boisvert, M., Bahdanau, D., and Bengio, Y. Babyai 1.1, 2020.

Hutchins, E. *Cognition in the Wild (Bradford Books)*. The MIT Press, September 1996. ISBN 0262581469. URL http://www.amazon.ca/exec/obidos/redirect?tag=citeulike04-20{&}path=ASIN/0262581469.

Jaques, N., Lazaridou, A., Hughes, E., Gülçehre, Ç., Ortega, P. A., Strouse, D., Leibo, J. Z., and de Freitas, N. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*, volume 97, pp. 3040–3049. PMLR, 2019. URL http://proceedings.mlr.press/v97/jaques19a.html.

Keupp, S., Behne, T., and Rakoczy, H. Why do children overimitate? normativity is crucial. *Journal of Experimental Child Psychology*, 116(2):392–406, 2013. ISSN 0022-0965. doi: https://doi.org/10.1016/j.jecp.2013.07.002. URL https://www.sciencedirect.com/science/article/pii/S0022096513001379.

Kosinski, M. Theory of mind may have spontaneously emerged in large language models. *arXiv preprint arXiv:2302.02083*, 2023.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25: 1097–1105, 2012.

Kumar, S., Correa, C. G., Dasgupta, I., Marjieh, R., Hu, M., Hawkins, R. D., Daw, N. D., Cohen, J. D., Narasimhan, K., and Griffiths, T. L. Using natural language and program abstractions to instill human inductive biases in machines. *ArXiv*, abs/2205.11558, 2022.

Le, M., Boureau, Y.-L., and Nickel, M. Revisiting the evaluation of theory of mind through question answering. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5872–5877, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1598. URL https://aclanthology.org/D19-1598.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

Lee, D., Jaques, N., Kew, J. C., Eck, D., Schuurmans, D., and Faust, A. Joint attention for multi-agent coordination and social learning. *CoRR*, abs/2104.07750, 2021. URL https://arxiv.org/abs/2104.07750.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. In *ICLR*, 2016.

Lindblom, J. and Ziemke, T. Social situatedness of natural and artificial intelligence: Vygotsky and beyond. *Adaptive Behavior*, 11(2):79–96, 2003. doi: 10.1177/10597123030112002. URL https://doi.org/10.1177/10597123030112002.

Lyons, D. E., Young, A. G., and Keil, F. C. The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, 104:19751 – 19756, 2007.

Meltzoff, A. N. Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental psychology*, 31 5:838–850, 1995.

Meltzoff, A. N. and Moore, M. K. Explaining facial imitation: A theoretical model. *Infant and child development*, 6(3-4):179–192, 1997.

Mirolli, M. and Parisi, D. Towards a vygotskyan cognitive robotics: The role of language as a cognitive tool. *New Ideas in Psychology*, 29 (3):298–311, 2011. ISSN 0732-118X. doi: https://doi.org/10.1016/j.newideapsych.2009.07.001. URL https://www.sciencedirect.com/science/article/pii/S0732118X09000348. Special Issue: Cognitive Robotics and Reevaluation of Piaget Concept of Egocentrism.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540): 529, 2015.

Moll, H. and Tomasello, M. Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, 24(3):603–613, 2006.

Nair, A., McGrew, B., Andrychowicz, M., Zaremba, W., and Abbeel, P. Overcoming exploration in reinforcement learning with demonstrations. In *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 6292–6299. IEEE, 2018.

Ndousse, K. K., Eck, D., Levine, S., and Jaques, N. Emergent social learning via multi-agent reinforcement learning. In *International Conference on Machine Learning*, pp. 7991–8004. PMLR, 2021.

Netanyahu, A., Shu, T., Katz, B., Barbu, A., and Tenenbaum, J. B. PHASE: physically-grounded abstract social events for machine social perception. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pp. 845–853. AAAI Press, 2021. URL https://ojs.aaai.org/index.php/AAAI/article/view/16167.

Oudeyer, P.-Y. and Kaplan, F. What is intrinsic motivation? a typology of computational approaches. *Frontiers in Neurorobotics*, 1, 2007. ISSN 1662-5218. doi: 10.3389/neuro.12.006.2007. URL https://www.frontiersin.org/articles/10.3389/neuro.12.006.2007.

