Object Empowerment-Driven Tool Selection in Reinforcement Learning

Faizan Rasheed
Adaptive Systems Research Group
University of Hertfordshire
Hatfield, UK
f.rasheed@herts.ac.uk

Daniel Polani
Adaptive Systems Research Group
University of Hertfordshire
Hatfield, UK
d.polani@herts.ac.uk

Kenzo Clauw
Adaptive Systems Research Group
University of Hertfordshire
Hatfield, UK
k.clauw@herts.ac.uk

Nicola Catenacci Volpi
Adaptive Systems Research Group
University of Hertfordshire
Hatfield, UK
n.catenacci-volpi@herts.ac.uk

***Abstract*—Tool use is a hallmark of intelligent behavior, both in animals and humans, enabling problem-solving beyond immediate sensorimotor capabilities. Yet, the cognitive and computational mechanisms that give rise to and govern effective tool use remain only partially understood in both cognitive science and artificial intelligence. Discovering and mastering tool use presents significant challenges for learning systems, because it involves delayed rewards and multi-step behaviors whose benefits are not immediately obvious. This has prompted calls for additional intrinsic drives or biases that can can guide learning systems toward the discovery of complex skills like tool use. In this paper, we investigate how artificial agents can successfully learn to use tools to interact with the environment by optimizing object-centric intrinsic motivations, specifically *object empowerment*. Object empowerment measures an agent's potential influence over specific objects of the environment and provides a grounded signal for discovering functional tool-object relationships. Through reinforcement learning (RL) experiments in grid world-like environments featuring multiple tools and objects, we demonstrate that object empowerment facilitates effective tool selection, and supports generalization to multi-object tasks. Moreover, it reveals the ability of tools to exert a never-ending influence over an object and the range of their interaction. Finally, we show that agents guided by object empowerment learn more efficiently in sparse reward conditions than vanilla RL agents.**

***Index Terms*—tool use, intrinsic motivation, empowerment, reinforcement learning.**

I. Introduction

Tool use is widely recognized as a hallmark of intelligent behavior across humans and other animals, reflecting advanced cognitive abilities such as planning, foresight, and causal reasoning [1]–[3]. From a developmental and learning perspective, the emergence of tool use presents a substantial challenge, as it often requires mastering multi-step action sequences with delayed and non-obvious benefits. Such behaviors demand the ability to anticipate future states, maintain intermediate goals, and coordinate actions over extended horizons. In both biological and artificial agents, these challenges have motivated interest in intrinsic motivational mechanisms [4], [39] that can drive exploration, support the discovery of functional affordances, and facilitate the acquisition of complex skills like tool use. Tools, by their very nature, extend an agent's embodiment and its control over its surroundings, enabling interactions that would otherwise be impossible [2], [3]. For these reasons, empowerment [6], [7], an information-theoretic measure that quantifies the potential of an agent to influence its environment, emerges as a natural intrinsic motivation (IM) signal to assess and quantify the impact of tools on their affordances.

In this regard, in [8] the concept of *object empowerment* was introduced as an extension of classical empowerment, quantifying an agent's influence over a specific object rather than over the entire environment. The study focused on scenarios with a single tool and a single object, where object empowerment was shown to facilitate the learning of effective object manipulation strategies for goal-directed behavior. However, in environments with *multiple* tools and objects, the challenge lies in evaluating the degree of control each tool affords over each object and in identifying the most effective tool-object combinations for the task at hand. The present study extends the object empowerment formalism to accommodate environments with multiple tools and objects, reflecting the ecological complexity of real-world settings. This enables the definition of multi-object empowerment, which generalizes the framework across multiple objects, and an empowerment-based tool selection mechanism.

In addition to determining which tools to use, another challenge for organisms is deciding where to use them. For instance, some tools only become effective when the agent is near a task-relevant object (e.g., chimpanzees use stones to crack nuts only near specific nut trees [9]), while others can act on objects from a distance (e.g., a remote control). In more complex scenarios, one may need to reason about the

A preliminary version of this work was presented as a poster at the CoLLAs 2025 Workshop Track and EWRL 2025. These were non-archival presentations without published proceedings.

downstream effects of tool use. The interaction with an object can depend on the satisfaction of specific preconditions or intermediate sub-goals, such as manipulating an object that lies behind a locked door. A tool may be essential for addressing such intermediary tasks (e.g., using a key to open the door), even when it is not directly involved in achieving the agent's primary objective, making the context of tool use harder to identify. Interestingly, this state-dependence of tool utility is mirrored in the nature of object empowerment, which is a function of the agent's state and can provide the necessary spatial context for meaningful tool-object interactions. As we will show, object empowerment landscapes can guide agents in discovering where a tool exerts maximum influence over its target object. Another crucial dimension of tool use is the temporal evolution of tool-object interactions. Some tools afford repeated or persistent transformations, while others have a one-time effect and then lose control over the object. For instance, unlocking a door with a key preserves long-term interaction possibilities (e.g., the door can be re-locked), whereas breaking the door eliminates future interactions. We will show that object empowerment enables the characterization tools in terms of the temporal extent to which they can continue interacting meaningfully with an object.

In what follows we will model an artificial agent, its environment and the interactions between the tools and the objects within it, using the Reinforcement Learning (RL) formalism [10]. A well-known limitation of RL is its reliance on extrinsic rewards, which are often sparse or delayed in real-world and ecological contexts [11]. This relates to a similar challenge when studying the emergence of tool use, where meaningful feedback may occur only after long sequences of interactions. By showing that with the guidance of object empowerment as an IM signal, RL agents can autonomously discover effective tool-use strategies we believe we can shed a light on the process that evolved tool use in the animal kingdom, suggesting that it may have been driven by an analogous optimization principle.

II. Related Work

Tool use has long fascinated researchers across fields such as cognitive science, developmental psychology, and robotics, each aiming to understand how autonomous agents, biological or artificial, interact adaptively with their environment. In cognitive science, studies of primates and birds have revealed spontaneous tool use, often interpreted as evidence of planning, foresight, and embodied intelligence [12]. In developmental psychology, studies have shown that infants develop tool-use skills through exploration and interaction with their environment, highlighting the role of sensorimotor experiences in cognitive development [13]. In robotics, research has focused on replicating tooluse behaviors in artificial agents, enabling them to perceive affordances and execute complex manipulation tasks. Studies have shown how robotic systems can autonomously identify and utilize tools to achieve specific goals, emphasizing the importance of environmental interaction and adaptive learning [14]. Across these fields, a common theme is that tool use emerges through interaction, exploration, and learned affordances, principles that are fundamental in RL. Our work builds upon these ideas by studying how RL agent can develop tool-use capabilities in environments where multiple tools and objects interact dynamically.

The formal modeling of tools use and learning is the object of a number of studies. In [15] a probabilistic framework is proposed that captures a triadic relationship between tools, actions, and their effects using Bayesian networks, enabling inference over tool-action-effect combinations. This is related to our focus on the number of the tool-object interaction's outcomes (i.e., tool to object empowerment), which we use to aid exploration in learning. Likewise, [16] demonstrates how tool-use capabilities can emerge from behavior-grounded exploration, where tools are represented through the effects they produce when manipulated by the agent. Other studies leverage RL to learn tool-use skills in manipulation tasks [17], in particular [18] use an auxiliary reward to optimize resource constraints. Closer to our motivation-driven framework, [19] develop a developmental robotics model where tool use emerges from IMs and planning.

