Understanding Tools like a Crow: an Ultimate Goal for Embodied Agents

Zhancun Mu
YuanPei College
Peking University
yhbylch@stu.pku.edu.cn

Abstract

The ability to create and use tools has long been considered a key distinguishing feature of human intelligence compared to other animals [7]. However, scientists have found that some animals, such as New Caledonian crows, also demonstrate a profound understanding of tools [4]. In AI research today, this ability is agreed to be very important, yet it remains largely unsolved. This essay will discuss the key components of tool understanding in order to inspire current agent research.

1 Introduction

The fact that adult Caledonian crows can use sticks to reach their prey is impressive enough. Even more astonishing is that they have developed the habit of carving a hook at the end of a stick to better accomplish this. Not only are crows tweaking and refining their old designs through exploration, they are also possibly sharing their knowledge with other crows. This suggests that crows have the capacity to develop a profound and nuanced understanding of tools, including the abilities to create, use, and improve them [4][11].

Understanding tools is a key aspect of human intelligence. It involves abilities such as identifying tools, manipulating tools, and manufacturing tools. Many complex tasks in intricate environments cannot be accomplished by hand alone. As such, this skill is crucial for robots to interact better with their surroundings and accomplish goals. However, enabling robots to understand tools remains a demanding, open challenge in current AI research.

In this essay, we will discuss the key components an agent needs to understand tools and integrate them into a unified framework. First, we will introduce the definition and taxonomy of tools proposed by [8]. Using this taxonomy as a guide, we will outline the requirements for such a framework and review current achievements in AI research related to tool understanding.

2 Definition and taxonomy of tools

While the concept of tool use is intuitive, providing a precise definition is challenging. One potential definition is a robot using objects to achieve a goal. However, a robot utilizing the floor to support itself would not be considered true tool use under this definition. Taking into account previous definitions, Qin et al. [8] identify three key elements:

- Tool use must have a goal.
- An agent utilizing a tool should understand the connection between the goal and the behavior, instead of achieving a goal through random exploration.
- The tool should satisfy specific physical criteria, such as being freely manipulable.

A formal definition by [8] is:

A robot attaches or secures to its end-effector an external, unanimated, freely available object or an object attached to another object, in order to achieve a goal of altering the state of another object, updating its own state, or other goals, through purposeful manipulations.

It is worth noting that the second point contrasts with the current paradigm of reinforcement learning, which requires agents to learn through trial and error without an understanding of the connection between the goal and behavior.

Taxonomies of tool use have also been widely discussed in both animal studies and robotics research. From the perspective of causal reasoning, [8] categorize robot tool use into *non-causal tool use* and *causal tool use*. The key difference is whether the agent needs to understand causal relationships. A more detailed taxonomy for *causal tool use* is provided in Tab. 1.

Type	Explanation
Basic Tool Use	The most basic form of tool use.
Transferable Tool Use	Transfer learned tool use skills to other intra-category objects that share common form factors.
Improvisatory Tool Use	Generalize learned tool use skills to novel inter-category objects which are generally not designed to perform these tasks
Deductive Tool Use	Deduct the usage of a tool from its physical knowledge about the world.
Combined Tool Use	Using two tools simultaneously.
Sequential Tool Use	Using multiple tools one after another.
Tool Selection	Choose the most appropriate tool among many options. in order to complete a tool use task.
Tool Manufacturing	Modify or Combine objects or existing tools, with or without the usage of other tools, to complete a tool use task, or to complete the task more efficiently.

Table 1: Taxonomy of causal tool use. The first four types belong to *Single-manipulation tool use* and the last four types belong to *Multi-manipulation tool use*. Credit: Qin et al. [8].

3 Towards an unified framework for tool understanding

Based on the above definitions, understanding tools encompasses three levels: perception, manipulation, and high-level cognition. Perception involves the ability to identify objects that can potentially serve as tools. Manipulation skills relate to using tools to accomplish goals. High-level cognition refers to reasoning about tools, comprehending their affordances, functions, and dynamics. This level of understanding is crucial for tasks like multi-tool planning, manufacturing, *etc*.

3.1 Perception

In contrast to traditional classification tasks in computer vision, objects can serve as different tools in various situations based on the goal. The purpose of tool use is to accomplish goals, which points to a task-oriented framework such as the one proposed by Zhu et al. [12] and shown in Fig. 1. Additionally, a deep understanding of tools' physical properties and related affordances is required. For instance, in the task of reaching prey, any elongated, slender object with a suitable end can function as a tool. Considering these factors, Zhu et al. [12] represent objects using a spatial-temporal parse graph. Saito et al. [10] utilize a deep neural network model to recognize object traits, obtain tool-object-action connections, and generate motions for tool selection and handling.

