Should artificial agents ask for help in human-robot collaborative problem-solving?

Adrien Bennetot, Vicky Charisi, Natalia Díaz-Rodríguez

Should artificial agents ask for help in human-robot collaborative problem-solving?

Adrien Bennetot[1,3], Vicky Charisi[2] and Natalia Díaz-Rodríguez[1]

***Abstract*— Transferring as fast as possible the functioning of our brain to artificial intelligence is an ambitious goal that would help advance the state of the art in AI and robotics. It is in this perspective that we propose to start from hypotheses derived from an empirical study in a human-robot interaction and to verify if they are validated in the same way for children as for a basic reinforcement learning algorithm. Thus, we check whether receiving help from an expert when solving a simple close-ended task (the Towers of Hanoï) allows to accelerate or not the learning of this task, depending on whether the intervention is canonical or requested by the player. Our experiences have allowed us to conclude that, whether requested or not, a Q-learning algorithm benefits in the same way from expert help as children do.**

I. INTRODUCTION

Developmental robotics [39] (and synonyms *cognitive developmental robotics*, *autonomous mental development* as well as *epigenetic robotics* [10]) is the interdisciplinary approach to the autonomous design of behavioural and cognitive capabilities in artificial agents that directly draws inspiration from developmental principles and mechanisms observed in children's natural cognitive systems [10], [39].

Autonomous agents in such settings learn in an open-ended [21] manner, where crucial components of such developmental approach consist of learning the ability to autonomously generate goals and explore the environment, exploiting intrinsic motivation [43] and computational models of curiosity [42], [37].

II. RELATED WORK

A. Development and learning in human child

The development of the executive functions (EF) in human infants and young children with rudimentary neurodevelopment of prefrontal cortex (PFC) refers to an array of organizing and self-regulating goal-directed behaviors that inhibit impulses and regulate behaviour from a very early age. These developments have been associated with both the PFC maturation and its connectivity with other brain areas [25] which is enabled by the individual's sustained interaction with the surrounding physical and social environment [55]. The initiation of these sensorimotor interactions in young children are exploratory in nature and often are embedded in playful activities with components of motor learning [1], [27]. Visual stimuli are also responsible for the elicitation of improved EF and cognitive organization which contributes to the development of perceptual learning.

Although exposure to visual stimuli can lead to perceptual learning, it is often insufficient to yield robust learning [40]. Research shows that additional factors, such as attention and reinforcement are needed to produce robust learning. Amount of exposure, strength of exposure, relation to attention, interactions of multiple sensory systems in perceptual learning are some of the factors that promote human learning; the underlying brain mechanisms that relate to these factors are among the most active targets of research into the complex mechanisms of child's learning and the association of EF development with visuo-motor integration [40].

In relation to the above-mentioned mechanisms, research has shown the interaction of memory and learning with mechanisms such as curiosity, appraisal, prediction and exploration [28], [31]. Gruber's PACE framework [31] suggests that curiosity is triggered by significant prediction errors that are appraised. This enhances memory which is encoded through increased attention, exploration and information seeking and contributes to the consolidation of information acquired while in a curious state through dopaminergic neuromodulation of the hippocampus. More on the dopamine neuromodulator from the intrinsic and extrinsic reward perspective of RL is in [51].

From a behavioural perspective, exploration has been previously identified as a special form of curiosity that refers to a drive that is either intrinsic or extrinsic [38]. Active experimentation with physical objects generates more accurate inferences about the latent properties of the object than passive observation [8]. Exploration of the physical world is considered as a phase in human transition from behavioural events towards symbolic and conceptual thinking. The developmental process of symbol and concept emergence has been associated to the relative frequency in which certain strategies are used and to the process of abandoning an old strategy and discovering new ones [50].