The study of intrinsic drives that enable the discovery of functional affordances is embedded in the broader theory of guided self-organization, which frames evolution and cognition as emergent processes driven by internal information gradients, rather than purely external supervision or reward [20]. One influential line of work views empowerment, as a foundational drive guiding adaptive behavior in both natural and artificial agents [6], [7]. In contrast to intrinsic motivations like curiosity [21], [22], which are mostly focused in measuring novelty in learning, empowerment is an information theoretic measure of how much control an agent has on its environment. Empowerment can be understood as a biologically plausible mechanism that shapes exploration and facilitates the emergence of intelligent behavior, including, we conjecture, tool use. In computational settings, empowerment has been applied in robotic control and RL to encourage structured exploration and skill acquisition, particularly in environments with sparse or delayed rewards [23], [24]. Recent work by [25] shows how empowerment can support open-ended skill discovery in tool-rich environments, further highlighting its relevance for tool-use scenarios.

III. Methodology

A. Tool Learning Framework

We model tool use within a RL framework [10], where an autonomous agent learns to handle tools by manipulating objects in its surrounding. The environment is represented as a Markov Decision Process (MDP), defined by a quadruple $(\mathcal{S}, \mathcal{A}, T, R)$. Here, $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, T is the transition function and R is the reward function. The agent aims to find policy that maximizes the expected return $\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$, where $\gamma \in [0, 1)$ is a discount factor that prioritizes immediate rewards over distant ones.

In RL, agents struggle to associate specific actions with long-term beneficial outcomes due to the lack of reward signal.

To overcome this, the concept of IM has been introduced, inspired by theories of developmental learning in psychology [4], [39]. Rather than relying solely on task-related (extrinsic) rewards, IMs provide internal signals that encourage agents to explore in an informed manner. The intrinsic signal employed in this paper is object empowerment, denoted as $\mathfrak{E}_{\mathfrak{O}}$. In our framework, assuming that the extrinsic reward $R(s)$ depends only on the state $s \in \mathcal{S}$, we combine the latter with object empowerment $\mathfrak{E}_{\mathfrak{O}}(s)$ into a single regularized reward function $\hat{R}(s)$ as follows

$$\forall s \in \mathcal{S} \;\; \hat{R}(s) \coloneqq R(s) + \beta \mathfrak{E}_{\mathfrak{O}}(s) \quad , \tag{1}$$

where $\beta \in \mathbb{R}_{\geq 0}$ is a weighting factor that balances the contribution of extrinsic and intrinsic reward. The maximization of the regularized reward $\hat{R}$ encourages the agent to explore actions and states that increase object empowerment even in the absence of immediate extrinsic rewards. A small β places greater emphasis on the completion of the task encoded by R, while a larger β pushes the agent to maintain its control over objects of the environment, even at the cost of not addressing the task at all when β is very large. A suitable trade-off can guide the agent to interact with objects during early learning, thereby facilitating task completion in later stages.

1) State Space: Formally, the environment consists of an agent, a set of n tools $\mathfrak{T} = \{\mathfrak{T}_1, \mathfrak{T}_2, \ldots, \mathfrak{T}_n\}$, and a set of m objects $\mathfrak{O} = \{\mathfrak{O}_1, \mathfrak{O}_2, \ldots, \mathfrak{O}_m\}$. Each of these entities contributes to the overall state space, defined as:

$$\mathcal{S} \coloneqq \mathcal{S}^{\mathfrak{A}} \times \left(\prod_{j=1}^{n} \mathcal{S}^{\mathfrak{T}_j} \right) \times \left(\prod_{i=1}^{m} \mathcal{S}^{\mathfrak{O}_i} \right) \times \mathcal{S}^{\mathfrak{W}} \tag{2}$$

Here, $\mathcal{S}^{\mathfrak{A}}$ is the agent's state space (e.g., its location in the environment), $\mathcal{S}^{\mathfrak{T}_j}$ is the state space of the j-th tool (e.g., its position or whether it is equipped by the agent), $\mathcal{S}^{\mathfrak{O}_i}$ is the state space of the i-th object (e.g., its location or condition), $\mathcal{S}^{\mathfrak{W}}$ includes other static components of the environment, such as walls or goal positions.

2) Action Space: Among all the actions in $\mathcal{A}$ that an agent can perform, we define now the subsets of actions that are executed while using the tools in $\mathfrak{T}$. We distinguish between the following subsets of $\mathcal{A}$: (i) the actions of the agent $\mathcal{A}^{\mathfrak{A}} \subseteq \mathcal{A}$ that do not involve the use of a tool; (ii) the actions $\mathcal{A}^{\mathfrak{T}_j} \subseteq \mathcal{A}$ that allow the agent to use the tool $\mathfrak{T}_j$, for $j = 1, 2, \ldots n$; (iii) the set $\mathcal{A}^{\mathfrak{A}\mathfrak{T}_j} \coloneqq \mathcal{A}^{\mathfrak{A}} \cup \mathcal{A}^{\mathfrak{T}_j}$ for $j = 1, 2, \ldots n$, containing both the actions of the agent that are not relevant to tool use and its actions specifically relevant to tool $\mathfrak{T}_j$. For instance, in a navigation task, $\mathcal{A}^{\mathfrak{A}}$ could contain the action "north", which moves the agent towards the north direction, where if the agent equips an axe, the set $\mathcal{A}^{\mathfrak{T}_{\text{axe}}}$ could include the action "chop".[1]

B. Object Empowerment

Empowerment [6] is defined as the Shannon capacity of an agent's actuation between action sequences and resulting states. Object empowerment extends the classical empowerment formalism by measuring an agent's influence over the state subspace $\mathcal{S}^{\mathfrak{O}_i}$ of specific objects of the environment $\mathfrak{O}_i$, rather than over the entire state space $\mathcal{S}$ [8]. Furthermore, given a tool $\mathfrak{T}_j$, object empowerment is defined by using the tool actions subset $\mathcal{A}^{\mathfrak{T}_j}$ as source of the agent's actuation channel, instead of the full agent action set $\mathcal{A}$ as in classical empowerment. This formulation not only quantify the degree of influence that an agent has over specific objects of the environment $\mathfrak{O}_i$, but also allows one to measure the impact of those interactions that are exclusively mediated via tool $\mathfrak{T}_j$.

Given a tool $\mathfrak{T}_j \in \mathfrak{T}$, let $a^h_{\mathfrak{T}_j} \coloneqq (a_1^{\mathfrak{T}_j}, a_2^{\mathfrak{T}_j}, \ldots, a_h^{\mathfrak{T}_j}) \in \mathcal{A}^h_{\mathfrak{T}_j}$ be a tool action sequence of length h, where $\mathcal{A}^h_{\mathfrak{T}_j}$ denotes the set of all possible sequences of h tool $\mathfrak{T}_j$ actions. Let S_t be a random variable representing the agent's state at time t, and $A^h_{\mathfrak{T}_j}$ the random variable for the h-step tool $\mathfrak{T}_j$ action sequence starting at time t. [2] Let $S^{\mathfrak{O}_i}_{t+h}$ be the random variable representing the state of object $\mathfrak{O}_i$ at time $t+h$. The h-step *object empowerment* $\mathfrak{E}^h_{\mathfrak{T}_j\mathfrak{O}_i}(s)$ of state $s \in \mathcal{S}$ from tool $\mathfrak{T}_j$ to object $\mathfrak{O}_i$ is defined as the Shannon capacity of the channel between the tool $\mathfrak{T}_j$ action sequence and the resulting state of object $\mathfrak{O}_i$, conditioned on the current state s:

$$\mathfrak{E}^h_{\mathfrak{T}_j\mathfrak{O}_i}(s) \coloneqq \max_{P(a^h_{\mathfrak{T}_j}|s)} I(S^{\mathfrak{O}_i}_{t+h}; A^h_{\mathfrak{T}_j} \mid S_t = s) \tag{3}$$

where $I(X;Y)$ denotes the mutual information between the random variables X and Y. $\mathfrak{E}^h_{\mathfrak{T}_j\mathfrak{O}_i}$ measures how much the actions of the tool $\mathfrak{T}_j$ can reliably influence the state of the object $\mathfrak{O}_i$. To capture an agent's control over multiple objects jointly, we extend the above formulation to define *multi-object empowerment*. Let $\mathfrak{D} = \{\mathfrak{D}_1, \mathfrak{D}_2, \ldots, \mathfrak{D}_q\} \subseteq \mathfrak{O}$ be a subset of objects. The h-step multi-object empowerment from tool $\mathfrak{T}_j$ to objects $\mathfrak{D}$ is then defined as:

$$\mathfrak{E}^h_{\mathfrak{T}_j\mathfrak{D}}(s) \coloneqq \max_{P(a^h_{\mathfrak{T}_j}|s)} I(S^{\mathfrak{D}_1}_{t+h} \ldots S^{\mathfrak{D}_q}_{t+h}; A^h_{\mathfrak{T}_j} \mid S_t = s) \tag{4}$$