Over, H. and Carpenter, M. The social side of imitation. *Child Development Perspectives*, 7(1): 6–11, 2013. doi: https://doi.org/10.1111/cdep.12006. URL https://srcd.onlinelibrary.wiley.com/doi/abs/10.1111/cdep.12006.

Parker-Holder, J., Jiang, M., Dennis, M., Samvelyan, M., Foerster, J. N., Grefenstette, E., and Rocktaschel, T. Evolving curricula with regret-based environment design. In *International Conference on Machine Learning*, 2022.

Pathak, D., Agrawal, P., Efros, A. A., and Darrell, T. Curiosity-driven exploration by self-supervised prediction. In *ICML*, 2017.

Perez, E., Strub, F., de Vries, H., Dumoulin, V., and Courville, A. C. Film: Visual reasoning with a general conditioning layer. *CoRR*, abs/1709.07871, 2017. URL http://arxiv.org/abs/1709.07871.

Portelas, R., Colas, C., Weng, L., Hofmann, K., and Oudeyer, P. Automatic curriculum learning for deep RL: A short survey. *CoRR*, abs/2003.04664, 2020. URL https://arxiv.org/abs/2003.04664.

Rabinowitz, N. C., Perbet, F., Song, H. F., Zhang, C., Eslami, S. M. A., and Botvinick, M. Machine theory of mind. In Dy, J. G. and Krause, A. (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.

Raileanu, R. and Rocktäschel, T. RIDE: rewarding impact-driven exploration for procedurally-generated environments. *CoRR*, abs/2002.12292, 2020. URL https://arxiv.org/abs/2002.12292.

Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S. G., Novikov, A., Barth-Maron, G., Gimenez, M., Sulsky, Y., Kay, J., Springenberg, J. T., Eccles, T., Bruce, J., Razavi, A., Edwards, A. D., Heess, N. M. O., Chen, Y., Hadsell, R., Vinyals, O., Bordbar, M., and de Freitas, N. A generalist agent. *ArXiv*, abs/2205.06175, 2022.

Richerson, P. J. and Boyd, R. *Not by Genes Alone: How Culture Transformed Human Evolution*. University Of Chicago Press, June 2006. ISBN 0226712125. URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0226712125.

Ruis, L., Khan, A., Biderman, S. R., Hooker, S., Rocktaschel, T., and Grefenstette, E. Large language models are not zero-shot communicators. *ArXiv*, abs/2210.14986, 2022.

Sap, M., Rashkin, H., Chen, D., Bras, R. L., and Choi, Y. Socialiqa: Commonsense reasoning about social interactions. *CoRR*, abs/1904.09728, 2019. URL http://arxiv.org/abs/1904.09728.

Sap, M., Bras, R. L., Fried, D., and Choi, Y. Neural theory-of-mind? on the limits of social intelligence in large lms. *ArXiv*, abs/2210.13312, 2022.

Savinov, N., Raichuk, A., Marinier, R., Vincent, D., Pollefeys, M., Lillicrap, T. P., and Gelly, S. Episodic curiosity through reachability. *ArXiv*, abs/1810.02274, 2018. URL http://arxiv.org/abs/1810.02274.

Scao, T. L., Fan, A., Akiki, C., Pavlick, E.-J., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., Gallé, M., Tow, J., Rush, A. M., Biderman, S. R., Webson, A., Ammanamanchi, P. S., Wang, T., Sagot, B., ..., and Wolf, T. Bloom: A 176b-parameter open-access multilingual language model. *ArXiv*, abs/2211.05100, 2022.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal policy optimization algorithms. *ArXiv*, abs/1707.06347, 2017. URL http://arxiv.org/abs/1707.06347.

Shu, T., Kryven, M., Ullman, T. D., and Tenenbaum, J. Adventures in flatland: Perceiving social interactions under physical dynamics. In Denison, S., Mack, M., Xu, Y., and Armstrong, B. C. (eds.), *Proceedings of the 42th Annual Meeting of the Cognitive Science Society - Developing a Mind: Learning in Humans, Animals, and Machines, CogSci 2020, virtual,* *July 29 - August 1, 2020*. cognitivesciencesociety.org, 2020. URL https://cogsci.mindmodeling.org/2020/papers/0723/index.html.

