Workshops at 18th International Conference on Intelligent Environments (IE2022)
H.H. Alvarez Valera and M. Luštrek (Eds.)

doi:10.3233/AISE220050

Self-Development and Causality in Intelligent Environments

Matteo MARTINELLI [a,1], Stefano MARIANI [a], Marco LIPPI [a] and Franco ZAMBONELLI [a]

[a] *University of Modena and Reggio Emilia, Reggio Emilia, Italy*

Abstract. Future intelligent environments will operate in dynamic and unpredictable situations. Thus, they will have to be able to dynamically learn how to act, interact, and adapt, with little or no a priori knowledge and without human intervention. That is, such systems should become able to self-develop causal models of themselves and of the environment in which they act (i.e., what their actions imply and what actions induce what effects on the environment), and of their social relationships (i.e,, what interactions induce what impact on other systems). In this paper, we introduce key concepts of self-development in intelligent environments, both at the individual and collective level, by framing its key concepts and its relation with causal models. Then, we introduce two case studies, focus of our current (preliminary) experiments. Finally, we discuss related work and some key research challenges.

Keywords. Self-development, causal models, learning, self-adaptation, coordination

1. Introduction

Infants, since their early weeks, start experiencing with their own body, moving hands, touching objects, and interacting with people around. Such activities are part of an overall process of *self-development* (aka autonomous development), which lets them gradually develop cognitive and behavioural capabilities [1]. These skills include the capability to recognise situations around, the sense of self, the so called “sense of agency” (i.e., understanding the causal effect of own actions in an environment), which subsequently enable the capability to act purposefully towards a goal, and some social capabilities (i.e., knowing how to act in the presence of others). Building machines capable to replicate these learning mechanisms is increasingly recognised as a key challenge in many areas of artificial intelligence (AI), such as robotics [2], autonomous vehicles management [3], and of course intelligent IoT systems and environments [4].

For small-scale and rather static environments, it is possible to “hardwire” a model of the environment within a system, alongside some pre-designed plans of action to achieve specific and well-defined goals. However, for larger and dynamic environments, and for more complex tasks, individual components of the system should be able to autonomously (i.e., without human supervision and with little or no innate knowledge): *(i)*

[1] Corresponding Author: Matteo Martinelli, DISMI – University of Modena and Reggio Emilia, Via Giovanni Amendola, 2 42122 Reggio Emilia, (RE), Italy; E-mail: matteo.martinelli@unimore.it.

build environmental models and continuously update them as situations evolve; *(ii)* develop the capability of recognising and modelling causal relations, and specifically the causal effect of their own actions on the environment (which variables of the environment can or cannot be directly affected by which actuators, which variables and actuators relate to each other); *(iii)* learn to achieve goals on this basis and depending on the current situation; *(iv)* learn how to organise and coordinate actions among multiple distributed components whenever necessary (e.g., by learning that a desired effect on the environment can be caused only by cooperating with others).

Building systems enriched with fully-fledged self-development capabilities is definitely an ambitious objective. Yet, the idea per se is not new, and its worthiness has been already advocated since several years [5]. However, the topic is now even more timely. Many recent research results in areas such as unsupervised and self-learning, causal analysis, multi-agent systems, and collective behaviours, have started shedding light on the various mechanisms that have to be involved in the overall process of self-development, hinting at the fact that the vision (at least in specific application areas) is close to become reality. Furthermore, unfolding the key concepts and mechanisms underlying self-development can also somewhat contribute to understand the many mental mechanisms behind artificial general intelligence [6].

In this context, the contribution of this paper is to frame the key concepts behind self-development and its relation with causality. Specifically, we first introduce a general conceptual framework for the continuous and adaptive process of self-development, both at the individual and at the collective level. Then, we illustrate two representative case studies, subject of some preliminary experiments we have performed toward self-development. Finally, we discuss related works and identify some key research challenges to be attacked.

