A summary of the recent studies on physical tool use

Depei Bao

1. Review and some ideas on the papers by our group members

As far as I know, there are four papers/technical reports on general topics of tool use. They are:
St. Amant(2002), Bicici(2003), Wood(2005) and St. Amant & Horton(2007). Most other papers focus on a specific aspect or a restatement of the ideas in these papers.
In the seminal paper by St. Amant (2002), he discussed possible issues about the structure and mechanism of tool use in the physical world in four aspects: procedural structure, a taxonomy of tools, ecological issues and visual and motor issues. Then in Bicici (2003), tool use was further clearly defined as reasoning to get the functionality of tools, and the ideas in St. Amant (2002) were further elaborated under the problem definition. Wood(2005) discussed further the functional components in the design of a habile agent and its implementation on a robot platform. Finally, St. Amant & Horton(2007) gave a new definition of animal tool use contrasted with Beck's definition on the basis of past research.

The guideline for past research seems like first precisely defining tool use. Then we can view it as a problem, and hope to find a general structure in it. Moreover, if there is hardly a general structure in the general problem, we can separate the whole problem by different kinds and find a structure in each category.

Apparently, tool use is not an external problem (Note 1). It is a human/animal behavior and an exhibition of one's cognitive process. Thus, tool use can only be defined by patterns or modes of human/animal behaviors (e.g., a functional way). Only a small part of the definition is about human cognition, i.e., how does our mind decide the behavior of tool use? Because the problem itself is to find cognitive processes under tool use behaviors. Even those small parts can reach only the conceptual level but the mechanism level, e.g., tool use is goal-driven, but how do goals direct behaviors?

On the other hand, understanding of a concept is individual because concept definition itself is an exhibition of cognitive processes. The sense of a concept stems from individual knowledge and reasoning process. We hope a definition to cover all sides and cases of a concept. This is possible for defining some objective concepts like motion because perhaps some "gist" of the concept exists in the external world. But as for defining one's own behaviors, it's unlikely that a single neat definition can cover all cases and rule out all non-cases. For example, for scenario 2c in St. Amant & Horton (2007): A chimpanzee cups water in its hand and splashes its body to cool off. In the paper, this is not counted as tool use. But such intuition may be based on the assumption that tools are normally discrete objects. If we ignore the assumption and think in another way that a tool is an apparatus to extend our own ability, then it seems 2c is a tool use. Simply put, a single definition can not cover all behaviors of tool use and eliminate all others, because there are many ambiguous cases.

Since much of the past work describes tool use on the behavior level, some of them only reflect our ability to extract a principle from instances. However, in the underlying reasoning process for tool use, they are simply not considered. For example, in St. Amant(2002), procedural structure in tool use includes four parts: *direct action, amplifies existing behavior, goal-directed activity* and *effective behavior*. It's true that tool use behaviors are direct actions, but when it comes to how to use tools to reach one's goal, this may not be a consideration. Similarly, amplifying existing behavior may be considered on occasion, e.g., when use a stick to extend one's reach. But in other cases, this idea does not go into our mind. For example, we decide to use a hammer not because we don't think our hands can put a nail into the wall but because we match a hammer-nail pattern. Even if we use a block of iron instead of a hammer, it's most probably the result of analogy. Finally, the last two parts actually raise problem rather than provide solution. For goal-directed activity, the question is how minds find an appropriate way to bridge the gap between the goal and the current situation. Effective behavior involves how minds reach an effective tool use, or more generally, how minds know the effective way to reach a goal. Two interesting views are force transformation and geometrical transformation. These two transformations appear in reasoning processes since we do have force sense and spatial geometry sense in our reasoning.

A similar case occurs in the effort to categorize tools. A single taxonomy is sure to follow some principles which can only represent one side of reasoning. For example, the basic principle of tool use taxonomy in St. Amant (2002) is functionality of tools, which reflects the potential effect of tools we can expect (a consideration in reasoning). How about a categorization in terms of manners that a tool is used? Or about the domain a tool is involved? Admittedly, all of these categorizations are ambiguous because functionality, manner and domain are abstract concepts which are hard to define precisely (this problem will be discussed later), and a category in one taxonomy may constrain categories in another taxonomy. But the point here is a universal hierarchical categorization can not accommodate all thinking process. Thus, concepts (Note 2) are not organized in that simple way -- the meaning of a concept is closely related to our ways of thinking.