In deterministic settings, where transitions and observations are uniquely determined by actions and state, the mutual information in (3) reduces to the log-cardinality of the set of distinct states of object $\mathfrak{O}_i$ that the agent can observe from s after executing all possible h-step tool action sequences $a^h_{\mathfrak{T}_j}$:

$$\mathfrak{E}^h_{\mathfrak{T}_j\mathfrak{O}_i}(s) = \log_2 \left(\left| \mathcal{S}^h_{\mathfrak{T}_j\mathfrak{O}_i}(s) \right| \right) \tag{5}$$

where $\mathcal{S}^h_{\mathfrak{T}_j\mathfrak{O}_i}(s) \coloneqq \{s_{t+h} \mid a^h_{\mathfrak{T}_j} \in \mathcal{A}^h_{\mathfrak{T}_j}, s_{t+h} = T^h(s, a^h_{\mathfrak{T}_j})\}$ is the set of states reachable from s after applying all $a^h_{\mathfrak{T}_j}$ in $\mathcal{A}^h_{\mathfrak{T}_j}$, and $T^h(s, a^h)$ denotes the h-step transition function.

C. Tool Selection Mechanism

In scenarios with multiple tools and objects an agent may benefit by knowing which tools enables the largest control over each object of the environment. While more than one tool may

[1] In this paper, interactions between objects and tools $\mathfrak{T}_j$ are always performed using actions from the set $\mathcal{A}^{\mathfrak{A}\mathfrak{T}_j}$. For notational simplicity, we denote this set as $\mathcal{A}^{\mathfrak{T}_j}$ throughout the text.

[2] When writing $A^h_{\mathfrak{T}_j}$ we omit the time index t to have a more compact notation.

exert some influence on a certain object, the level of influence may vary, with some tools that may be very effective while others may be useless. To represent all possible tool-object relationships, we define the *tool-object empowerment matrix*. It contains the state-averaged object empowerment $\hat{\mathfrak{E}}^h_{\mathfrak{T}_j\mathfrak{O}_i}$ of each tool to each object in the environment (see Table I). We define the h-step tool-object empowerment matrix $\mathbb{T} \in \mathbb{R}^{n\times m}$ as

$$\mathbb{T}[j,i] = \hat{\mathfrak{E}}^h_{\mathfrak{T}_j\mathfrak{O}_i} \quad j = 1,\dots,n,\ i = 1,\dots,m.$$

TABLE I
TOOL-OBJECT EMPOWERMENT MATRIX $\mathbb{T}$ SHOWING THE STATE-AVERAGED EMPOWERMENT $\hat{\mathfrak{E}}^h_{\mathfrak{T}_j\mathfrak{O}_i}$ FOR EACH TOOL-OBJECT PAIR. VALUES INDICATE THE DEGREE OF INFLUENCE EACH TOOL HAS OVER EACH OBJECT AND i^* INDICATES THE OBJECT OF INTEREST.

	$\mathfrak{O}_1$	$\cdots$	$\mathfrak{O}_{i^*}$	$\cdots$	$\mathfrak{O}_m$
$\mathfrak{T}_1$	$\hat{\mathfrak{E}}^h_{\mathfrak{T}_1\mathfrak{O}_1}$	$\cdots$	$\hat{\mathfrak{E}}^h_{\mathfrak{T}_1\mathfrak{O}_{i^*}}$	$\cdots$	$\hat{\mathfrak{E}}^h_{\mathfrak{T}_1\mathfrak{O}_m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathfrak{T}_{j^*}$	$\hat{\mathfrak{E}}^h_{\mathfrak{T}_{j^*}\mathfrak{O}_1}$	$\cdots$	$\hat{\mathfrak{E}}^h_{\mathfrak{T}_{j^*}\mathfrak{O}_{i^*}}$	$\cdots$	$\hat{\mathfrak{E}}^h_{\mathfrak{T}_{j^*}\mathfrak{O}_m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathfrak{T}_n$	$\hat{\mathfrak{E}}^h_{\mathfrak{T}_n\mathfrak{O}_1}$	$\cdots$	$\hat{\mathfrak{E}}^h_{\mathfrak{T}_n\mathfrak{O}_{i^*}}$	$\cdots$	$\hat{\mathfrak{E}}^h_{\mathfrak{T}_n\mathfrak{O}_m}$

Note: i^* denotes the object of interest.

Tools with non-zero average object empowerment with certain objects constitute candidates tools for interacting with those objects. On the contrary, if an item exhibits zero average object empowerment toward all objects, it can not be considered a tool for that environment. Finally, there is the tool with maximum average object empowerment for an object. Given an object of interest $\mathfrak{O}_{i^*}$, this is defined as follows:

$$\mathfrak{T}_{j^*} \coloneqq \arg\max_j \ \hat{\mathfrak{E}}^h_{\mathfrak{T}_j\mathfrak{O}_{i^*}}. \tag{6}$$

Equation (6) enables the design of a *tool selection mechanism* that can be used by artificial agents to automatically select tools when interacting with specific objects $\mathfrak{O}_{i^*}$. One can expect that, without prior knowledge about the tools, on average, the tool selected by (6) has the largest chance of being useful when interacting with the task object $\mathfrak{O}_{i^*}$. Thus, in our RL experiments, we used (6) to choose the object empowerment $\mathfrak{E}^h_{\mathfrak{T}_{j^*}\mathfrak{O}_{i^*}}$ from the selected tool $\mathfrak{T}_{j^*}$ to the task objects $\mathfrak{O}_{i^*}$ as intrinsic reward in (1). We will show that the object empowerment of the selected tool can guide exploration toward meaningful object interactions.

IV. EXPERIMENTS

We conduct our experiments in MiniHack environments [26], which allows complex interactions between tools and objects in a grid-based world. Let represent with $\mathcal{W}$ all possible cells of the grid-word. The employed state space $\mathcal{S}$ includes the grid location of the agent $s^{\mathfrak{A}} \in \mathcal{W}$, the positions of all the tools, $s_p^{\mathfrak{T}_1},\dots,s_p^{\mathfrak{T}_n} \in \mathcal{W}^n$, and the locations of all the objects, $s_p^{\mathfrak{O}_1},\dots,s_p^{\mathfrak{O}_m} \in \mathcal{W}^m$. In addition, the tools states $\mathcal{S}^{\mathfrak{T}_j}$ include an equipped status, indicating whether the tool has been picked up by the agent and added to its inventory, and a hidden status that says whether a tool is visible from the agent's point of view. Similarly, the object states $\mathcal{S}^{\mathfrak{O}_i}$ include a hidden status and a flag that indicates whether an object has been destroyed by the agent. The employed action space $\mathcal{A}$ includes the agent movements in the grid $\mathcal{A}^{\mathfrak{A}}$ (i.e., north, south, east, west) and the tools actions $\mathcal{A}^{\mathfrak{T}_j}$, which only take effect when a tool is equipped. Tools are equipped automatically when the agent moves onto the the cell where the tool is located (i.e., $s^{\mathfrak{A}} = s_p^{\mathfrak{T}_j}$). Tool actions are based on the MiniHack game mechanics, where the use of a tool involves the following three transitions: first, the agent needs to decide that it wants use a tool by executing the "apply" action; then, it chooses which tool from its inventory to use via tool identifiers actions; finally, the agent specifies one of the four cardinal directions to which apply the tool. For example, applying an axe to the north may destroy a tree located in that direction. The grid-world dynamics T is deterministic, so we have used (5) to compute object empowerment $\mathfrak{E}^h_{\mathfrak{T}_j\mathfrak{O}}$. Each experiment is formulated as an episodic MDP. The employed reward structure is sparse: agent is rewarded only for achieving the task objective, such as destroying designated objects, when the agent receives a reward of $+1$ and transitions to a terminal state. Otherwise, each other transition incurs a reward of 0. To solve the MDPs considered in our experiments, we used the Proximal Policy Optimization (PPO) algorithm [27], as implemented in the Ray's RLlib library [28].