2. Conceptual Framework

Self-development, besides being the process that infants carry out during the early stages of their life [1], may also involve any "agent" whenever it is incarnated in a new body and immersed in a new environment. As an example to quickly and intuitively introduce our general framework (Figure 1) let us consider what we do whenever we start playing a new video-game. At first, we observe the game environment and the commands (potential causes) available; that is, we get acknowledged with our *embodiment and perception* on the video-game. We try the commands to assess their effects; that is, we try to acquire a *sense of agency*, i.e., a causal model of ourselves in the environment. Then, we understand what is the goal of the game and how we can use the sense of agency to achieve it; that is, we start acting in a *goal-oriented* way.

Typically, we recognise the presence of other "agents", virtual characters that are not under our control; that is, we distinguish between *self and non-self*. This implies that we acknowledge that we should act also in dependence of the actions of other agents (*strategic thinking*). All this process is typically repeated in a cyclic way (i.e., when reaching a new level in the game) to adapt to new environments, new situations, new tools available, new goals, and new enemies.

In the case of multiplayer games, besides recognising the presence of other players, and the need to act also accounting for them, we should understand: whether we have

Figure 1. The conceptual framework of self-development.

communication tools available and how to use them to affect and influence the actions of others, i.e., to *coordinate* with them, so that eventually *institutional* ways to act together towards a goal can be established. Again, this process may be cyclically repeated as the game advances. Truly intelligent and adaptive ICT systems should undergo a similar process and autonomously develop through similar phases.

2.1. The Individual Level

Let us consider a single agent X (purely software or physically embodied) immersed in a (virtual or physical) environment. The agent can observe a set of environmental variables $\mathscr{V} = \{v_1, v_2, \ldots, v_m\}$. As the agent is part of the environment, internal variables of the agent itself (i.e., its current status and configuration) are included in the set. In addition, the agent has a set of possible actions $\mathscr{A} = \{a_0, \ldots, a_{n-1}, null\}$, including the *null* action.

Embodiment and Perception. In this early phase, the agent should autonomously recognise the existence of $\mathscr{A}$ and $\mathscr{V}$, that is, it should get acknowledged to its actuation (causation) and sensorial capabilities. Without resorting to complex AI techniques, methods from the reflective and self-adaptive programming systems can effectively apply in this phase [7] to let the agent dynamically self-inspect its capabilities and start analysing the observed variables. The agent can also start acquiring some understanding of the relations between the observed variables over time, as well as some simple prediction skills.

Sense of Agency. In this exploratory phase, the agent starts trying to understand what are the effects of $\mathscr{A}$ on $\mathscr{V}$, by trying to apply actions (even without any goal in mind) to see their effects. That is, it will eventually recognise that, given the *current* state $\mathscr{V}_{curr}$, the application of an action a_i (or of a sequence of actions) will eventually lead (with some probability) to state $\mathscr{V}_{next}$. This mechanism enables the construction of the basic sense of agency [1], and of the sense of causality.

Goal-orientedness. In this exploitation phase, the agent starts applying $\mathscr{A}$ with goals in mind. That is, given the current state $\mathscr{V}_{curr}$ and a desired future state $\mathscr{V}_g$ (the goal, aka the desired "state of the affairs"), the agent exploits its sense of agency by applying the action that can possibly cause the environment to move to state $\mathscr{V}_g$. This implies achieving the capability of planning the required sequence of actions to achieve a specific goal.

Self and Non-Self. As soon as an agent starts exploring its own actions $\mathscr{A}$, and recognizes that they have effect on the environment, it also understands that there are effects that are not under its own control. That is, there are "non-self" entities acting in the environment, too. By learning how to apply $\mathscr{A}$, the agent also learns the limits of such actions because of the non-self entities affecting some variable v_i.

Strategic Thinking. The agent has built a causal model of the world, that is, of how $\mathscr{A}$ affects $\mathscr{V}$, and it starts somehow including in such model the models of others (non-self) [8] while acting and while designing strategies. That is, it can recognize that there are goals it can possibly (or hopefully) attain only by accounting for the actions of others.