It should also be noted that robots with different configurations may have different perceptions of tools. For instance, a robot with a gripper could use a stick to reach prey, whereas a robot with a suction cup would be unable to do so. This relates to the correspondence problem, which will be discussed later.

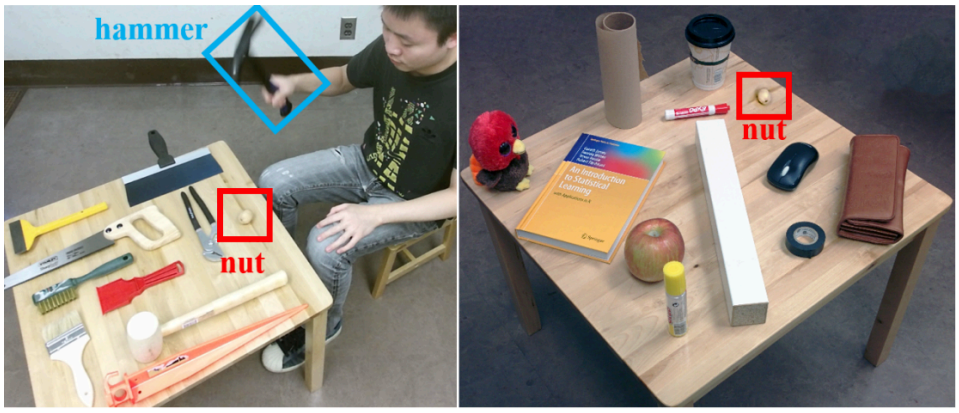


Figure 1: Task-oriented object recognition. (a) In a learning phase, a rational human is observed picking a hammer among other tools to crack a nut. (b) In an inference phase, the algorithm is asked to pick the best object (i.e. the wooden leg) on the table for the same task. This generalization entails physical reasoning. Image and caption credit: Zhu et al. [12].

3.2 Manipulation

While manipulation encompasses a broad scope, we will concentrate on generating low-level actions. Formally, we require a policy $\pi(a|s, g, t)$ to accomplish a goal g using a tool t in state s. This could be achieved by programming a wide variety of tool use actions. However, we desire generalized, robust, and transferable solutions. Current approaches include reinforcement learning, imitation learning, etc. However, manipulating tools with the proficiency of crows remains extremely challenging for present RL algorithms, which depend heavily on reward shaping and are data inefficient. Here we will focus on imitation learning.

The ability for *Learning from Demonstration* (LfD) is observed in many animals. To formulate this problem, we denote the expert demonstrations as a collection of trajectories $\mathcal{D} = \tau_i i = 1^N$. Each trajectory τ_i comprises a sequence of states, actions (and rewards) $\tau_i = (s_t, a_t, (r_t))t = 1^{T_i}$. Three approaches are summarized by Argall et al. [2] in Fig. 2.

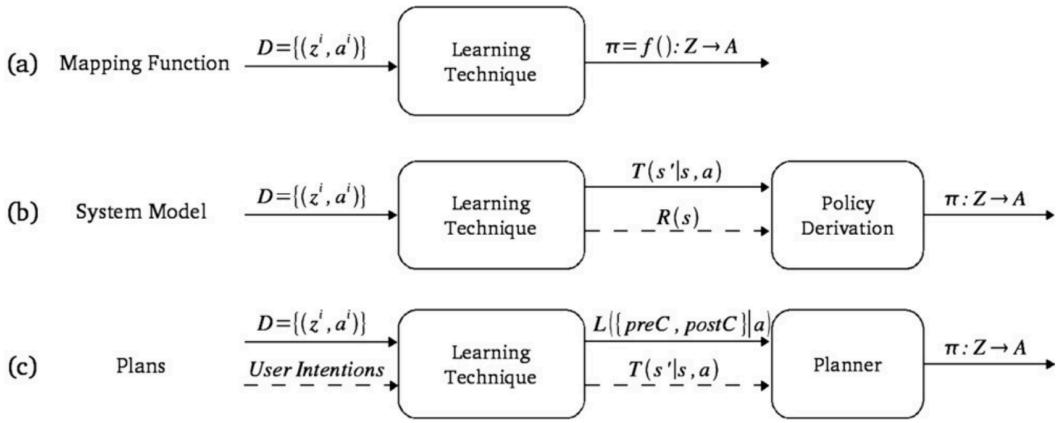


Figure 2: Policy derivation using the generalization approach of determining (a) an approximation to the state action mapping function, (b) a dynamics model of the system and (c) a plan of sequenced actions. Image and caption credit: Argall et al. [2].