In problem-solving tasks these mechanisms have been correlated with child's ability to inhibit a certain action while considering an alternative one that would be more appropriate for the optimal performance of a task [6]. The developmental process that leads from sensori-motor events to abstract learning and the acquisition of the optimal strategy

[1] U2IS, ENSTA, Institut Polytechnique Paris, Inria FLOWERS team, Palaiseau, France. {adrien.bennetot, natalia.diaz}@ensta-paris.fr

[2] European Commission, JRC, Centre for Advanced Studies, Seville, Spain. Vasiliki.Charisi@ec.europa.eu

[3] Segula Technologies, Parc d'activité de Pissaloup - Trappes, France

for a specific task can be measured by behavioural indicators such as task performance speed and accuracy level. [19]. However, this process appears more complex in the case of collaborative problem-solving where the child interacts with a more knowledgeable social agent. This includes the process of selective social learning and relies on child's social motivation aspects for learning [34].

Research shows that humans have the ability to explicitly communicate their uncertainty to others at a very early stage of their life. Infants are capable of monitoring and communicating their own uncertainty non verbally to gain knowledge from others [29]. While playing in unstructured and uncertain environments that lack clear extrinsic reward signals, they actively seek help from other humans. In early childhood, however, children might be aware of their uncertainty, but they do not proceed always to help-seeking [57] which shows the complexity of extrinsic and intrinsic motivation.

In this complex context, the examination of the learning outcome often is not adequate for the understanding of children's problem-solving activity. An emphasis on how children move from early to later levels of competence within an EF component allows the depiction of their developmental trajectories [50], [18], [6]. A mapping of the developmental trajectories reveals inter-individual differences in cognitive mechanisms such as inhibition of prepotent responses, mental shifting [26] and generalization [3]. These changes have been associated with changing brain connectivity which is considered as both cause and consequence of the developmental changes [52]. An additional input towards the understanding of child's developmental process comes from the field of child-robot interaction in which the child can take advantage of the robot's appropriate interventions.

B. Child development inspired artificial agent learning

Child learning has vastly inspired how to build learning machines [35]. A sample of cognitive architecture to teach robots in the way infants learn is in [33], demonstrating how exploiting sensitivity to sensorimotor contingencies/ affordances in developmental psychology, combined with the notion of goal allows an agent to develop new sensorimotor skills in open ended learning settings [21], [20]. An example of new discovered contingency is, e.g., touching a bell to generate a sound.

Inspired by developmental psychology, in [22] interactive learning (active imitation learning and goal-babbling) is combined with autonomous exploration in a strategic learner to reuse previously learned tasks or "procedures" in a *Socially Guided Intrinsic Motivation with Procedure Babbling* (SGIM-PB) able to determine the representation of a hierarchy of interrelated tasks. In hard-exploration games, novelty seeking agents [14], curiosity meta-learning [2] and remembering promising states and exploring from them [24] are powerful approaches to learn artificial agents.

Essential robotics scenarios for open-ended learning making use of brain inspired models are Long-Term Memory for Artificial Cognition [23], for robots to learn to operate in different worlds under different goals when the occurrence of experiences is intertwined. In this context, a Baxter robot demonstrates to learn control tasks, segmenting the world into semantically loaded categories associated with contexts, that in order, can allow higher level reasoning and planning. Architectures for lifelong learning by evolution in robots are MDB (Multilevel Darwinist Brain) [5], [4].

Some of the modulation based mechanism embedded within a cognitive architecture for robots combine long-term memory and a motivational system in order to select candidate primitive value functions for transfer and adaptation to new situations through modulatory ANNs. These progressively conform new parameterized value functions able to address more complex situations in a developmental manner in a Baxter robot, which must solve different tasks in a cooking setup [46], or simplify the utility space in continuous state spaces [47].

Charisi et. al. [12] take inspiration from inhibitory control in developmental psychology and examine child-robot collaborative problem-solving with a focus on the process rather than the outcome of child's acquisition of a certain strategy. The task of Tower of Hanoï is used to study the initiation of voluntary request for help in a child-robot interaction setting with child-initiated robot interventions. They observe children's trajectories of problem-solving and the needs for exploratory actions. We extend this work [12] to test if robotics learning processes and agent learning from an expert can be child-development inspired. Since their analysis of when and why asking for help helps solving collaborative tasks in inhibitory processes, in this paper we contrast the hypotheses tested in kids with those mimicking the same situations in an artificial agent learning to solve the same task, with reinforcement learning [53].