As in the videogame example, self-development is not to be conceived as a "once-and-for-all" process. Rather, it is a continuous, never-ending process: environmental conditions can change, new sensors may become available, and new actions become feasible (or vice versa, some sensors and actions may no longer be available). This requires the agents to re-tune their learned causal sense of agency, and re-think how to achieve goals in isolation and in the presence of non-self entities.

In the related work section, we discuss how individual self-development roughly correspond to the Pearl's "ladder of causation" [9].

2.2. The Collective Level

In the presence of multiple agents acting in the same environment, an agent recognises that there are goals that cannot be achieved in isolation or by simply applying strategic thinking. Thus as part of their self-development, they should collectively develop some forms of "autonomous social engagement". Formally, this corresponds to considering a set of K agents $X_0,\ldots,X_{K-1}$, where *(i)* each agent can choose the actions to perform from its own set (either disjoint or partially overlapping with those of the others); *(ii)* not necessarily all the agents can observe the whole set of environmental variables, but more likely each agent X_j can perceive and/or control a subset of them. Thus, for specific goals $\mathscr{V}_g$ to be achieved, there is the need of properly combining and sequencing actions by different agents, e.g. X_i executes a^i_w whereas X_j executes a^j_z, and so on.

Communication. To overcome the limitations of strategic thinking, agents should be provided with a specific set of *communication actions*, i.e., actions that are devoted to cause a change in the actions of others. These could take the form of explicit communication acts, i.e., messages, that the agent should learn how to receive and send as an additional – social – form of perception and action. However, they could also take the form of implicit actions aimed at affecting the behaviour of others, i.e., leaving signs in the environment (stigmergy) or adopting peculiar behaviours aimed at being noticed by others (behavioural implicit communication) [10]. All these cases can be formalized by including in the $\mathscr{A}$ set communication actions, and in $\mathscr{V}$ observable signs in the environment.

Coordination. By exploring their available communication actions, agents start understanding how such acts can be used to get access and to affect some of the variables of the environment, and in particular those that are not observable and controllable by themselves. For instance, they can learn how to use communication acts to get access to the value of some non-observable variables v_i or to direct other agents in executing the actions that can affect its values as required for a goal to be achieved. In other words, such explorations enable learning basic forms of coordination, which can be thought of as a social form for the sense of agency establishing causal relationships amongst agents issuing and receiving communications.

Institution. Eventually, after exploring coordination protocols, the agents can "institutionalise" their patterns of interaction towards collective actions. That is, they will learn those acceptable social patterns of coordination, and the set of social norms and social incentives that enables them to systematically achieve goals together [11]. Formally, this corresponds to having agents in the collective recognise and adhere to a set of constraints $\mathscr{C}(\mathscr{A},\mathscr{V})$ ruling the way (communication) actions $\mathscr{A}$ can be performed in specific conditions $\mathscr{V}$, as well as the commitments and expectations (causes) motivating agents' compliance to the communication protocol (effect).

As for the case of the individual level, the dynamics of the environment or of the agent population may require the above process to assume a continuous cyclic nature. We emphasise that communication, coordination, and institutions are not strictly necessary to promote complex goal-oriented collective actions [12]. Nevertheless, whenever communication mechanisms are available, learning to exploit them is a natural part of the self-development process, and can facilitate collective action.