However, a concept in our mind does show such categorical and hierarchical properties. The meaning of a word or a phrase does have a comparison implication with other entities in the same category. For example, the meaning of RUN is not merely described by its own properties but also by its comparison with other similar ways of behavior, say, WALK. Also, a concept also has the sense of its upper level concept, which exhibits a hierarchical way for concept organization. Thus, a hierarchy can be seen as a part of the entire organization. These discussions raise two problems to be solved. One is what the static concept organization is. How it is stored in memory. The other is what the dynamic search method is. How can searching appropriately locate the concept we use to describe a problem and its solution. The meaning of a concept may be implied in these two parts. When it comes to reasoning for tool use, these two problems will also be critical.

Why a concept system is so important to problem solving? There are many reasons, perhaps one of them is a concept system is critical to represent the problem domain. When we do reasoning, we can find a very suitable representation of a problem without much effort (this is not our language but an internal representation). Such representation relies on how we associate the semantics of related concepts with the external world we want to describe. It is impossible to tell how reasoning is carried out until such representation mechanism is established, or part of reasoning may be actually in representation mechanism (Note 3). We all have experience that after find a good perspective to see a problem, e.g., a mathematics problem, the reasoning to solve it is handy. Surely, from a cognitive perspective, reasoning is not homogenous, which may include many different processes and mechanisms. I will discuss this latter.

This point that representation has priority over reasoning can be illustrated by expanding Bicici (2003) which proposed several issues in reasoning the functionality of tools. The problem is how do we represent functionality of tools? It's obvious when we talk about functionality we have many distinct ways to express it. The function of a hammer may be "pound a nail into wall" or "a lever". Also, a simple abstract level description may not be sufficient to describe the "functionality". A hammer's size also determines its functionality. The current situation which decides if it can be used could also be part of its functionality. After all, we may express it in many a manner in various situations. When an agent reasons about the functionality of a tool, how does he describe the result of his reasoning? The description is dependent on his purpose and his understanding of the circumstance. When the purpose is just answering a question or planning, the abstract level description might be enough. But when it comes to actually use a tool, more details of representation should be involved.

A recommendation in Bicici (2003) for further research also reflects the purpose to represent the functionality in a unique way:

"Therefore, we need to search for where these tools come from and what is the underlying functionality that we achieve while using them. For example, we can think of a hammer as a tool that changes the direction of the force and the momentum applied to it and we can figure out that its functionality is based on the basic functionality of a lever. The human body has itself many levers; for this reason classifying tools according to their lever types seems appealing."

"Underlying" here may be misleading. For one thing, it's hard to get exactly what the "underlying" is. For the hammer example, it is a restatement of the functionality in physics. Can this translation meet all description needs for using a tool? There are too many "underlying" things to consider. When we extend our reach by stick, we use many "underlying" properties of a stick, like shape, rigidity, length, etc. Even there are many other properties which may not have directly to do with its functionality, like we won't put sugar stick into water. Sometimes, the "underlying" is not only represented qualitatively but also quantitatively (e.g., length of sticks). For the other thing, reasoning on higher levels may need different information on lower levels, i.e., even if the "underlying" does exist, it is not unique. As for hammer use, few of us think it as a lever as we consider if use it. Normally, we first consider if attaining our goal needs the function of the hammer on an abstract functional level (though the functionality is not described uniquely, which depends on the representation of the current situation and one's goal), and then evaluating if hammer head matches, say, nail will go into our mind. Meanwhile, other issues like if I have enough strength to raise the hammer may also get the attention. But all of them are considered on a conceptual level. When it comes to actually using it, reasoning on lower levels (e.g., motion control) might adjust our control system the same way we use lever, but it's by no means to say the “underlying” to use a hammer is a lever.

Another important notion mentioned both in St. Amant (2002) and Bicici (2003) is affordances of tools which is the “designed relationships between their physical/dynamic properties and the properties/abilities of their intended users”. The papers also point out “the Physical affordances (are) closely related to constraints” and “Tool use involves establishing and exploiting constraints between the user and the tool, the user and the environment, and the tool and the environment”. All these demonstrate that “affordances” is a specialization of constraints to tool use. The problem is if this constraint viewpoint can help us find the principle in reasoning for tool use.

It’s true that there exist constraints between tools, tool users and environments. However, I have to argue that knowing what constraints are is the universal cognitive capability of human beings. All our perception, reasoning and judgment just reflect constraints of the external world. Human have a very good insight to these constraints, can reason in terms of these constraints, and can act under these constraints. In another word, constraints are how knowledge is represented and organized and how reasoning proceeds. However, at a certain moment, we can’t consider every constraint of the world, human develop a very good mechanism to narrow down which constraint is relevant to our goals and current situations. How this mechanism works is the real problem, which is, in turn, related to how knowledge is organized and how reasoning proceeds. Perhaps the constraints notion can give us a clue for knowledge representation and organization.

