Running head: Consciousness and ACT-R

Consciousness in the ACT-R architecture

Niels A Taatgen

Carnegie Mellon University

Consciousness in the ACT-R architecture

ACT-R (Adaptive Control of Thought, Rational) is an architecture of cognition (Anderson *et al.*, 2004), which is used to simulate and explain human performance and learning on a wide range of tasks, from typical psychological experiments to complex, dynamic tasks. The constraints that have shaped the architecture are functional (what cognition should be capable of), behavioral (the simulation predictions have to match human data) and, more recently, neuropsychological (components of the architecture correspond to areas in the brain). The consequence of this is that little attention has been paid to more global phenomena like consciousness. Nevertheless, two aspects of ACT-R are relevant for the study of consciousness. The first is what parts of the knowledge representation are open to awareness, and the second is the distinction between implicit (unconscious) and explicit (conscious) learning.

Awareness in ACT-R

ACT-R is made up of several modules (visual, manual, intentional, declarative, procedural, and others) that can all operate independently (Figure 1). These modules communicate through buffers, each of which can hold a single piece of information. The system can be considered “aware” of the contents of all these buffers, i.e., it is aware of the currently attended visual stimulus, it is aware of the current action that is being taken, it is aware of the current goal, and the currently active fact in declarative memory. A central part of the theory is taken by two long-term memory systems: procedural and declarative. Declarative memory is open to inspection, meaning that when the system successfully retrieves knowledge from declarative memory it is then aware of this knowledge. Procedural knowledge, on the other hand, is not open to direct conscious inspection. The awareness status of the other modules is less well defined: it is quite likely that awareness in the visual module should be beyond the single item that is in the visual buffer, and there is a notion that there is awareness in the form of a *feeling-of-knowing* when declarative retrieval is busy or not entirely successful. Nevertheless the overall picture of awareness in ACT-R is one in which the focus of awareness is on the communication between systems, and in which both processing in the central production system and in the modules is unconscious. In that sense it is consistent with Jackendoff's notion of intermediate-level consciousness (Jackendoff, 1987). According to Jackendoff, people become aware of perceptions after some levels of processing, for example the 2.5D representation in visual perception, or a phonological representation in language perception. He also states that people are unaware of their core cognition processes.

Implicit versus Explicit learning

The difference between implicit (unconscious) and explicit (conscious) learning is traditionally explained by assuming separate memory systems for implicit and explicit knowledge (Tulving et al., 1982). Although this seems to map very well onto ACT-R's notions of procedural and declarative memory, it is not consistent with the way ACT-R learns from experience. Many models of implicit learning use storage and retrieval of examples in declarative memory (instance theory) to explain that performance improves without participants being able to explain why (Taatgen and Wallach, 2002). An alternative account of implicit learning is that all learning is basically implicit (consistent with ACT-R's notion that learning is unconscious), and that explicit learning consists of (learned) strategies that are tied to goals (Taatgen, 1999). This means that a certain type of explicit learning is only possible if the knowledge that implements that strategy is available, which explains why individual differences in explicit learning are so large. It can explain that people are aware of their learning, because the acquired knowledge is associated with a learning goal. For example, *rehearsal* (repeatedly retrieving an item from declarative memory) is an explicit learning strategy to drive the implicit learning mechanism that increases activation of an item each time it is retrieved.

This way of looking at the distinction enables explanations for several of the implicit learning phenomena, for example an explanation for Tulving et al.'s (Tulving et al., 1982) memory experiment (Taatgen, 1999). In that experiment participants had to learn a list of words, and were tested after an hour and then after a week explicitly (by being asking whether a particular word was on the list), and implicitly (through a word-completion task in which words were sometimes from the list and sometimes not). Performance on the explicit task turned out to decrease after a week, but performance on the implicit task remained constant. In ACT-R the explicit task requires the creation of a separate memory chunk to record that a word has been studied in the context of the experiment, while the implicit task only relies on the word having received extra action by studying it. The experimental results can be explained by the fact that the new memory chunk decays quickly and cannot be recalled after a week, but the extra activation still gives the word chunks an edge in the competition with other words.

References

ANDERSON, J. R., BOTHELL, D., BYRNE, M., DOUGLASS, S., LEBIERE, C. & QIN, Y. (2004) An integrated theory of cognition. *Psychological Review,* 111**,** 1036-1060.

JACKENDOFF, R. (1987) *Consciousness and the computational mind,* Cambridge, MA, MIT-Press.

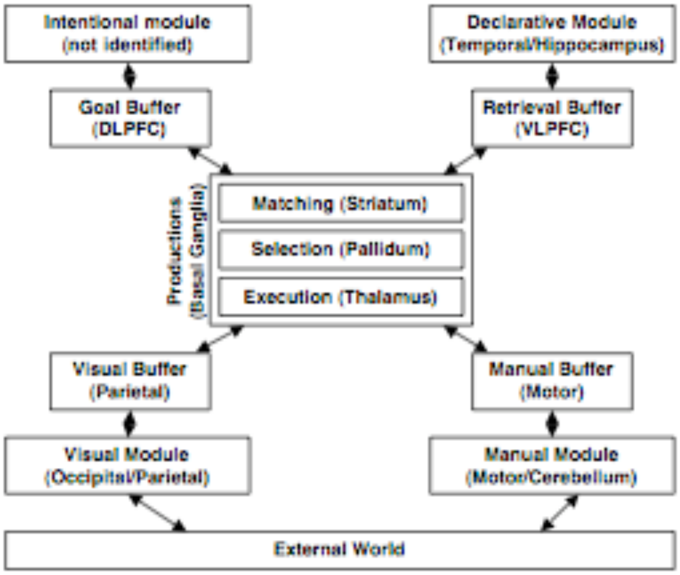
TAATGEN, N. A. (1999) Implicit versus explicit: an ACT-R learning perspective; Commentary on Dienes & Perner: A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences,* 22**,** 785-786.

TAATGEN, N. A. & WALLACH, D. (2002) Whether skill acquisition is rule or instance based is determined by the structure of the task. *Cognitive Science Quaterly,* 2**,** 163-204.

TULVING, E., SCHACTER, D. L. & STARK, H. A. (1982) Priming effects in word-fragment completion are independent of recognition memory. *Journal of experimental psychology: learning, memory, and cognition,* 8**,** 336-342.

Figure Captions

Figure 1. An overview of the ACT-R architecture


Intentional module
(not identified)
Declarative Module
(Temporal/Hippocampus)
Goal Buffer
(DLPFC)
Retrieval Buffer
(VLPFC)
Productions
(Basal Ganglia)
Matching (Striatum)
Selection (Pallidum)
Execution (Thalamus)
Visual Buffer
(Parietal)
Manual Buffer
(Motor)
Visual Module
(Occipital/Parietal)
Manual Module
(Motor/Cerebellum)
External World