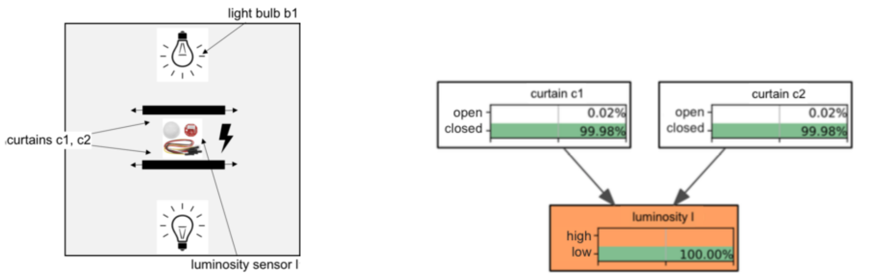
3. Case Studies

There are diverse application scenarios that can potentially take advantage of systems capable of self-development. In the following, two case studies will be presented. The first one has been also physically implemented with a proof-of-concept framework, representing a smart home scenario, where the application of the described technology demonstrates its potential. The second one, only theoretically designed, illustrates the possible benefits that such solutions can bring to an industrial environment.

3.1. Smart Homes

Smart homes can facilitate our interaction with the environment and increase our safety and comfort. We envision that once a new home is built, its smart devices could start exploring their own individual and collective capabilities, so as to eventually learn how they can causally affect the environment, and apply such capabilities once users will start populating it. This will also require to continuously adapt to habits and preferences of users, accommodate new devices and services, tolerate partial failures. Our preliminary experience suggests the feasibility of the vision [4].

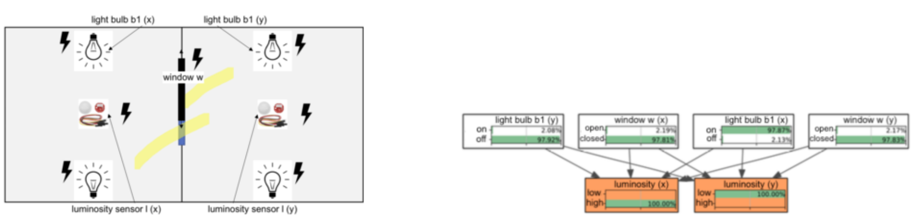
We designed and performed several experiments, following the steps outlined in our proposed self-development framework. The experiments focus on both the individual and collective level there described. We hereby summarise the two most relevant experiments, while the reader can refer to [4] for a comprehensive description. We employ a Bayesian Network to learn dependencies between the observed variables. We remark that Bayesian Networks are not causal models, since edges only represent conditional dependencies between nodes. Nevertheless, in the considered domain, one can reasonably assume that the connections between the actions performed by the home actuators and the relevant effects on the sensors can be considered not only as a correlation, but as a cause-effect relation. Clearly, this is a simplifying assumption that will need to be refined in more complex scenarios.



(a) (b)

Figure 2. *Learning sense of agency*. We show the smart home setup (a) and the associated Bayesian network for learning sense of agency (b). The lightning bolt represents the actuator controlled by the room agent. The room eventually learns the actions to be taken (cause) to achieve the goal "complete darkness" (effect).

- *Learning sense of agency*. This experiment applies at the individual (i.e., single-room) level. The goal is to assess whether the agent of a single room is able to learn the effect of an action (sense of agency), which is the necessary precondition for learning goal-orientedness. The experiment has been performed using a room luminosity sensor *l*, two light bulbs *b1* and *b2* which cannot be controlled by the agent, and two controllable curtains *c1* and *c2*, positioned to hide the light of the related light bulb when in *closed* position. During training, the agent continuously performs actions in order to move the curtains and to understand the relationship between its actions and the sensed environment. Once the learning phase is completed, the agent is eventually able to understand what to do to reach the state where the light in the room is low. In other words, it has acquired a sense of agency, according to our conceptual framework.
- *Learning to coordinate*. This experiment applies at the collective level (i.e., with two connected rooms). In this setup, two agents are involved, each monitoring its room. Rooms are connected by a common window, so that the light of sensors in one room also affects the other. The window can be controlled by either agent. By learning from the joint set of observations, the two agents learn that they need to cooperate in order to reach a common state of the affairs: for example, closing the window together in order to obtain the dark. In terms of our conceptual framework, they need to coordinate actions.