A. Experiment 1: Empowerment-Guided Tool Selection in Single-Object Task

The environment reported in Fig. 1 includes two manipulable objects, a tree and a wall, and four available tools: an axe, a pickaxe, a tin opener and a key. We considered the task of chopping the tree and the one of destroying the wall. We start with the first one, hence, here the tree is the task-relevant object $\mathfrak{O}_{\text{tree}^*}$. To support tool selection, we compute the tool-object empowerment matrix $\mathbb{T}$ for this environment and report it in Table II. Among all the tools of this environment, only the axe has an influence over the state of the tree ($\hat{\mathfrak{E}}^h_{\mathfrak{T}_{\text{axe}}\mathfrak{O}_{\text{tree}^*}} = 4.233 \times 10^{-8}$ bits). [3] The pickaxe has only an effect on the state of the wall ($\hat{\mathfrak{E}}^h_{\mathfrak{T}_{\text{pickaxe}}\mathfrak{O}_{\text{wall}}} = 4.233 \times 10^{-8}$ bits). The tin opener and the key have no impact on any object ($\hat{\mathfrak{E}}^h_{\mathfrak{T}_{\text{tinop}}\mathfrak{O}_{\text{tree}^*}} = 0$ bits, $\hat{\mathfrak{E}}^h_{\mathfrak{T}_{\text{key}}\mathfrak{O}_{\text{wall}}} = 0$ bits)), so they should not be considered tools for this environment. Since the axe yields the highest object empowerment for the tree, $\mathfrak{T}_{\text{axe}^*}$ is selected through the tool selection mechanism of (6).

To illustrate the spatial distribution of object empowerment in this environment, we examine the empowerment landscape before and after the axe is equipped. In Fig. 2, we report the landscape of $\mathfrak{E}^8_{\mathfrak{T}_{\text{axe}^*}\mathfrak{O}_{\text{tree}^*}}$ for when the tool is not equipped. When the axe is not equipped, $\mathfrak{E}^8_{\mathfrak{T}_{\text{axe}^*}\mathfrak{O}_{\text{tree}^*}}$ is nonzero only in the cell where the axe is located, indicating that from

[3] The MDP representing this environment has more than 70000 states, due to the combinatorial contribution of the states of all the tools an objects in the environment, each one having three possible states. For this reason, and the sparsity of the landscape, in this experiment, and the following ones, the state averaged object empowerment is a very small number.

Fig. 1. Initial state of the environment of experiment 1. Black cells represent unobserved areas hidden from the current agent's field of view.

TABLE II
STATE-AVERAGED TOOL-TO-OBJECT EMPOWERMENT $\hat{\mathfrak{E}}^h_{\mathfrak{T}_j\mathfrak{O}_i}$ IN BITS FOR EACH TOOL-OBJECT COMBINATION OF EXPERIMENT 1.

	$\mathfrak{O}_{\text{tree}*}$	$\mathfrak{O}_{\text{wall}*}$
$\mathfrak{T}_{\text{axe}*}$	4.233×10^{-8}	0
$\mathfrak{T}_{\text{pickaxe}*}$	0	4.233×10^{-8}
$\mathfrak{T}_{\text{tinop}}$	0	0
$\mathfrak{T}_{\text{key}}$	0	0

there in 8 steps the agent can reach tree and chop it. There, the value of $\mathfrak{E}^8_{\mathfrak{T}_{\text{axe}*}\mathfrak{O}_{\text{tree}*}}(s_p^{\mathfrak{T}_{\text{axe}*}})$ is 1 bit, because the agent can either chop the tree or leave it intact. When $\mathfrak{E}^8_{\mathfrak{T}_{\text{axe}*}\mathfrak{O}_{\text{tree}*}}$ is used as intrinsic reward, this acts as a beacon towards the tool location, helping the agent to find the axe while exploring the environment. In Fig. 3 we report the landscape of $\mathfrak{E}^3_{\mathfrak{T}_{\text{axe}*}\mathfrak{O}_{\text{tree}*}}$ for when the tool is equipped. The landscape shows non-zero values (i.e., 1 bit) of $\mathfrak{E}^3_{\mathfrak{T}_{\text{axe}*}\mathfrak{O}_{\text{tree}*}}$ in locations adjacent to the tree. This indicates that from those positions, the agent can chop the tree in the 3 steps necessary to execute the "apply"→"choose"→"direction" sequence of actions illustrated in the previous section. This landscape reflects the fact that the axe is a tool whose influence is highly localized and effective only when the agent is next to its target object. When used as intrinsic reward, $\mathfrak{E}^3_{\mathfrak{T}_{\text{axe}*}\mathfrak{O}_{\text{tree}*}}$ acts as beacon that attracts the agent to the tree once the axe is equipped. Since the agent needs more steps to interact with an object when a tool is unequipped (i.e., additional steps are necessary to reach the tool and pick it up), in our experiments we have used a longer horizon h for states where the tool is unequipped and a shorter horizon for states where the tool is equipped (here, $h = 8$ and $h = 3$ respectively).

To evaluate learning performance under sparse reward conditions, we compare a standard PPO agent with an intrinsically motivated agent whose reward is regularized with object empowerment. Fig. 4 shows the average cumulative reward across training episodes, computed over 10 independent runs. The agent using $\mathfrak{E}^h_{\mathfrak{T}_{\text{axe}*}\mathfrak{O}_{\text{tree}*}}$ ($\beta = 0.0009$) shows a faster convergence to optimal performance compared to the baseline PPO agent. This improvement highlights how object empowerment

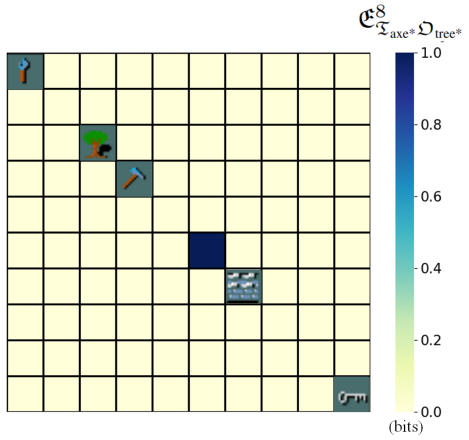

Fig. 2. 8-step axe to tree empowerment $\mathfrak{E}^8_{\mathfrak{T}_{\text{axe}*}\mathfrak{O}_{\text{tree}*}}$ landscape for all possible agent's locations (in bits), when the agent is not equipped with the axe.