III. METHODOLOGY

As in [12] we are evaluating the learning agent (LA) on the Tower of Hanoï game, but instead of the LA being a child, our agent is a Q-learning algorithm [58] with a learning rate $\alpha = 1$, a discount factor $\gamma = 0.8$ and an exploration $\epsilon = 0.05$. As it can be seen on Fig. 3 in the Appendix, the Tower of Hanoï game with 3 disks is a simple close-ended task with 27 possible states and, at most, 3 possible actions associated to each state. Each element of the reward matrix used for the Q-learning represents the reward from moving from the current state to the next one. Moves leading to the goal state are assigned a reward of 100, illegal moves a reward of $-\infty$ and others a reward of 0.

A. Hypotheses

In order to explore if algorithms benefit from asking for help in human-robot collaborative problem-solving, in the same manner as kids do, we further formulate two hypotheses:

- *H1*: Canonical interventions from an expert speed up learning.
- *H2*: Getting help *on demand* from an expert accelerates finding the optimal solution compared to not *on demand*.

B. Research Design

We manipulate the expert intervention with 2 different scenarios:

- The LA1 solves the task in collaboration with the expert in a "turn-taking" scenario, which results in a canonical cognitive intervention by the expert.
- The LA2 solves the task independently, having the option to ask for help of the expert whenever (if) this is needed, which results in an *on demand* intervention by the expert.

In order to test the different variations among teacher-driven and learner-driven interaction [16] in our HRI setting, we vary two main parameters:

- The **canonical intervention rate**, i.e. the frequency of the expert's intervention during the canonical scenario.
- The **ask-for-help** parameter, i.e. how much the LA asks the expert to do the next movement, as a proxy to simulate the needs for help, during the *on demand* scenario.

Our evaluation metric is the number of movements required to solve the task after a variable number of training episodes. To make these results robust, all the experiments were repeated 100 times.

IV. RESULTS

We used the above-mentioned parameters to test our hypotheses as follows.

A. Task Performance with and without Turn-Taking

The first configuration consists of a LA1, a Q-learning agent, playing in collaboration with an expert that knows exactly what is the optimal movement in each configuration. Every two turns, the expert will play instead of the LA1 and perform the optimal action. We compare this with the performance of the LA1 when it solves the task alone, and with the one of a random policy.

As it can be seen in Fig. 1, the LA1 is directly more efficient when it is helped by the expert in a turn taking scenario, going from an order of 10^2 moves to solve the task without help without training, to 10^1 with canonical interventions. This can be explained by the fact that the agent is directly placed by the expert on the optimal sequence of actions (the left side diagonal from Fig. 3) to solve the task. In fact the expert is able to solve the task in 7 moves starting from any state, so after it has played, the LA1 is necessarily only 6 moves away from victory rather than 7. Thus during the first episodes of Q-learning, when the LA1 is not yet aware of the optimal path and acts somewhat randomly, it

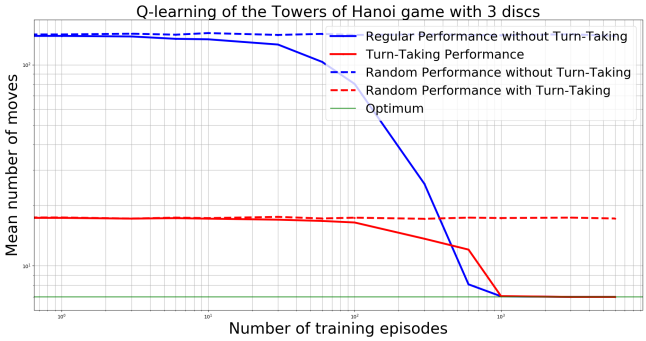

Fig. 1. Canonical intervention scenario: the LA1 solves the task in collaboration with the expert as they play alternatively. In cyan and yellow the performance when the LA1 follows a random policy, with and without help. Log. scale used on both axes.