The above discussion about constraints focuses mainly on cognition on higher levels. But on lower levels, especially motor control (this issue is very important since our research is on physical tool use), constraints will be represented in a way different from those on upper levels. One example is, for high-level reasoning, a constraint is perhaps just expressed as “can touch”. But when it comes to low-level motor control, there is more constraint information needed for “touch”, e.g., the permitted force. It’s apparent that when we go to touch an object, we don’t consciously know the information and the reasoning on lower levels. However, we do get this information from our sensor to enable an appropriate muscle control. How low-level information is represented and what information is needed for our control system? These two questions need further research.

Spatial self-representation takes a critical role in motor control system since it determines the interaction between body and environment. In Wood (2005), the robot’s body is represented as a “Four-Segment Model”, and is controlled by the MMC net. This scheme may oversimplify motor control system. First, do we have a central physical body model? When touching an object by hand, I don’t need to calculate to set the entire body state. We just adjust a subsystem for arm control. In another word, we don’t have a centralized control system. Control is distributed across several subsystems. Second, the organization and coordination of these subsystems have to do with upper-level cognition. If my goal is touching an object by hand, I know I should use arm control subsystem. Third, is body represented as a geometric model? This is open to debate. I speculate that body constraints are implied as parameters in those subsystems.

In summary, we don’t have a centralized body representation. Instead, such “representation” spreads over our mind and control system. Our control system is a distributed and hierarchical organization (those subsystems are also built layer by layer) governed by cognition system. In upper level cognition, body representation is likely to be organized on conceptual level while in lower level, body constraints are implied in subsystem parameters.

2. What’s a possible solution look like?

2.1. Problem constraints

By the previous discussion, tool use is a human behavior above human cognition. To find the way of using tools requires an in-depth explanation of human cognition. The problem seems huge but we can still limit the tool use problem from several dimensions.
First, the majority of reasoning for physical tool use is only relevant to physical domain, which can drastically reduce the magnitude of knowledge base. Second, for a first trial, tool use scenario can be restricted to one scene, i.e., goals and the way to reach it have only to do with the present situation. For example, an agent won’t think he can buy a hammer from a grocery when he wants to knock in a nail. He can only try to find a hammer-like tool at the scene. Third, from a decision making perspective, a tool use can be seen as the result of answering the following four questions:

-*What the effect the subject wants to achieve (the description of subject’s goal)*
-*Does the subject need tools?*
-*Are there tools to rely on?*
-*How can the subject use the tools to reach the goal?*

The answer to these questions is determined by the environment, tools’ properties, and the subject’s capability. But when these constraints are applied to answer these questions relies on the cognitive processes we take. Moreover, how the constraints are found depends on our knowledge and current understanding of the situation. I will try to explain these below.

2.2 The elements of cognition

There is no doubt that human behavior is goal-directed. However, motivations behind goals are heterogeneous. Some of them stems from our instinct and body need, which is beyond cognition. Other’s can almost be explained by cognition. A simulated agent has no self demand. So we can explain his motivation solely by cognition. Also, he can only take orders from a user, which can be seen as the first drive of his behavior (behavior in real life will be far more complex). A goal makes sense only at a specific situation or context, and a specific situation might in turn motivate a goal. So situation awareness (the term implies the meaning) can be seen as another elements. The third is knowledge (Note 2). Interactions among these three elements decide our behaviors.

The most common theoretical framework of situation awareness is provided by Endsley (1995), which illustrates three stages of SA formation: perception, comprehension, and projection.

SA starts from percepting status, attribute and dynamics of relevant elements in the environment. Elements, e.g., objects, events, people, systems and other environmental factors with typical states and attributes like locations, conditions, modes, actions, etc. Then on the second level—comprehension, these fundamental elements are recognized, interpreted and evaluated under user's goals and objectives. Finally, a projection phase extrapolates information like status knowledge, dynamics and comprehension results of level 1 and 2 forward in time to predict the further status. This three-level SA is put forward for decision making problems, so two steps after SA is decision and action. After that, the change to the environment feeds back to perception to initiate a new SA cycle.

Endsley's SA framework is powerful and can be used for many applications. However, it is a simplified model which doesn't go in depth about the interactions between SA, goals and knowledge. The next section will explain these interactions.