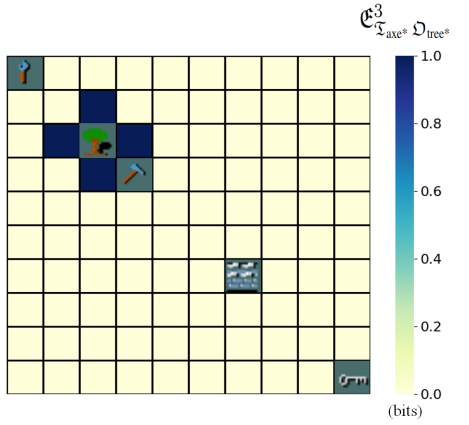

Fig. 3. 3-step axe to tree empowerment $\mathfrak{E}^3_{\mathfrak{T}_{\text{axe}*}\mathfrak{O}_{\text{tree}*}}$ landscape for all possible agent's locations (in bits), when the agent is equipped with the axe.

helps guide exploration in sparse reward environments where useful tool-object interactions must be discovered. Similar results were obtained when the objective was to destroy the wall $\mathfrak{O}_{\text{wall}*}$ and the the pickaxe $\mathfrak{T}_{\text{pickaxe}*}$ was selected as a tool, confirming the generality of the proposed approach.

B. Experiment 2: Empowerment-Guided Tool Selection in Multi-Object Task

In this experiment, we explore a more challenging scenario where the agent is required to destroy two distinct objects: a boulder and a door (i.e., with multiple targets $\mathfrak{O}_{\text{bould}*}\mathfrak{O}_{\text{door}*}$). The agent receives a reward of 1 for each object successfully destroyed. The environment (see Fig. 5) contains four tools: a wand, an axe, a tin opener, and a katana. Here, the wand can destroy both the boulder and the door, the axe is capable of only destroying the door, while the tin opener and katana serve as distractors with no effect on the environment's objects. In addition, the environment includes walls that act as static barriers, preventing agent movement, which do not serve as manipulable objects.

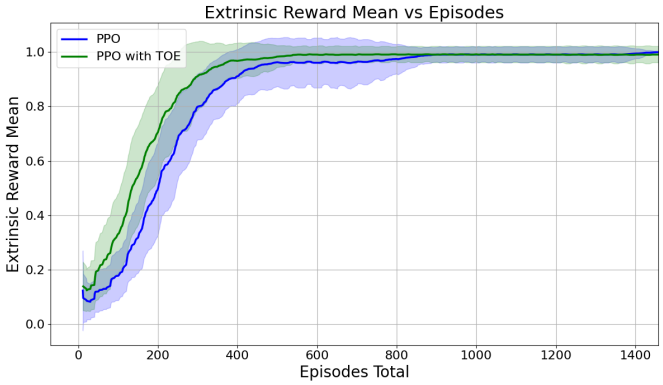

Fig. 4. The agent using the axe to tree empowerment $\mathfrak{E}^{h}_{\mathfrak{T}_{\text{axe}^*}\mathfrak{O}_{\text{tree}^*}}$ as a regularizer (green) shows faster convergence compared to the standard PPO (blue). Shaded regions represent standard deviation across 10 independent runs.

Fig. 5. Initial state of the environment of experiment 2.

We report the tool-object empowerment matrix $\mathbb{T}$ for this environment in Table III. In addition to the average tool to object empowerment of the individual objects, the Table III also reports the average tool to object empowerment $\hat{\mathfrak{E}}^{h}_{\mathfrak{T}_j\,\mathfrak{O}_{\text{bould}^*}\mathfrak{O}_{\text{door}^*}}$ of the two objects considered together (see (4)). Being $\hat{\mathfrak{E}}^{h}_{\mathfrak{T}_{\text{wand}}\,\mathfrak{O}_{\text{bould}^*}\mathfrak{O}_{\text{door}^*}}$ the largest average object empowerment for both targets, our tool selection method chooses the wand $\mathfrak{T}_{\text{wand}^*}$ and its boulder-door empowerment as intrinsic reward for RL.

TABLE III
STATE-AVERAGED TOOL-TO-OBJECT EMPOWERMENT $\hat{\mathfrak{E}}^{h}_{\mathfrak{T}_j\mathfrak{O}_i}$ FOR EACH TOOL-OBJECT COMBINATION. THE LAST COLUMN REFLECTS THE MULTI-OBJECT EMPOWERMENT $\hat{\mathfrak{E}}^{h}_{\mathfrak{T}_j\,\mathfrak{O}_{\text{BOULD}^*}\mathfrak{O}_{\text{DOOR}^*}}$.

	$\mathfrak{O}_{\text{bould}}$	$\mathfrak{O}_{\text{door}}$	$\mathfrak{O}_{\text{bould}^*}\mathfrak{O}_{\text{door}^*}$
$\mathfrak{T}_{\text{wand}^*}$	5.292×10^{-7}	6.138×10^{-7}	9.564×10^{-7}
$\mathfrak{T}_{\text{axe}}$	0	3.281×10^{-7}	3.281×10^{-7}
$\mathfrak{T}_{\text{tinop}}$	0	0	0
$\mathfrak{T}_{\text{kata}}$	0	0	0

In this experiment we use $h = 5$ when the wand is unequipped for reward regularization, because this horizon yields an object empowerment landscape peaked in the location of the wand, and $h = 6$ when the wand is equipped. We report the wand-equipped landscape of $\mathfrak{E}^{6}_{\mathfrak{T}_{\text{wand}}\,\mathfrak{O}_{\text{bould}^*}\mathfrak{O}_{\text{door}^*}}$ in Fig. 6. Unlike the tools in Experiment 1, which can affect objects only when adjacent to it, here the wand's area of influence spans a larger portion of the grid. This is because in MiniHack, the wand can strike objects at arbitrary distances along the orthogonal directions from the agent's position. This observation suggests that object empowerment formalism could be used to characterized tools-object range of interaction. The 2.0 bits peaks of $\mathfrak{E}^{6}_{\mathfrak{T}_{\text{wand}}\,\mathfrak{O}_{\text{bould}^*}\mathfrak{O}_{\text{door}^*}}$ are in the two cells where in 6 steps the agent can destroy both the boulder and the door (i.e., 3 steps to destroy one plus 3 steps to destroy the other). Intermediate values, such as 1.58 and 1.0 bits, appear in locations where the agent can affect either one of the two objects or only one of the two, respectively.

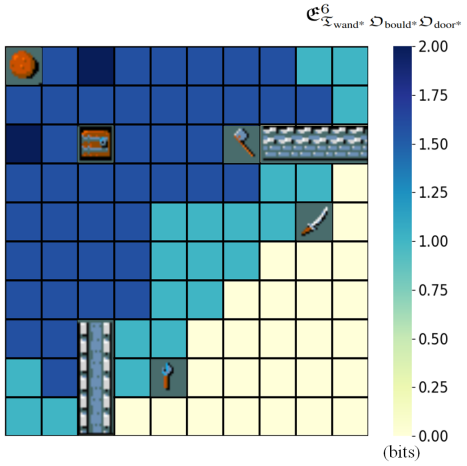

Fig. 6. 6-step wand to boulder-door empowerment $\mathfrak{E}^{6}_{\mathfrak{T}_{\text{wand}^*}\,\mathfrak{O}_{\text{bould}^*}\mathfrak{O}_{\text{door}^*}}$ landscape for all possible agent's locations, when the agent is equipped with the wand.

We compare learning performance using standard PPO and PPO regularized with $\mathfrak{E}^{h}_{\mathfrak{T}_{\text{wand}^*}\,\mathfrak{O}_{\text{bould}^*}\mathfrak{O}_{\text{door}^*}}$. Fig. 7 reports the the cumulated reward mean per episode averaged over 10 independent runs. The TOE-augmented agent converges rapidly and attains higher final performance compared to the baseline. In contrast, the standard PPO agent frequently plateaus at suboptimal values, indicating it gets trapped in local optima, for instance by learning to destroy only one object. TOE-based regularization helps overcome this limitation by encouraging policies that expand future influence towards both objects, driving the agent toward broader interaction strategies that ultimately solve the full task.