is still closer to the resolution of the task when it receives help than when it does not, because in the worst case it would be 6 moves away from the resolution instead of 7. In other words, the help of an expert improves the performance of the random policy. However the LA1 is moving away from the random policy after only 10 episodes when it does not receive help and it takes 100 training episodes to start drastically reducing its mean number of moves. At the same time, the performance still seems to be random in the turn-taking configuration and it takes to the agent 300 training episodes before it starts to converge to the optimal solution. The curves intersect after 400 training episodes when the LA1 without help starts to outperform the helped LA1. The LA1 needs 3,000 episodes of training to reach the optimal solution with canonical interventions, whereas it only needs 1,000 episodes when it is not helped. We can therefore conclude that being helped every 2 rounds by an expert agent does not speed up the learning process, on the contrary it slows it down.

This is somehow not really surprising because the expert giving the optimal solution every two rounds prevents the agent from exploring every possible state. As shown in Fig. 3, the objective is to reach the *222* state at the bottom left and each move of the expert will therefore lead the game to a state further to the left or further down than the previous state. This makes some states hard to reach (such as *121*) or even impossible (such as those below the *223*), thus delaying the convergence towards the optimal solution as the agent will still waste time trying to get in there even if it is not possible. This is a drawback of the learning system used. In contrast to some state-of-the-art methods such as Policy Shaping [30], [11], our Learning Agent is guided by an expert user and the feedback is not formulated as policy advice, as the goal is not to optimize the human feedback but to mimic how a kid learns solve the task with a Learning Agent with a Q-learning algorithm, instead of a child as in the settings of [12]. The learning system could be improved by optimizing the teaching [9] by not always giving the

optimal action but the one that will teach the agent the most.

A solution that would not deviate from the initial experimental setup could therefore be to let the LA1 explore the different states by involving the expert less frequently, by modifying the canonical intervention rate. This is what we did in Fig. 4 in the Appendix, letting the expert play every 3 and 4 turns.

B. On demand or canonical intervention by the expert

The second configuration consists of a LA2, a Q-learning agent, trying to solve the task independently. It has the opportunity to ask for help to an expert agent whenever it needs to. To do this, we added an *ask for help* parameter to the Q-learning. At each turn, if the best policy value is lower than the *ask for help* parameter, the expert will play instead of the LA2. As we can see on Fig. 2, the LA2 is directly more effective, because he is always asking for help as it does not know yet what do to. After asking for help many times during the first 10 episodes it starts solving the task by itself, resulting in a loss of efficiency. We interpret this as the LA2 gaining confidence in movements which, while not perfect, still allows the task to progress towards its resolution through state exploration and trial. Compared to the LA1 without help, the LA2 asking for help is much more efficient but there is not a lot of variation between the canonical and the *ask for help* configuration. This is probably due to the rather simple simulation of the ask for help trigger.

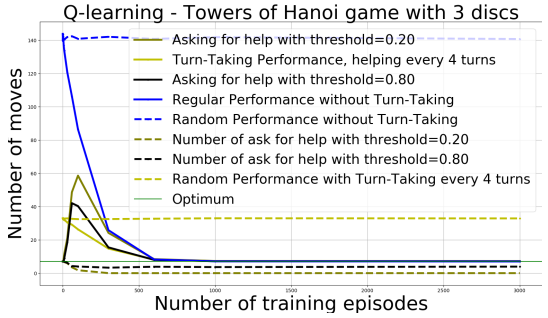

Fig. 2. Asking for help scenario with different *ask for help* values: the LA2 tries to solve the game alone while being able to ask for help whenever its best action is not good enough (plot not on logarithmic scale as the agent asks for help at most 7 times).

V. DISCUSSION AND FUTURE WORK

This paper presents the initial work towards the understanding of problem-solving process with two artificial agents by simulating a child-robot interaction experimental study. We acknowledge that the simulation of child's behaviour is a complex task and more emphasis is needed on accurate description of the multidimensional child behaviour.