(a) (b)

Figure 3. *Learning to coordinate*. We show the smart home setup (a) and the associated Bayesian network for learning to coordinate (b). The window can be actuated only through joint action of both room agents, which have to agree on the desired state of the affairs. Agents eventually learn how to cooperate.

The described experiments show how using Bayesian Networks merged with explorative actions it is possible to pave the way for systems that autonomously learn cause-effects relations, making a step towards the development of an individual and collective sense of agency (coordination).

3.2. Smart Manufacturing

Manufacturing operations in Industry 4.0 also represent an ideal scenario to model and evaluate our framework of self-development. Let us take as an example two manufacturing nodes called A and B, as depicted in Figure 4. A single process carried out in node A takes 1 batch of base modules and 3 batches of 3 kind of components to be attached on it. Then, the processed base modules are taken from node A and moved to node B where it is assembled with another component. The base modules are moved by an automated guided vehicle (AGV) while the components are moved by human picking operators. To illustrate the importance of self-development within this kind of domain, we exemplify two possible situations.

Causality. Consider the following scenario: a workstation places a request to an AGV. Sometimes it may occur that the request is not fulfilled in the expected time. Slowly the delivery performance may start to deviate from its average, causing more and more delays. In a traditional scenario, time delays are only tracked at workstations: if the personnel is well trained and disciplined, the only information available at the given workstation is that a logistic delay occurred. In the opposite scenario – for example with new operators not trained or sufficiently skilled – the information would be not so reliable. Therefore, the workstation performance tracking is only the starting point for the root cause investigation, which may involve not only the aforementioned workstation, but also other workstations and equipment. With the help of a self-organised and autonomous system empowered by causal relations capabilities, the workstation may compute the causal relation with all the entities that interacted with them and that may have affected its performance in the last time span. If the strongest causal relation for its delays is with the AGV, then the workstation may share the information with the AGV, asking for a diagnosis check. The AGV may then find an unexpected value produced by a wheel sensor which, in turn, may suggest a broken bearing forcing the AGV to move slowly. The AGV may then also share this information with other workstations, that might be affected by the same delays as well. The root cause analysis would thus result much faster and effective.

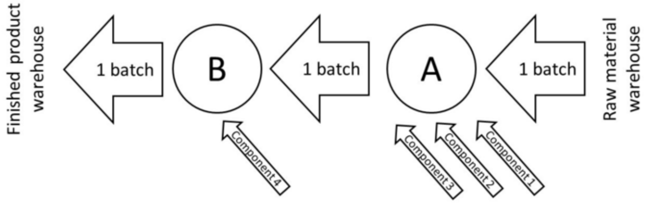


Figure 4. The considered production setting scheme in the smart manufacturing scenario.

Coordination. Following the planning control system, coordinating and orchestrating a set of heterogeneous resources to complete the manufacturing process, implementing the planned actions as well as respecting the time constraints are all hard tasks. The involved entities are: the raw material warehouse; the finished product warehouse; the AGV transport system; the picking logistic operators; processing nodes A and B. When a production order is released from the production plan, processing node A has to perform the requested tasks. To do so, it has to check whether the needed raw material is present in the workstation, whether the workstation is equipped with the relevant tools to complete the task, whether the tools and machines are in a healthy state. Let us assume that the material is not present: to retrieve it, the AGV and the human logistic picker have to be involved. Therefore, the node has to choose the best AGV to ask for retrieving the base modules, and the best operator for retrieving components. In parallel, a task to an operator is scheduled, to perform the machine setup as required by the order planning. While the setup is in process, the AGV and the operator move to the raw material warehouse asking the warehouse entity to release the requested material. Therefore, the warehouse releases the material in the correct quantities and then the AGV and the operator come back to the node to feed it as requested. Then, the manufacturing tasks start.