C. Experiment 3: Tool Use to Achieve Sub-goals

In this last experiment, we examine a more complex scenario, where tools enable task completion not through direct manipulation of the goal object, but via the interaction with another object whose manipulation is a pre-condition to reach the goal object. We depict the environment in Fig. 8. The environment contains an axe and a key as tools, and a door and a boulder as objects. The task for the agent is to move the

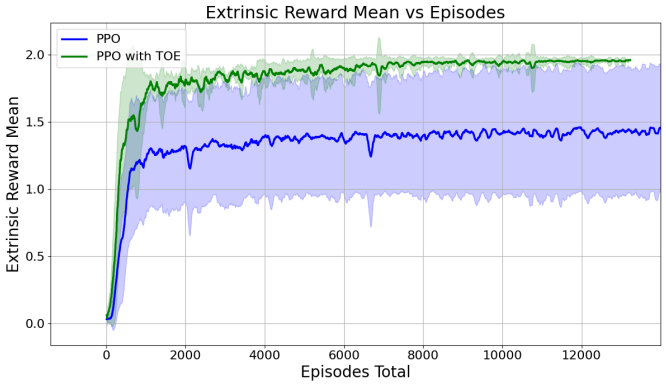

Fig. 7. The agent using $\mathfrak{E}^{h}_{\mathfrak{T}_{\text{wand}^*}\,\mathfrak{O}_{\text{bould}^*}\mathfrak{O}_{\text{door}^*}}$ as a regularizer with $\beta = 0.0009$ (green) learns faster and more reliably than the standard PPO agent (blue).

Fig. 8. Initial state of the environment of experiment 3.

boulder onto a designated goal location (highlighted as a blue square). To "push" the boulder the agent must occupy a cell adjacent to a it and executes a movement action toward it, then both the agent and the boulder are displaced by one cell in the same direction. Hence, the boulder in this environment can be moved directly by the agent without requiring any tool. However, the boulder is initially inaccessible, positioned behind a locked door and surrounded by walls that restrict movement. To reach and push the boulder, the agent must first move through the door, either by opening it with the key or by destroying it using the axe. These tools therefore provide instrumental affordances: they do not act directly on the boulder but instead enable access to it by modifying the environment. Although the axe and the key do not impact the state of the boulder directly, the average axe to boulder empowerment $\hat{\mathfrak{E}}^{h}_{\mathfrak{T}_{\text{axe}}\,\mathfrak{O}_{\text{bould}^*}}$ and key to boulder empowerment $\hat{\mathfrak{E}}^{h}_{\mathfrak{T}_{\text{key}}\,\mathfrak{O}_{\text{bould}^*}}$ are non-zero for $h \geq 7$: they emerge indirectly, through a causal chain of actions where the agent alters the door state using a tool and subsequently moves the boulder using its own body.

In Fig. 9, we report the landscape of 7-step axe to boulder empowerment $\hat{\mathfrak{E}}^{7}_{\mathfrak{T}_{\text{axe}}\,\mathfrak{O}_{\text{bould}^*}}$ when the axe is initially not equipped. Fig. 10 presents the landscape of 5-step axe to boulder empowerment $\hat{\mathfrak{E}}^{5}_{\mathfrak{T}_{\text{axe}}\,\mathfrak{O}_{\text{bould}^*}}$ when the axe is equipped. In Fig. 11 we report the landscape of 5-step axe to boulder

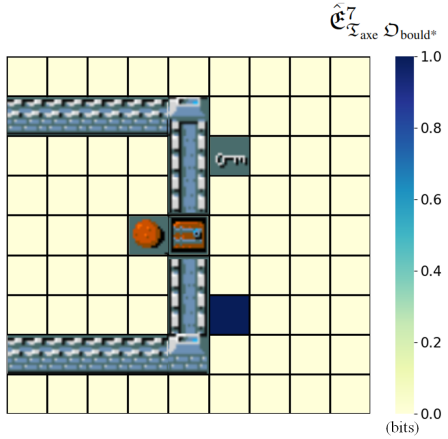

Fig. 9. 7-step axe to boulder empowerment $\hat{\mathfrak{E}}^{7}_{\mathfrak{T}_{\text{axe}}\,\mathfrak{O}_{\text{bould}^*}}$ landscape for all possible agent's location, when the agent is not equipped with the axe in experiment 3.

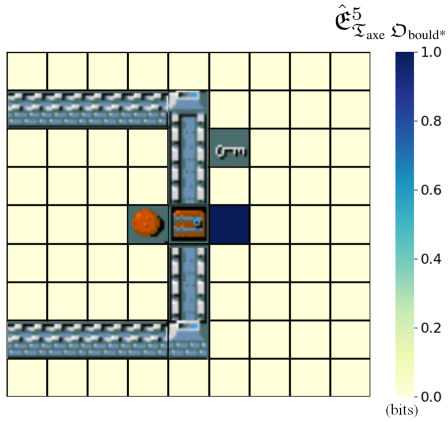

Fig. 10. 5-step axe to boulder empowerment $\hat{\mathfrak{E}}^{5}_{\mathfrak{T}_{\text{axe}}\,\mathfrak{O}_{\text{bould}^*}}$ landscape for all possible agent's location, when the agent is equipped with the axe in experiment 3.

empowerment $\hat{\mathfrak{E}}^{5}_{\mathfrak{T}_{\text{axe}}\,\mathfrak{O}_{\text{bould}^*}}$ when the axe is equipped and the room is accessible. Differently from the objects of the previous experiments, which had only two possible states, the boulder can be repeatedly pushed in multiple directions and transit in always more states as the number of interactions with it increases, creating a richer object empowerment landscape. As a result, $\hat{\mathfrak{E}}^{5}_{\mathfrak{T}_{\text{axe}}\,\mathfrak{O}_{\text{bould}^*}}$ increases with h and with the proximity to the boulder, two features that, when used as intrinsic reward, not only make the boulder empowerment to act as beacon for the object, but also as a sort of gradient towards it, which in turn facilitates learning.

The tools-door interactions can themselves be characterized by their axe to door empowerment $\mathfrak{E}^{h}_{\mathfrak{T}_{\text{axe}}\,\mathfrak{O}_{\text{door}}}$ and key to door empowerment $\mathfrak{E}^{h}_{\mathfrak{T}_{\text{key}}\,\mathfrak{O}_{\text{door}}}$, which enables downstream influence to the boulder empowerment $\mathfrak{E}^{h}_{\mathfrak{T}_{\text{axe}}\,\mathfrak{O}_{\text{bould}^*}}$ and $\mathfrak{E}^{5}_{\mathfrak{T}_{\text{key}}\,\mathfrak{O}_{\text{bould}^*}}$, respectively. When the agent is located in the cell in front of the door, both tools yield a 3-step door empowerment of $\mathfrak{E}^{3}_{\mathfrak{T}_{\text{axe}}\mathfrak{O}_{\text{door}}} = \mathfrak{E}^{3}_{\mathfrak{T}_{\text{key}}\mathfrak{O}_{\text{Door}}} = 1.0$ bit (the door is either cleared,

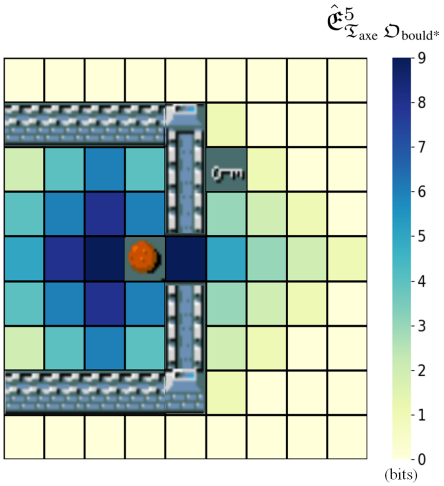

Fig. 11. 5-step boulder empowerment landscape after the room becomes accessible in experiment 3.