The aspect of intrinsic motivation [43], indirectly related with solving a concrete task, but concerned with learning a set of reusable skills, should be further studied when rising the level of abstraction, specially in the context of solving different tasks and taking as input larger state spaces and of larger dimensionality [45], [44] in order to simplify problem-solving in an end-to-end learning manner. State representation learning [36] may come into use for a more realistic, less preprocessing demanding setting, i.e., not requiring human annotations of each game state when involving human collaboration.

Our point is to verify if hypotheses derived from an empirical study in an HRI setting are valid when translated from children to a RL Agent. Thus, the difficulty lies in the simulation of the child's behaviour by an artificial agent. The addition of an intrinsic motivation, on the desire for the LA to solve the game by itself, could increase the accuracy of this comparison. Our LA asks for help when it considers that a movement is not good enough to be played (i.e. when the largest Q-value among all available states fall under a pre-set threshold), whereas in reality, the mechanisms pushing the child to ask for help are much more complex [35], [7].

One of the challenges to explore is to validate the hypotheses tested with more complex tasks. More elaborated manners should be devised to more faithfully model uncertainty in the agent while acting. Future work could better mimic the presented and other human learning inspired behaviours. For instance, one could quantify (aleatoric and epistemic) uncertainty [54], [13], [15] of the agent's next action so we can better simulate the *ask for help* setting when an agent is not certain enough. An accurate assessment should be made of the mechanisms that lead a child to ask for help when solving a task independently. This would allow it to be represented in the LA's behaviour so that it could ask for help in a more human-like natural way.

Future work includes the expansion of collaborative problem-solving settings with triadic interactions e.g. two children and a robot, in order to examine features of collective problem solving accounting for social dynamics [56]. In addition to this, we are planning to examine the shifting processes [41], i.e. the processes of strategy generalization in a different task in human and artificial agents. Future work could also consider the possibility of trading-off between the gain generated for the agent by asking versus the disruption it causes to the human, using principles of mixed initiative interaction [32]. CoBots approaches[1] to ask for help are a related field to further explore, e.g., planning approaches for the LA to distinguish actions that it can complete autonomously from those that it cannot [49], [48].

Finally, in order to better understand child's developmental trajectories, we aim to replicate a similar child-robot interaction setting with a larger sample by manipulating additional variables such as the agent's social behaviour. This would inform our testing of more complex algorithms than Q-learning, using other dopamine based distributional RL signals [17], and as little training data as people need [35].

[1]CoBots http://www.cs.cmu.edu/~coral/projects/cobot/

VI. ACKNOWLEDGEMENT

We thank Cristina Conati for giving feedback on this work.

REFERENCES

[1] K. E. Adolph and J. E. Hoch. Motor development: Embodied, embedded, enculturated, and enabling. *Annual review of psychology*, 70:141–164, 2019.

[2] F. Alet, M. F. Schneider, T. Lozano-Perez, and L. P. Kaelbling. Meta-learning curiosity algorithms. *arXiv preprint arXiv:2003.05325*, 2020.

[3] S. M. Barnett and S. J. Ceci. When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, 128(4):612, 2002.

[4] F. Bellas, J. Becerra, and R. Duro. Using promoters and functional introns in genetic algorithms for neuroevolutionary learning in non-stationary problems. *Neurocomputing*, 72(10):2134 – 2145, 2009. Lattice Computing and Natural Computing (JCIS 2007) / Neural Networks in Intelligent Systems Designn (ISDA 2007).

[5] F. Bellas, G. Varela, and R. J. Duro. A Cognitive Developmental Robotics Architecture for Lifelong Learning by Evolution in Real Robots. 2010.

[6] J. R. Best and P. H. Miller. A developmental perspective on executive function. *Child development*, 81(6):1641–1660, 2010.

[7] M. Biehl, C. Guckelsberger, C. Salge, S. C. Smith, and D. Polani. Expanding the active inference landscape: More intrinsic motivations in the perception-action loop. *Frontiers in Neurorobotics*, 12:45, 2018.

[8] N. R. Bramley, T. Gerstenberg, J. B. Tenenbaum, and T. M. Gureckis. Intuitive experimentation in the physical world. *Cognitive psychology*, 105:9–38, 2018.