Generally, buffers for raw materials and finished goods are not infinite, so the logistic requests will be rolling, being released when a buffer is getting empty with a lead time equal or close to consumption time of raw material. In the same way, the output buffer has to be emptied when the batch is full, in order to accommodate a new empty batch ready to accept other processed products without stopping the process. Hence, the logistic service has to be called again in a rolling fashion to move the finished product of node A (that is, the base module raw materials for node B) from the output buffer, and, ideally, directly to node B, that has to be almost ready with all the preparatory tasks previously described. In the middle of all those processes, reaction to unexpected situations is a key factor for maintaining the flow of processing material tight with minimum wastes. Eventually, node B should be able to coordinate again effectively with the logistic services in order to free its output buffers before a block occurs, letting the logistics move the finished material to the finished product warehouse.

The described vision is very ambitious and simply describes the need (and potential) of self-organising individual and collecting system empowered by causality models in industrial environments. Expecting to hardwire all the possibilities in a reliable and effective codebase is a very hard task even with only the involvement of a few resources, as depicted. As the number of resources grows, the complexity of the system and the number of possible unexpected situations will inevitably grow as well.

4. Related Works

The idea of self- or autonomous development, at both the individual and collective level, has been widely investigated in areas such as cognitive psychology, neuroscience, philosophy, and ethics [5]. We hereby focus on the computational perspective, reporting related works that can contribute to unfold the mechanisms involved in the self-development vision and to eventually realise it.

The area of reinforcement learning shares with our vision the objective of self-training machines to act to achieve a specific goal, given a specific context. The best in class approach is deep Q-Learning [13], which, however, does not aim at building a system with a causal sense of agency and an interpretable world model. Approaches based on *intrinsic* rewards [14], instead, more closely exploit the idea of exploring the world to develop a sense of agency. In fact, intrinsic rewards are developed by the agent itself to satisfy its curiosity (i.e., when it discovers how to achieve specific tasks) [15]. However, causality is typically still out of the picture in most researches.

Curriculum-based approaches to machine learning go somewhat in the direction of gradually developing the capability to act in complex scenarios [16]. In the same vein, recent approaches based on the theory of affordances [17] propose to have agents gradually learn the effects of their actions. With this approach, they eventually develop an explicit sense of agency, i.e., a model of how their actions affect the environment that is interpretable under causal lenses.

The issue of understanding and leveraging *causality* is increasingly recognized as a key challenge for AI [18]. In particular, Judea Pearl [9] has proposed the idea of a "causal hierarchy" (also named "ladder of causation") to define different levels of causality recognition and exploitation by an intelligent agent. Such ladder corresponds to some of the phases of the self-development loop we defined: the first one is mostly involved in the perception phase, whereas the second one is associated to the development of a sense of agency and to recognition of self and non-self. The final layer clearly enables goal-oriented behavior, strategic thinking, and collective coordination. Bayesian and causal networks are among the models that are most widely exploited in order to build interpretable models of the world. A recent contribution is the application of curriculum learning to the problem of learning the structure of Bayesian networks [19]. On a pure sub-symbolic level, on the other hand, another recent work proposes to learn causal models in an online setting [20], with the aim to find (and strengthen) causal links between variables.

Recently, autocurricula-based approaches have produced stunning results in multiagent environments, both cooperative and competitive, like in the hide-and-seek scenario [21]). However, providing agents with an *explicit* modeling (possibly in *causal* terms) of others' behavior and overall societal behavior, may be necessary [8]. Also, autocurricula approaches do not currently account for the possibility of explicitly interacting (e.g., through speech acts) with other agents, which may be indeed fundamental to improve collective learning.

5. Discussion and Open Challenges

The general vision of self-development is still far to be reality. Several ideas in the areas of unsupervised and self-learning, causality, multi-agent systems, are already showing its potential feasibility and applicability, at least in specific application areas. However, many open challenges are still to be faced:

- Although our general vision ideally assumes that all the knowledge is self-acquired by a system, real-world system may require least some innate knowledge about the world. We think that the identification of the most proper and effective trade-off between innate and acquired knowledge will be a key challenge.