Shu, T., Bhandwaldar, A., Gan, C., Smith, K. A., Liu, S., Gutfreund, D., Spelke, E. S., Tenenbaum, J. B., and Ullman, T. D. AGENT: A benchmark for core psychological reasoning. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9614–9625. PMLR, 2021. URL http://proceedings.mlr.press/v139/shu21a.html.

Siposova, B. and Carpenter, M. A new look at joint attention and common knowledge. *Cognition*, 189:260–274, 2019.

Tang, H., Houthooft, R., Foote, D., Stooke, A., Chen, X., Duan, Y., Schulman, J., Turck, F. D., and Abbeel, P. Exploration: A study of count-based exploration for deep reinforcement learning, 2017.

Tennie, C., Call, J., and Tomasello, M. Push or pull: Imitation vs. emulation in great apes and human children. *Ethology*, 112(12):1159–1169, 2006.

Tennie, C., Walter, V., Gampe, A., Carpenter, M., and Tomasello, M. Limitations to the cultural ratchet effect in young children. *Journal of experimental child psychology*, 126:152–160, 2014.

Tomasello, M. *The Cultural Origins of Human Cognition*. Harvard University Press, 1999. ISBN 9780674000704. URL http://www.jstor.org/stable/j.ctvjsf4jc.

Tomasello, M. Becoming human. In *Becoming Human*. Harvard University Press, 2019.

Tomasello, M. The role of roles in uniquely human cognition and sociality. *Journal for the Theory of Social Behaviour*, 50(1):2–19, 2020. doi: 10.1111/jtsb.12223.

Tomasello, M., Kruger, A. C., and Ratner, H. H. Cultural learning. *Behavioral and brain sciences*, 16(3):495–511, 1993.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

Trott, S., Jones, C. J., Chang, T. A., Michaelov, J. A., and Bergen, B. K. Do large language models know what humans know? *ArXiv*, abs/2209.01515, 2022.

Vollmer, A.-L., Wrede, B., Rohlfing, K. J., and Oudeyer, P.-Y. Pragmatic frames for teaching and learning in human–robot interaction: Review and challenges. *Frontiers in Neurorobotics*, 10:10, 2016. ISSN 1662-5218. doi: 10.3389/fnbot.2016.00010. URL https://www.frontiersin.org/article/10.3389/fnbot.2016.00010.

Vygotsky, L. S. and Cole, M. *Mind in society : the development of higher psychological processes*. Harvard University Press Cambridge, 1978. ISBN 0674576284 0674576292.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E. H., Le, Q., and Zhou, D. Chain of thought prompting elicits reasoning in large language models. *CoRR*, abs/2201.11903, 2022. URL https://arxiv.org/abs/2201.11903.

Whiten, A., Horner, V., Litchfield, C. A., and Marshall-Pescini, S. How do apes ape? *Animal Learning & Behavior*, 32:36–52, 2004.

Whiten, A., McGuigan, N., Marshall-Pescini, S., and Hopper, L. M. Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528):2417–2428, 2009.

Wood, D., Bruner, J. S., and Ross, G. The role of tutoring in problem solving. *Journal of child psychology and psychiatry, and allied disciplines*, 17(2):89–100, 1976. doi: 10.1111/j.1469-7610.1976.tb00381.x.

Wood, D., Bornstein, M., and Bruner, J. Interaction in human development, 1989.

Wyman, E., Rakoczy, H., and Tomasello, M. Normativity and context in young children's pretend play. *Cognitive development*, 24(2):146–155, 2009.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., and Cao, Y. React: Synergizing reasoning and acting in language models. *ArXiv*, abs/2210.03629, 2022.

Zhang, T., Xu, H., Wang, X., Wu, Y., Keutzer, K., Gonzalez, J. E., and Tian, Y. Bebold: Exploration beyond the boundary of explored regions. *CoRR*, abs/2012.08621, 2020. URL https://arxiv.org/abs/2012.08621.

Zhang, Z., Zhang, A., Li, M., Zhao, H., Karypis, G., and Smola, A. J. Multimodal chain-of-thought reasoning in language models. *ArXiv*, abs/2302.00923, 2023.