[9] M. Cakmak and M. Lopes. Algorithmic and human teaching of sequential decision tasks. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, AAAI'12, page 1536–1542. AAAI Press, 2012.

[10] A. Cangelosi and M. Schlesinger. From babies to robots: The contribution of developmental robotics to developmental psychology. *Child Development Perspectives*, 2018.

[11] T. Cederborg, I. Grover, C. L. Isbell, and A. L. Thomaz. Policy shaping with human teachers. In *IJCAI*, 2015.

[12] V. Charisi, E. Gomez, G. Mier, L. Merino, and R. Gomez. Child-robot collaborative problem-solving and the importance of child's voluntary interaction: A developmental perspective. *Frontiers in Robotics and AI*, 7:15, 2020.

[13] W. R. Clements, B.-M. Robaglia, B. Van Delft, R. B. Slaoui, and S. Toth. Estimating risk and uncertainty in deep reinforcement learning. *arXiv preprint arXiv:1905.09638*, 2019.

[14] E. Conti, V. Madhavan, F. P. Such, J. Lehman, K. Stanley, and J. Clune. Improving exploration in evolution strategies for deep reinforcement learning via a population of novelty-seeking agents. In *Advances in neural information processing systems*, pages 5027–5038, 2018.

[15] F. L. Da Silva, P. Hernandez-Leal, B. Kartal, and M. E. Taylor. Uncertainty-aware action advising for deep reinforcement learning agents. 2019.

[16] F. L. da Silva, G. Warnell, A. H. R. Costa, and P. Stone. Agents teaching agents: a survey on inter-agent transfer learning. *Autonomous Agents and Multi-Agent Systems*, 34:1–17, 2019.

[17] W. Dabney, Z. Kurth-Nelson, N. Uchida, C. K. Starkweather, D. Hassabis, R. Munos, and M. Botvinick. A distributional code for value in dopamine-based reinforcement learning. *Nature*, pages 1–5, 2020.

[18] D. DeMarie-Dreblow and P. H. Miller. The development of children's strategies for selective attention: Evidence for a transitional period. *Child Development*, pages 1504–1513, 1988.

[19] A. Diamond. Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. *Principles of frontal lobe function*, pages 466–503, 2002.

[20] S. Doncieux, N. Bredeche, L. L. Goff, B. Girard, A. Coninx, O. Sigaud, M. Khamassi, N. Díaz-Rodríguez, D. Filliat, T. Hospedales, A. Eiben, and R. Duro. DREAM Architecture: a Developmental Approach to Open-Ended Learning in Robotics, 2020.

[21] S. Doncieux, D. Filliat, N. Díaz-Rodríguez, T. Hospedales, R. Duro, A. Coninx, D. M. Roijers, B. Girard, N. Perrin, and O. Sigaud. Open-ended learning: a conceptual framework based on representational redescription. *Frontiers in Neurorobotics*, 2018.

[22] N. Duminy, S. M. Nguyen, and D. Duhaut. Learning a set of interrelated tasks by using a succession of motor policies for a socially guided intrinsically motivated learner. *Frontiers in Neurorobotics*, 12:87, 2019.

[23] R. J. Duro, J. A. Becerra, J. Monroy, and F. Bellas. Perceptual generalization and context in a network memory inspired long-term memory for artificial cognition. *International Journal of Neural Systems*, 29(06):1850053, 2019. PMID: 30614325.

[24] A. Ecoffet, J. Huizinga, J. Lehman, K. O. Stanley, and J. Clune. First return then explore. *arXiv preprint arXiv:2004.12919*, 2020.

[25] A. Fiske and K. Holmboe. Neural substrates of early executive function development. *Developmental Review*, 52:42–62, 2019.

[26] N. P. Friedman and A. Miyake. Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, 86:186–204, 2017.

[27] E. J. Gibson. Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual review of psychology*, 39(1):1–42, 1988.

[28] J. Gottlieb and P.-Y. Oudeyer. Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, 19(12):758–770, 2018.

[29] L. Goupil, M. Romand-Monnier, and S. Kouider. Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences*, 113(13):3492–3496, 2016.

