Biologically Inspired Cognitive Architectures (2014) **8**, 132–139

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bica

INVITED ARTICLE

Between architecture and model: Strategies for cognitive control

Niels A. Taatgen [*]

University of Groningen, Institute of Artificial Intelligence, Nijenborgh 9, 9747 AG Groningen, Netherlands

Received 14 February 2014; received in revised form 20 March 2014; accepted 24 March 2014

KEYWORDS
Cognitive transfer;
Skill acquisition;
Cognitive control;
Cognitive architecture;
Cognitive training

Abstract
One major limitation of current cognitive architectures is that models are typically constructed in an ''empty'' architecture, and that the knowledge specifications (typically production rules) are specific to the particular task. This means that general cognitive control strategies have to be implemented for each specific model, which means a lack of consistency and constraint. Alternatively, cognitive control can be implemented as a part of the architecture itself, which is often implausible, because many strategies are learned and differ among individuals. A third solution is to assume cognitive control consists of learned strategies that can be used for many different tasks. The PRIMs theory (Taatgen, 2013) provides a modeling framework for this type of reuse. The approach is discussed using the example of working memory control in which I show that three different working memory tasks share the same strategic components to store and recall items. The broader impact of the work is that it shows that general cognitive skills may play a much more important role in understanding many aspects of cognition that are traditionally considered to be part of either the cognitive architecture or specific task knowledge, and therefore provides an important stepping stone towards the larger goal of unified theories of cognition.

Introduction

The enterprise of cognitive architectures is connected to grand ambitions. Both Newell (1990) and Anderson (1983) proposed the cognitive architecture as the great unifier in cognitive science and cognitive psychology. To underline these ambitions, Newell proposed a list with 13 items in his 1990 book that cognitive architectures should strive to accomplish. I will not reiterate the list here, but it contains items such as ''Behave robustly in the face of error'', ''Use vast amounts of knowledge'', up to ''Arise through evolution''.

---

[*] Tel.: +31 503636435.
E-mail address: n.a.taatgen@rug.nl

http://dx.doi.org/10.1016/j.bica.2014.03.010
2212-683X/

By the end of the book Newell revisits the list, and concludes 6 of the goals have been satisfied (at least in some fashion). Anderson and Lebiere (2003) revisit the list, calling it the *Newell Test*, and conclude their ACT-R architecture satisfies 5 out of 12 criteria, pitting it against connectionism, which they also rate at 5 out of 12 (although on different items). Despite the seeming lack of progress on the Newell test scale, cognitive architectures have made great strides in providing explanations for a wide variety of cognitive phenomena. The huge body of publications on the ACT-R website (act-r.psy.cmu.edu) shows that researchers using ACT-R do not shy away from any area of experimental psychology. As a consequence, models based on cognitive architectures have been successful in being published in mainstream cognitive science and psychology journals, and play a role in the development of cognitive theory. However, because of the alignment with standard practice some of the original goals have been neglected, which may be why progress on the Newell Test has been limited, because it has more ambitious goals than cognitive psychology generally pursues.

A typical modeling paper discusses a particular phenomenon using experimental data and a model of these data, and proposes a particular explanation or theoretical position. Although this is much better than verbal explanations, this general research strategy is one that at least Newell argued against strongly (Newell, 1973). According to him, psychology would never make progress in understanding cognition if it would persist in studying single phenomena using a research paradigm that he mockingly called ''The Twenty Questions Game''. He pointed out that pursuing this ''Game'' would be a fine choice to advance your career as a scientist, but would, in the end, not advance science. Unfortunately, it seems even cognitive modelers have partially fallen into this trap. The main problem is that models are constructed for particular experimental tasks or paradigms, but that generalization from one model to another is often very limited. Part of this problem lies in the current cognitive architectures themselves, because they promote thinking about cognition in terms of separable tasks.

Task-specific models

The construction of a cognitive model for a particular task starts with an ''empty'' cognitive architecture. The modeler adds knowledge components that are specific to that task, or, in the case of learning models, trains the model using items specific to that task. The underlying assumption is that a real cognitive system already has a large body of knowledge about other tasks, but this knowledge has no or limited influence on the new task, and can therefore safely be omitted.

As a consequence, the common element between different models within a particular cognitive architecture is just the architecture itself: its representations, its mechanisms to handle knowledge, and its modules to interact with the outside world. The assumption of these architectural elements is that they are the same, no matter what the task is. Even stronger, there is the assumption that all architectural components are innate, because they are properties of the brain itself.

This leaves a gap for a third component: knowledge that is not part of the architecture, but also not task-specific. In particular I want to focus here on procedural knowledge, because the importance of more factual general knowledge has already been acknowledged and discussed elsewhere (e.g., Salvucci, 2013). A first question concerning general procedural knowledge is one of transfer: if we already know a particular text editor, is it easier to learn a new one? The answer is yes, much easier (Singley & Anderson, 1985). It is also quite plausible that skills build upon each other. For example, it is easier to learn multicolumn subtraction after learning multicolumn addition first, even though the two differ enough to require separate specific rules.

But is there reuse of procedural knowledge beyond tasks that resemble each other? One domain to look at in more detail is what is generally referred to as *cognitive control*. This includes the handling of goals and tasks, organization of working memory, activating relevant information and suppressing irrelevant information, interleaving multiple tasks, and handling interruptions, among others. There is always a certain awkwardness in how cognitive architectures handle cognitive control. It is a challenge to make control part of the architecture, because this implies it operates in the same way in every possible situation, which it often does not. On the other hand, if control is part of the (task-specific) model, control feels ad-hoc, and ''programmed'' by the modeler. Let me give two examples.

Organization of goals

In production system models, the organization of goals is sometimes handled by a goal stack. Although it is a convenient mechanism that works very well in many models, it has several problems. One category of problems is functional. Sometimes problems are not suitable for goal stacks, in tasks where goals are created and abandoned often, requiring great efforts in breaking down and rebuilding the goal stack. A second problem is behavioral plausibility: the human cognitive system does not act as if it implements a perfect goal stack. For example, both Altmann and Trafton (2002) and Anderson and Douglass (2001) have shown that errors people make in the iconic goal-stack task, the Towers of Hanoi, are not consistent with an architectural goal stack. Instead, the handling of goals in their alternative models uses a strategy that partially resembles the goal stack, but that is part of the task model. Although this solution is more satisfactory in the sense that it accounts much better for the empirical data, it is also implausible that the handling of goals has to be reinvented for every single new task.

Working memory control

A second example is working memory control. There is an ongoing debate in cognitive psychology about the nature of working memory, but researchers almost all take the architectural stance that working memory is a structural component of our information processing system (e.g., Baddeley, 1986). Nevertheless, many aspects of working memory are not part of the architecture. For example, the use of rehearsal to maintain elements in memory is unlikely to be architectural, because young children do