i.e. opened by the key or destroyed by the axe, or closed). Furthermore whether the door has been opened by the key or destroyed by the axe does not make any difference w.r.t. enabling the interaction of the agent with the boulder. But, although the axe and the key seems to act in a similar manner here, there is a fundamental difference between the two tools. The key operates reversibly: the door can be repeatedly opened and re-closed. In contrast, when the axe irreversibly destroys the door, it precludes any further interaction with it. From a purely object empowerment-based perspective, both interactions have the same door empowerment $\mathfrak{E}^{h}_{\mathfrak{T}_j\,\mathfrak{O}_{\text{door}}}$ of 1 bit for $h >= 3$. However, if we condition the object empowerment on the door not being closed, these two quantities dramatically differ. As shown in Fig. 12, when conditioned, the $\mathfrak{E}^{h}_{\mathfrak{T}_j\,\mathfrak{O}_{\text{door}}}$ of 1 bit for $h >= 3$ evolves differently depending on whether the agent uses a key or an axe. In the Fig. 12, the x-axis represents time steps t, and the y-axis encodes the possible state of the door (i.e., open, closed, or destroyed). Object empowerment values are represented by the color of the curves (red for 1 bit and blue for 0 bits). For the key, the door empowerment remains at a steady value of 1 bit even as the door changes state, highlighting the reversibility and temporal persistence of the key's influence on the door. By contrast, when using the axe, the agent can only destroy the door once and, after that, no further state transitions are possible, and empowerment drops and stays at 0 bits.

In Fig. 13 we show that the agent learning with $\mathfrak{E}^{h}_{\mathfrak{T}_{\text{axe*}}\,\mathfrak{O}_{\text{door*}}\mathfrak{O}_{\text{bould*}}}$ as intrinsic rewards ($h = 7$ for states with the axe unequipped and $h = 5$ for equipped states) and ($\beta = 0.00006$) learns quicker and reaches higher asymptotic performance compared to the PPO baseline. This improvement demonstrates how object empowerment can encourage policies that account for multi-step dependencies, such as clearing intermediate obstacles to eventually reach the goal object.

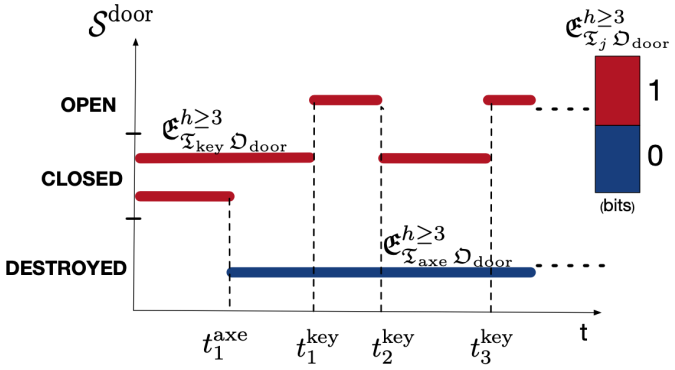

Fig. 12. Temporal evolution of object empowerment $\mathfrak{E}^{h\geq 3}_{\mathfrak{T}_j,\mathfrak{O}_{\text{Door}}}$ as the agent interacts with the door using either the key or the axe.

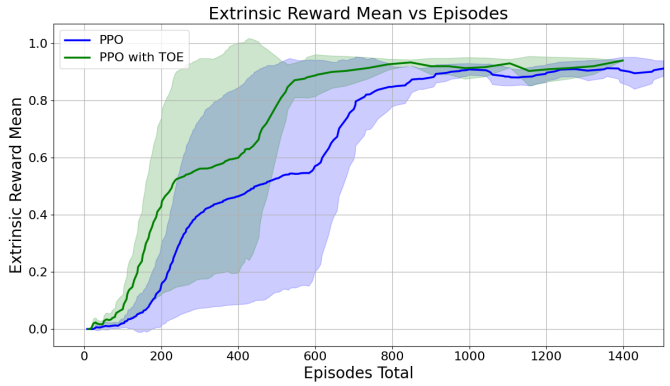

Fig. 13. The agent using $\mathfrak{E}^{h}_{\mathfrak{T}_{\text{axe*}}\,\mathfrak{O}_{\text{door*}}\mathfrak{O}_{\text{bould*}}}$ as a regularizer with $\beta = 0.0006$ (green) learns faster compared to the standard PPO agent (blue).

V. Conclusions

This work has explored how object empowerment, an object-centric intrinsic motivation, can guide learning of tool-use behaviors in environments with multiple tools and objects. Rooted in cognitive science-inspired principles, our approach reflects the idea that adaptive organisms seek to maximize their potential influence over the objects that surround them. Across increasingly complex experiments, we demonstrated that object empowerment enables to: (i) identify tools with the highest potential influence over task-relevant objects, (ii) characterize tools interactions that support interactions with multiple objects, that are long ranged, or that are persistent and reversible, (iii) reason about downstream effects of tool use over sequence of objects. Our experiments showed that agents guided by object empowerment consistently outperform vanilla baseline RL agents, converging faster and more reliably despite sparse rewards, suggesting that this mechanism could be one of the drives that pushed the evolution of tool use in the animal kingdom.

For clarity and ease of interpretation, our experiments were conducted in simple grid-world environments that are fully observable, deterministic, and discrete. Nevertheless, our framework can be readily extended to more complex and realistic scenarios by adapting existing extensions of the

classical empowerment formalism to the case of object empowerment. In stochastic discrete environments, our definition of object empowerment in (3) remains fully valid. However, for its computation, instead of using equation (5), one would need to employ the Blahut–Arimoto algorithm [29], [30] for an exact solution. Furthermore, empowerment was originally defined for partially observable environments [6], [7], where the receiver variable of the actuation channel corresponds to the agent's observations $\mathcal{O}$ rather than the state variable $\mathcal{S}$. This leads to a measure of the amount of influence that an agent has on its environment that it can actually perceive. In the case of object empowerment, this translates into measuring how much change in the object the agent can observe through its sensors. In continuous domains, empowerment has been successfully estimated via Gaussian-channel-based approximations with known [31], [32] and unknown [33] dynamics, as well as via variational approximations [23], [34], [35]. These families of methods has led to interesting applications of the empowerment framework in robotics [24], [36], [37] and can be directly employed for the computation of object empowerment. The application of object empowerment to tool use in robotics is an especially promising direction for future work. While the computation of object empowerment in discrete deterministic environments presented here (5) scales exponentially with the action horizon, a UCT-like pruning method has been proposed to reduce the complexity of the exact computation of empowerment in this context [38]. In addition, the same aforementioned variational estimations used for continuous domains also provide an efficient solution to approximate empowerment in large or high-dimensional discrete environments, which could be readily applied in the context of object empowerment.

This study focused on object empowerment as an intrinsic motivation. Previous work [8] compared the performance of object empowerment with that of classical empowerment in similar tool-use RL settings, showing that there are cases where classical empowerment can also be beneficial for tool-use learning. While we believe that object empowerment provides a particularly transparent and interpretable account of how an agent's actions influence specific objects in its environment, it would be informative to compare its effects with other prominent intrinsic motivation approaches, such as novelty [39], curiosity-driven exploration [21], and random network distillation [40]. Such a comparison would help assess whether the observed gains arise specifically from object empowerment or represent a more general benefit of intrinsic reward shaping, and how these results compare with the performance obtained using other intrinsic motivations. Conducting a systematic comparison between object empowerment and such approaches constitutes a promising direction for future research, which we plan to pursue in subsequent work.

In future work we intend to explore how object empowerment can be further extended to characterize key structural properties of tools, by pushing forward a complete formalization of the notion of a tool's range of interaction, and by introducing a measure of reliability of their control over objects. Finally, although object empowerment quantifies the "volume" of influence that a tool has over an object, it is indifferent to the specific semantic role that an object has in a given task. Thus, the combination of object empowerment with more goal-directed extensions of the empowerment formalism [41] could provide a promising direction of research.

REFERENCES

[1] S. L. Washburn, "Speculations on the interrelations of the history of tools and biological evolution," Human Biology, vol. 31, pp. 21–31, Feb 1959.