[30] S. Griffith, K. Subramanian, J. Scholz, C. Isbell, and A. Thomaz. Policy shaping: Integrating human feedback with reinforcement learning. *Advances in Neural Information Processing Systems*, 01 2013.

[31] M. J. Gruber and C. Ranganath. How curiosity enhances hippocampus-dependent Memory: The prediction, appraisal, curiosity, and exploration (PACE) framework. *Trends in cognitive sciences*, 2019.

[32] E. Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, page 159–166, New York, NY, USA, 1999. Association for Computing Machinery.

[33] L. Jacquey, G. Baldassarre, V. G. Santucci, and J. K. O'Regan. Sensorimotor contingencies as a key drive of development: From babies to robots. *Frontiers in Neurorobotics*, 13:98, 2019.

[34] M. A. Koenig and M. A. Sabbagh. Selective social learning: New perspectives on learning from others. *Developmental Psychology*, 49(3):399, 2013.

[35] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. Building machines that learn and think like people. *Behavioral and brain sciences*, 40, 2017.

[36] T. Lesort, N. Díaz-Rodríguez, J.-F. Goudou, and D. Filliat. State representation learning for control: An overview. *Neural Networks*, 108:379 – 392, 2018.

[37] T. Lesort, V. Lomonaco, A. Stoian, D. Maltoni, D. Filliat, and N. Díaz-Rodríguez. Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. *Information Fusion*, 58:52 – 68, 2020.

[38] G. Loewenstein. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, 116(1):75, 1994.

[39] M. Lungarella, G. Metta, R. Pfeifer, and G. Sandini. Developmental robotics: a survey. *Connection Science*, 15(4):151–190, 2003.

[40] M. M. McClelland and C. E. Cameron. Developing together: the role of executive function and motor skills in children's early academic lives. *Early Childhood Research Quarterly*, 46:142–151, 2019.

[41] A. Miyake, N. P. Friedman, M. J. Emerson, A. H. Witzki, A. Howerter, and T. D. Wager. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology*, 41(1):49–100, 2000.

[42] P. Oudeyer. Computational theories of curiosity-driven learning. *CoRR*, abs/1802.10546, 2018.

[43] P.-Y. Oudeyer, F. Kaplan, and V. Hafner. Intrinsic motivation systems for autonomous mental development. *Evolutionary Computation, IEEE Transactions on*, 11(2):265–286, April 2007.

[44] A. Prieto, A. Romero, F. Bellas, R. Salgado, and R. J. Duro. Introducing separable utility regions in a motivational engine for cognitive developmental robotics. *Integrated Computer-Aided Engineering*, 26(1):3–20, 2019.

[45] A. Romero, J. A. Becerra, F. Bellas, and R. J. Duro. Modulation based transfer learning of motivational cues in developmental robotics. In

2019 International Joint Conference on Neural Networks (IJCNN), pages 1–8, July 2019.

[46] A. Romero, F. Bellas, J. A. Becerra, and R. J. Duro. Producing parameterized value functions through modulation for cognitive developmental robots. In M. F. Silva, J. Luís Lima, L. P. Reis, A. Sanfeliu, and D. Tardioli, editors, *Robot 2019: Fourth Iberian Robotics Conference*, pages 250–262, Cham, 2020. Springer International Publishing.

[47] A. Romero, A. Prieto, F. Bellas, and R. J. Duro. Simplifying the creation and management of utility models in continuous domains for cognitive robotics. *Neurocomputing*, 353:106–118, 2019.

[48] S. Rosenthal, M. Veloso, and A. K. Dey. Is someone in this office available to help me? *Journal of Intelligent & Robotic Systems*, 66(1-2):205–221, 2012.

[49] S. Rosenthal, M. M. Veloso, and A. K. Dey. Task behavior and interaction planning for a mobile service robot that occasionally requires help. In *Automated Action Planning for Autonomous Mobile Robots*, 2011.

[50] R. S. Siegler. Microgenetic analyses of learning. *Handbook of child psychology*, 2, 2007.