2.3 Interactions between cognition elements

Realization of goals and selection of knowledge will depend on the understanding of current situation. Thus, SA lies in the heart of their interactions. However, what is situation? Current perception may not be current situation. Suppose my goal is going to school when I am at my kitchen, my current situation related to my goal is at my house rather than at the kitchen. This is a representation problem under the goal I have. This problem can be set aside for a while if the problem is restricted to one scene, i.e., only what the agent can perceive has direct relation with his goal. For example, only in the kitchen has to do with my goal or it doesn't matter where I am. Limitation to one scene often makes the representation unique.

In Endsley's SA framework, goals are regarded as one-way determinants of comprehension of current situation (level 2). However, her framework does not intend to explain human behaviors, instead, its purpose is providing information for human decision making. Thus, one or several goals in her model exist independently of current situation. However, comprehension of current situation may yield new goals which may or may not have to do with the agent's original goals. There are many scenarios to illustrate such bidirectional determination between situation awareness and goals.

-*1. I am cooking in my kitchen. All of a sudden, the fire alarm rings. I have to run out of my house.*
This can be seen as I detect the sound at perception, and interpret it as fire alarm. Then a new goal (run off) is derived without direct relationship with cooking. The reason for such interpretation and action is perhaps some routine.
-*2. A bowl almost fall off, I place it back.*
I detect this event and have a sense of the result of the event. Then my goal is to prevent it happening.
-*3. I know I have a bowl at hand, but suddenly I hear a cracking sound and I realize that my bowl is broken.*
I expect that the bowl's position does not change, but my perception is against my expectation, so my attention is drawn to it. This is different from the first one mainly because the bowl is an

object of my present awareness while the fire alarm is not an element of perception, though I know I have a device.

-4a. I need to use a turning shovel while I am cooking (goal 1), 4b. I need to find it (goal 2).

Goal 1 might be a subgoal of another goal and motivated by a specific situation. At the next cycle, I realize I don't have a shovel at hand, which yields goal 2.

-4c. I remember where I put the shovel last time, so I search there.

To find the appropriate action for the goal, past situations in memory may also be of use.

-4d. I have not been in the kitchen before, so I can only search several possible places for it.

This behavior pattern works as follows: I will choose an action in a certain manner, which may be from clues by the current situation and a pattern matching. At the next cycle, after the action I may gain more information, and choose another action until I reach the goal or I have a new goal (Note 4).

These scenarios show that goals are continuously raised and attained, which works as a scheduling system. What is a possible framework to integrate this goal scheduling into the situation awareness? A good idea may consider how the goal is raised and attained at various stages of SA.

2.4 A new situation awareness framework

In the new model, there are only two stages. The action selection is at the comprehension stage. Thus, projection of action results is also at this stage. Extrapolation of other environmental elements happens at the phase two as well since projection can raise new goals. At the perception stage, it is impossible to fire new goals since all perceived are primitive elements. Thus, all goals arise at the comprehension stage. Another difference is that the end of a cycle may not be an action. If a new goal is fired at the second stage, no action will be taken since we need the next comprehension cycle to understand the situation concerning the new goal to choose an appropriate action. Also, an emergency will raise a goal to react and I may overlook what I perceive right now.

There are several possibilities to raise a goal. First, I may comprehend an event as a pattern to trigger a reaction (scenario 1, a sound is explained as fire alarm). Second, an event's result is projected, and it is against one's wish (some built-in rules or habits). The new goal is to prevent it (scenario 2, a bowl is almost dropped). Third, at the previous stage, all perceived elements are projected according to their dynamics (at the second phase). In this cycle, however, an event which is against expectation occurs, this will draw one's attention (scenario 3, a bowl has been dropped). A specific situation pattern may raise a goal (scenario 4a, I need to use a shovel). If there is a gap between a goal and the current situation, a subgoal is derived to find an action to resolve the gap (scenario 4b, I need to use a shovel but I don't have it, so I have to find a way to have it). At the next cycle, I may find a clue from past experience (scenario 4c) or I just choose an action according to my habits—a behavior pattern (scenario 4d). Then I will take the action. Meanwhile, I have an expectation of my action, so once after my action, I don't reach my expectation, I will choose another action or just give up.