[2] A. Seed, and R. Byrne, R, "Animal tool-use," Current biology, vol. 20, pp. R1032–R1039, Dec 2010.

[3] R. S. Amant, and A. B. Wood, "Tool use for autonomous agents," in AAAI, pp. 184-–189. Jul 2005.

[4] P. Y. Oudeyer, and F. Kaplan, "What is intrinsic motivation? a typology of computational approaches," Frontiers in neurorobotics, vol. 1, pp. 108 Nov 2007.

[5] A. Barto, M. Mirolli and G. Baldassarre, "Novelty or surprise?" Frontiers in psychology, vol. 4, pp. 907, Dec 2013.

[6] A. S. Klyubin, D. Polani, and C, L. Nehaniv, "Empowerment: A universal agent-centric measure of control," IEEE congress on evolutionary computation, vol. 1, pp. 128-–135, Sep 2005.

[7] C. Salge, C. Glackin, and D. Polani, "Empowerment–an introduction," Guided Self-Organization: Inception Berlin, Heidelberg: Springer Berlin Heidelberg pp. 67—114, Oct 2014.

[8] F. Rasheed, D. Polani, and N. C. Volpi, "Leveraging empowerment to model tool use in reinforcement learning." in IEEE International Conference on Development and Learning (ICDL), IEEE, pp. 28—36, Nov 2023.

[9] M. Günther, and C. Boesch, "Energetic cost of nut-cracking behaviour in wild chimpanzees," Hands of primates, Vienna: Springer Vienna, pp. 109—129, 1993.

[10] A. Barto, and S. R. Sutton, "Reinforcement learning: an introduction," Cambridge: MIT press, 2018.

[11] M. Riedmiller, R. Hafner, T. Lampe, M. Neunert, J. Degrave, T. Wiele, V. Mnih, N. Heess, and J. T. Springenberg, "Learning by playing solving sparse reward tasks from scratch," in International conference on machine learning, pp. 4344-–4353. PMLR, Jul 2018.

[12] D. Biro, M. Haslam, and C. Rutz, "Tool use as adaptation," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 368 pp. 20120408, Nov 2013.

[13] C. Baber, "Cognition and tool use: Forms of engagement in human and animal use of tools," CRC Press, Jul 2003.

[14] M. Qin, J. Brawer, B. Scassellati, "Robot tool use: A survey," Frontiers in Robotics and AI, vol. 16, pp. 1009488, Jan 2023.

[15] R. Jain, R. and T. Inamura, "Learning of tool affordances for autonomous tool manipulation," in IEEE/SICE international symposium on system integration (SII), IEEE, pp. 814-–819, Dec 2011.

[16] A. Stoytchev, "Behavior-grounded representation of tool affordances," in Proceedings of the 2005 IEEE international conference on robotics and automation, IEEE, pp. 3060—3065, April 2005.

[17] S. Wenke, D. Saunders, M. Qiu, and J. Fleming, "Reasoning and generalization in rl: A tool use perspective," arXivpreprint arXiv:1907.02050, Jul 2019.

[18] Z. Liu, S. Tian, M. Guo, C. K. Liu, and J. Wu, "Learning to design and use tools for robotic manipulation," arXiv preprint arXiv:2311.00754, Nov 2023.

[19] K. Seepanomwan, D. Caligiore, K. J. O'Regan, and G. Baldassarre, "Intrinsic motivations and planning to explain tool-use development: A study with a simulated robot model," IEEE Transactions on Cognitive and Developmental Systems, vol. 14, pp. 75–89, May 2020.

[20] M. Prokopenko, "Guided self-organization: Inception," Springer Science & Business Media, vol. 9, Dec 2013.

[21] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, "Curiosity-driven exploration by self-supervised prediction," in International conference on machine learning, pp. 2778-– 2787. PMLR, Jul 2017.

[22] N. Bougie, and R. Ichise, "Skill-based curiosity for intrinsically motivated reinforcement learning," Machine Learning, vol. 109, pp. 493-–512, Mar 2020.

[23] S. Mohamed, and J. D. Rezende, "Variational information maximisation for intrinsically motivated reinforcement learning," in Advances in neural information processing systems, 28, 2015.
[24] S. Dai, W. Xu, A. Hofmann, and B. Williams, "An empowerment-based solution to robotic manipulation tasks with sparse rewards," Autonomous Robots, vol. 47, pp. 617–633, Jun 2023.
[25] A. Lidayan, Y. Du, E. Kosoy, M. Rufova, P. Abbeel, and A. Gopnik, "Intrinsically-motivated humans and agents in open-world exploration," arXiv preprint arXiv:2503.23631, Mar 2025.
[26] M. Samvelyan, R. Kirk, V. Kurin, J. Parker-Holder, M. Jiang, E.Hambro, F. Petroni, H.Küttler, E. Grefenstette, and T. Rocktäschel, "Minihack the planet: A sandbox for open-ended reinforcement learning research," arXiv preprint arXiv:2109.13202, Sep 2021.
[27] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, Jul 2017.
[28] E. Liang, R. Liaw, R. Nishihara, P. Moritz, R. Fox, K. Goldberg, J. Gonzalez, M. Jordan, and I. Stoica, "Rllib: Abstractions for distributed reinforcement learning," in International conference on machine learning, pp. 3053-–3062, PMLR, Jul 2018.
[29] R. Blahut, "Computation of channel capacity and rate-distortion functions," IEEE Transactions on Information Theory, vol. 18, pp. 460-473, Jan 1972.
[30] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," IEEE Transactions on Information Theory, vol. 18, pp. 14-20, Jan 1972.
[31] C. Salge, C. Glackin, and D. Polani, "Approximation of empowerment in the continuous domain," Advances in Complex Systems, vol. 16, pp. 1250079, May 2013.
[32] S. Tiomkin, I. Nemenman, D. Polani, and N. Tishby, "Intrinsic motivation in dynamical control systems," PRX Life, vol. 2, pp. 033009, Aug 2024.
[33] R. Zhao, K. Lu, P. Abbeel, and S. Tiomkin, "Efficient empowerment estimation for unsupervised stabilization," arXiv preprint arXiv:2007.07356, Jul 2020.
[34] K. Gregor, D.J. Rezende, and D. Wierstra, "Variational intrinsic control," arXiv preprint arXiv:1611.07507, Nov 2016.
[35] M. Karl, P. Becker-Ehmck, M. Soelch, D. Benbouzid, P. van der Smagt, and J. Bayer, "Unsupervised real-time control through variational empowerment," In The International Symposium of Robotics Research, pp. 158-173, Oct 2019.
[36] H. Cao, F. Feng, J. Huo, and Y. Gao, "Causal action empowerment for efficient reinforcement learning in embodied agents," Science China Information Sciences, vol. 68, pp. 150201, May 2025.
[37] N.C. Volpi, D. De Palma, D. Polani, and G. Indiveri, "Computation of empowerment for an autonomous underwater vehicle," IFAC-PapersOnLine, vol. 49, pp. 81-87, Jan 2016.
[38] C. Salge, C. Guckelsberger, R. Canaan, and T. Mahlmann, "Accelerating empowerment computation with UCT tree search," In 2018 IEEE Conference on Computational Intelligence and Games (CIG), IEEE, pp. 1-8, Aug 2018.
[39] A. Barto, M. Mirolli, and G. Baldassarre, "Novelty or surprise?," Frontiers in psychology, vol. 4, pp. 907, Dec 2013.
[40] Y. Burda, H. Edwards, A. Storkey, and O. Klimov, "Exploration by random network distillation," arXiv preprint arXiv:1810.12894, Oct 2018.
[41] N. C. Volpi, and D. Polani, "Goal-directed empowerment: combining intrinsic motivation and task-oriented behavior," IEEE Transactions on Cognitive and Developmental Systems, vol. 15, pp. 361-–372, Dec 2020.