[51] S. P. Singh, A. G. Barto, and N. Chentanez. Intrinsically motivated reinforcement learning. In *Advances in neural information processing systems*, pages 1281–1288, 2005.

[52] L. Smith, L. Byrge, and O. Sporns. Beyond origins. developmental pathways and the dynamics of brain networks. In A. J. Lerner, S. Cullen, and S.-J. Leslie, editors, *Current Controversies in Philosophy of Cognitive Science*, pages 49–62. Routledge, 2020.

[53] R. S. Sutton. *Introduction to reinforcement learning*, volume 135. 1998.

[54] N. Tagasovska and D. Lopez-Paz. Single-model uncertainties for deep learning. In *Advances in Neural Information Processing Systems*, pages 6414–6425, 2019.

[55] F. J. Varela, E. Thompson, and E. Rosch. *The embodied mind: Cognitive science and human experience*. 2016.

[56] S. Wallkotter, S. Tulli, G. Castellano, A. Paiva, and M. Chetouani. Explainable agents through social cues: A review. *arXiv preprint arXiv:2003.05251*, 2020.

[57] A. M. Was and F. Warneken. Proactive help-seeking: Preschoolers know when they need help, but do not always ask for it. *Cognitive Development*, 43:91–105, 2017.

[58] C. Watkins and P. Dayan. Technical note: Q-learning. *Machine Learning*, 8:279–292, 05 1992.

APPENDIX

A. Tower of Hanoï game

All possible states of the Tower of Hanoï game are in Fig. 3.

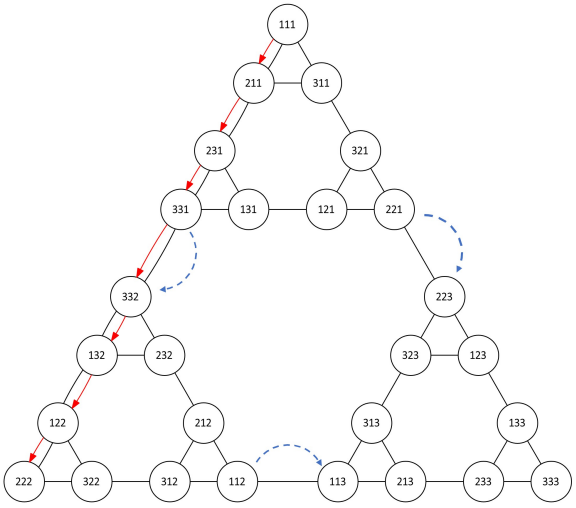

Fig. 3. Abstract graph of the Tower of Hanoï game states for 3 disks. Each node represents one possible state of the game. The starting configuration is on the top, the final one in the bottom left. In red the optimal sequence of actions, in blue dashed movements between sub-graphs leading toward the solution (retrieved from [12].)

B. Additional Results

The LA1 who receives help is, regardless of the number of training episodes, always more efficient than the one who does not receive help. We can therefore conclude that an agent is more efficient when it receives help, as long as this help does not block its exploration.

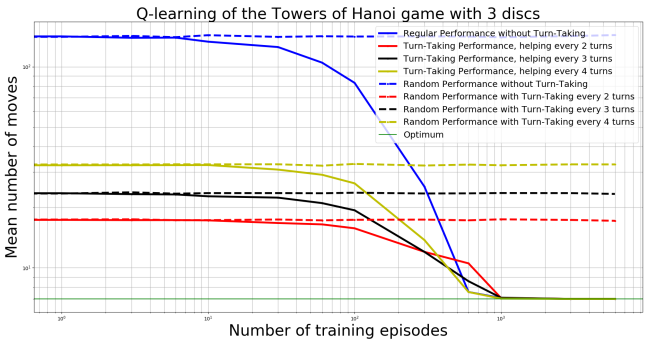

Fig. 4. Canonical intervention scenario with different intervention rates: the LA1 and the expert solve the game in collaboration but the expert is only playing every 2, 3 and 4 turns. It means that, e.g., in the last configuration, the LA1 will play 3 times before the expert plays. Log. scale used on both axes.