- Can we control the evolution of behaviours during the self-development process? We argue that an explicit meta-level modelling causality underlying system evolution will be necessary to support such a controlled evolution.
- The more technologies based on self-development will pervade our everyday environments, the more humans will have to continuously interact with them. This interaction will raise technical issues (will we have “handles” to control or block such systems?) and ethical problems (will we be rather “handled” by these systems and subjected to their decisions?).

References

[1] Rochat P. Self-perception and action in infancy. Experimental brain research. 1998;123(1-2):102-9.

[2] Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. Building machines that learn and think like people. Behavioral and brain sciences. 2017;40.

[3] Yang Y, Taylor M, Luo J, Wen Y, Slumbers O, Graves D, et al. Diverse Auto-Curriculum is Critical for Successful Real-World Multiagent Learning Systems. In: 20th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS; 2021. .

[4] Lippi M, Mariani S, Zambonelli F. Developing a Sense of Agency in IoT Systems: Preliminary Experiments in a Smart Home Scenario. In: 17th CoMoRea workshop at PerCom. IEEE; 2021. .

[5] Weng J, McClelland J, Pentland A, Sporns O, Stockman I, Sur M, et al. Autonomous Mental Development by Robots and Animals. Science. 2001;291(5504):599-600.

[6] Marcus G, Davis E. Insights for AI from the Human Mind. Commun ACM. 2021 Dec;64(1):38–41.

[7] Salehie M, Tahvildari L. Self-adaptive software: Landscape and research challenges. ACM transactions on autonomous and adaptive systems. 2009;4(2):1-42.

[8] Subagdja B, Tan AH. Beyond autonomy: The self and life of social agents. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems; 2019. p. 1654-8.

[9] Pearl J. The Seven Tools of Causal Inference, with Reflections on Machine Learning. Commun ACM. 2019 Feb;62(3):54-60.

[10] Mariani S, Omicini A. Anticipatory Coordination in Socio-Technical Knowledge-Intensive Environments: Behavioural Implicit Communication in MoK. In: AI*IA 2015. Springer; 2015. p. 102-15.

[11] Morris-Martin A, De Vos M, Padget J. Norm emergence in multiagent systems: a viewpoint paper. Autonomous Agents and Multi-Agent Systems. 2019;33(6):706-49.

[12] Leibo JZ, Hughes E, Lanctot M, Graepel T. Autocurricula and the emergence of innovation from social interaction: A manifesto for multi-agent intelligence research. arXiv preprint arXiv:190300742. 2019.

[13] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529-33.

[14] Schmidhuber J. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). IEEE Transactions on Autonomous Mental Development. 2010;2(3):230-47.

[15] Burda Y, Edwards H, Pathak D, Storkey A, Darrell T, Efros AA. Large-Scale Study of Curiosity-Driven Learning. In: International Conference on Learning Representations; 2018. .

[16] Bengio Y, Louradour J, Collobert R, Weston J. Curriculum learning. In: Proceedings of the 26th International Conference on Machine Learning; 2009. p. 41-8.

[17] Khetarpal K, Ahmed Z, Comanici G, Abel D, Precup D. What can I do here? A Theory of Affordances in Reinforcement Learning. In: International Conference on Machine Learning; 2020. p. 5243-53.

[18] Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, Goyal A, et al. Toward Causal Representation Learning. Proceedings of the IEEE. 2021.

[19] Zhao Y, Chen Y, Tu K, Tian J. Learning Bayesian network structures under incremental construction curricula. Neurocomputing. 2017;258:30-40.

[20] Javed K, White M, Bengio Y. Learning causal models online. arXiv preprint arXiv:200607461. 2020.

[21] Baker B, Kanitscheider I, Markov T, Wu Y, Powell G, McGrew B, et al. Emergent Tool Use From Multi-Agent Autocurricula. arXiv preprint arXiv:190907528. 2020.