However, problems remain with imitation learning approaches. A major issue is the *correspondence problem* [6], *i.e.*, the difference in embodiments between humans and robots. Meanwhile, from a neuroscience perspective, pure behavior cloning does not align with how animals learn. Mirror neurons activate both when a primate performs a goal-directed action and when they observe another primate carrying out the same action [9]. In fact, mirror neurons fire when a primate *understands* the action of another [9, 3]. Inspired by this, Liu et al. [5] perform imitation learning to achieve functional equivalence. Some properties of mirror neurons are also noteworthy - they are highly specialized

and action-specific, suggesting the existence of a neural codebook [9]. This aligns with the concept proposed in [1]. Other challenges include collecting demonstrations, stitching trajectories, and in-context and few-shot learning. We believe that rather than direct behavior cloning, greater focus on learning the underlying functionality and dynamics may be key to addressing these problems.

3.3 High-level cognition

High-level cognition is essential for transferring learned skills, combining different tools, tool selection, and tool manufacturing. From the perspective of FPICU (functionality, physics, intent, causality, and utility) [13], each domain is critical. Functionality provides understanding of what can be accomplished with a tool. Physics gives the dynamic model of the agent-tool-environment system. We express intentions as sequences of goals to be achieved through tools. For causality, we must determine the causal factors leading to functionality and how they enable achieving a goal. Learning utility or values can facilitate tool selection, high-level planning, and tool improvement. Notably, tool cognition should account for agents' embodiments. However, even a single type of causal tool use remains challenging for current approaches.

4 Conclusion

In conclusion, our discussion highlights that we are still at the very initial stages of developing true tool understanding in AI. The sophisticated tool use exhibited by animals like crows serves as an excellent example to motivate a paradigm shift in current AI. Significant advancements are needed across all levels – from basic perception and manipulation to high-level causal reasoning and planning. There remains a long path ahead, and it is time for use to learn from the crows.

References

[1] Anonymous. Pre-training goal-based models for sample-efficient reinforcement learning. In *Submitted to The Twelfth International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=o2IEmeLL9r`. under review. 4

[2] Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483, 2009. 3

[3] Valeria Gazzola, Giacomo Rizzolatti, Bruno Wicker, and Christian Keysers. The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage*, 35 (4):1674–1684, 2007. 3

[4] Gavin R Hunt. Manufacture and use of hook-tools by new caledonian crows. *Nature*, 379 (6562):249–251, 1996. 1

[5] Hangxin Liu, Chi Zhang, Yixin Zhu, Chenfanfu Jiang, and Song-Chun Zhu. Mirroring without overimitation: Learning functionally equivalent manipulation actions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8025–8033, 2019. 3

[6] Chrystopher L Nehaniv and Kerstin Dautenhahn. *Imitation in animals and artifacts*. MIT press, 2002. 3

[7] Kenneth P Oakley. Man the tool-maker. *Proceedings of the Geologists' Association*, 55(2): 115–118, 1944. 1

[8] Meiying Qin, Jake Brawer, and Brian Scassellati. Robot tool use: A survey. *Frontiers in Robotics and AI*, 9:1009488, 2023. 1, 2

[9] Giacomo Rizzolatti and Laila Craighero. The mirror-neuron system. *Annu. Rev. Neurosci.*, 27: 169–192, 2004. 3, 4

[10] Namiko Saito, Tetsuya Ogata, Satoshi Funabashi, Hiroki Mori, and Shigeki Sugano. How to select and use tools?: Active perception of target objects using multimodal deep learning. *IEEE Robotics and Automation Letters*, 6(2):2517–2524, 2021. 2

[11] Alex AS Weir, Jackie Chappell, and Alex Kacelnik. Shaping of hooks in new caledonian crows. *Science*, 297(5583):981–981, 2002. 1

[12] Yixin Zhu, Yibiao Zhao, and Song Chun Zhu. Understanding tools: Task-oriented object modeling, learning and recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 3

[13] Yixin Zhu, Tao Gao, Lifeng Fan, Siyuan Huang, Mark Edmonds, Hangxin Liu, Feng Gao, Chi Zhang, Siyuan Qi, Ying Nian Wu, et al. Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense. *Engineering*, 6(3):310–345, 2020. 4