Further Research

The framework in the second part addresses how an agent interacts with the external world. It does not specify sensor and motor mechanism. Also, it does not give how the knowledge is organized. Knowledge organization is important in the actually instantiation of the framework. To be specific, knowledge is important in three aspects. First, knowledge organization partially determines what a pattern is and how a pattern gets attention rather than another. In another word, the reason for a way of thinking lies in the knowledge organization. Second, what knowledge is needed for tool use? Since knowledge in physical tool use can be limited mainly to physical domain, it is relatively easy to organize. Third, as note 2 put, knowledge is not homogeneous, the interaction between different sorts of knowledge is also relevant to knowledge organization.

In a word, further research on tool use related knowledge organization will focus on two issues: 1. how the knowledge is organized to give meanings. 2. how the knowledge is organized to give our ways of thinking.

Notes

Note 1. In my opinion, there are two sorts of problems in AI. One exists out of our mind. Most of problems in a specific domain are fallen into this sort, such as chess and 8-puzzle. These problems have their own problem structures independent of human ways of thinking. So the solutions to them may be independent of human ways of thinking. On the contrary, some problems are the characterizations of human behaviors, e.g., general purpose problem solving. For such problem, there is no precise definition of problem domain. Thus no special structures exist in these problems.

Note 2. Concept organization is considered a part of knowledge. From a cognitive perspective, knowledge is not homogeneous. Knowledge can be classified into several kinds (so to reasoning) since reasoning is a process to use heterogeneous knowledge. Research on memory shows the possible types of knowledge.
The known memory types are listed below:
Long-term memory: Procedural memory/Implicit memory · Declarative memory/Explicit memory (Episodic memory, Semantic memory)
Short-term memory: Working memory
Sensory memory: Visual memory · Iconic memory · Echoic memory
Concept organization is regarded as the organization of semantic memory.

Note 3. Bicici (2003) mentioned Sowa's approach or an "egg-yolk theory of word meaning" for organizing concepts and recognizing objects, which classifies an object into a concept category based on similarity. The question is if recognition is just assigning an object to a concept. Imagine one seeing a chair. Do minds already classify it as a chair? It appears like that. However, I think the contrary since an object can belongs to numerous categories. Thus, concept assignment might be goal-oriented, i.e., when we perceive an object, we don't assign it into any specific concept category until we have some purpose to do so. For the chair example, after we have the motivation to sit, then we recognize a chair as something "sittable" rather than a "chair". When we want to know what the object is, we may assign it as a chair. Though both of them rely on the object's shape to classify, the former is on the basis of object's functionality while the latter is solely shape recognition. We seldom recognize a sofa as a chair, but when it comes to sit, both of them are "sittable". This also implies that we may recognize an object as various concepts in different levels of conceptual organization.

Note 4. These scenarios may imply a new view on planning or problem solving. Planning used to be seen as a search problem with action alternatives. Actions or operations have a result of the state. The search is to find a sequence of action to reach a final goal state. This view seems ignoring that the results of actions are also goals. Human behaviors, especially our everyday routines do not follow this explicit planning procedure. Rather, every action is selected in terms of the understanding of the current situation, and a new sub-goal derived from a pattern matching or reasoning on the comprehension level of situation awareness. Sometimes, it seems we do a planning in our mind. But inspect closely, this planning is perhaps an illusion. For everyday routine, we don't explicitly know what I should do next to reach my goal. Most of the action is taken without conscious thinking. Thus, the sense of knowing how to do something is more like a pattern stored in our memory. When some subgoal is derived e.g., I need to find a shovel, we reach the subgoal by pattern matching based on the understanding of the current situation to choose an action. To see this from another perspective, we don't plan all actions in advance. Instead, the current situation may motivate subgoals along my way to reach the final goals. For conscious thinking process, I speculate that our planning can be seen as a reflection of lower level pattern matching on a certain description level (e.g., when we consider how to go to school, I don't thinking I should first put on my shoes). We implicitly assume the achievement of an action, and then find a pattern based on the ideal situation I've imagined (ignoring some details).

Reference

1. St. Amant, R. A preliminary discussion of tools and tool use, April 25, 2002. (Also available as NCSU technical report TR-2002-06.)
2. Bicici, E. and St. Amant, R. Reasoning about the functionality of tools and physical artifacts. Technical Report TR-2003-22, Department of Computer Science, North Carolina State, 2003.
3. AB Wood, TE Horton, and R. St. Amant, "Effective tool use in a habile agent", In IEEE Systems and Information Engineering Design Symposium, 2005, pp 75
4. St. Amant, R., and Horton, T. E. (In press). Revisiting the definition of tool use. Animal Behaviour, doi: 10.1016/j.anbehav.2007.09.028.
5. Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors 37(1), 32-64.