Cooper, R. P. (2006). Cognitive architectures as Lakatosian research programmes: two case studies. *Philosophical Psychology* **19** (2) 199-220.

To appear in *Philosophical Psychology*.
Submitted 10/06/04; Revision submitted 01/04/05; Accepted 13/07/05.

Cognitive Architectures as Lakatosian Research Programmes: Two Case Studies

Richard P. Cooper
School of Psychology
Birkbeck, University of London

Abstract: Cognitive architectures—task-general theories of the structure and function of the complete cognitive system—are sometimes argued to be more akin to frameworks or belief systems than scientific theories. The argument stems from the apparent non-falsifiability of existing cognitive architectures. Newell (1990) was aware of this criticism and argued that architectures should be viewed not as theories subject to Popperian falsification, but rather as Lakatosian research programmes based on cumulative growth. Newell's argument is undermined because he failed to demonstrate that the development of Soar, his own candidate architecture, adhered to Lakatosian principles. This paper presents detailed case studies of the development of two cognitive architectures, Soar and ACT-R, from a Lakatosian perspective. It is demonstrated that both are broadly Lakatosian, but that in both cases there have been theoretical progressions that, according to Lakatosian criteria, are pseudo-scientific. Thus, Newell's defence of Soar as a scientific rather than pseudo-scientific theory is not supported in practice. The ACT series of architectures has fewer pseudo-scientific progressions than Soar, but it too is vulnerable to accusations of pseudo-science. From this analysis, it is argued that successive versions of theories of the human cognitive architecture must explicitly address five questions to maintain scientific credibility.

Keywords: Cognitive architecture; ACT-R; Soar; Falsification; Theory development

1. Introduction

The past twenty years of research within cognitive psychology has witnessed the development of a number of cognitive architectures, such as Soar (Laird *et al.*, 1987; Newell, 1990), ACT-R (Anderson, 1993; Anderson & Lebiere, 1998; Anderson *et al.*, 2004), 4-CAPS (Just *et al.*, 1999; Newman *et al.*, 2003) and EPIC (Kieras & Meyer, 1997; Meyer & Kieras, 1997). These are task-general theories of the structure and function of the complete cognitive system. They are general theories of how perceptual, cognitive and motor processes interact in producing behaviour, rather than specific theories of behaviour on a single task (e.g., the Stroop task) or behaviour in a single functional domain (e.g., working memory). Cognitive architectures seek to unify single-task theories by abstracting from those theories and specifying general computational or information processing mechanisms that hold across tasks. Equally they seek to unify single-domain theories by embedding them within a complete theory of cognitive processing. Once developed, a cognitive architecture can then be applied to a specific task by providing task-specific knowledge, in much the same way as a general purpose computer can be applied to different tasks (word processing, accounting applications, game playing, etc.) through the execution of different computer programs.

While some cognitive architectures have proved to be very successful in providing a task-general substrate to support the execution of a wide variety of specific cognitive behaviours (most notably ACT-R: Anderson & Lebiere, 1998; Anderson *et al.*, 2004), architectures themselves are sometimes argued to be more akin to frameworks or even belief systems than scientific theories (e.g., Hunt & Luce, 1992; Vere, 1992). Two factors contribute to this view. First, architectures can be used as general purpose programming languages to implement single-task theories, sometimes with little apparent input from the architecture. Second, architectures cannot be falsified in any simple-minded way. Since behaviour on any particular task is held to be the result of the interaction of a task-general architecture with task-specific knowledge, potentially falsifying empirical findings can never be unambiguously attributed to the architecture and can always be attributed to an erroneous theory of how the task is carried out within the architecture.

Newell (1990), in his presentation of the Soar cognitive architecture, was aware of both of these criticisms. With respect to the second, he argued that architectures should not be viewed as theories subject to Popperian falsification, but rather should be viewed as Lakatosian research programmes with success based on cumulative growth. The issue is one of scientific credibility, for this is what falsification is supposed to bring to a theory, and Newell's appeal to Lakatos is justified in the sense that Lakatos (1970) demonstrated that much scientific progress in physics and chemistry did not conform to a simple Popperian model of scientific theories being proposed, falsified, and discarded. The Lakatosian view is instead that theories develop over time, partly as theoreticians lay out a simplified theory and then elaborate its details, and partly in response to the results of empirical studies testing the fringes of the theory.

Newell's argument, however, is hollow because he failed to demonstrate that the development of Soar, his own candidate architecture, adhered to Lakatosian principles. In short, Newell's appeal to Lakatos was not explicitly reflected in the methodology of theory development adopted by the Soar community, and appeared to serve little more than to divert the attention of falsificationists. This interpretation is reinforced by the apparent rejection of Lakatosian principles in Newell's later writings (Newell, 1992). This paper aims to ground Newell's appeal to Lakatosian principles in the development of cognitive architectures in general and Soar in particular by presenting detailed case studies of the development of two cognitive architectures—Soar and ACT-R—from a Lakatosian perspective. It is demonstrated that while both are broadly Lakatosian both have also witnessed historical developments that, according to strict Lakatosian criteria, are pseudo-scientific. Such progressions have been relatively frequent in the case of Soar, so Newell's appeal to Lakatosian principles in defending the scientific credibility of Soar is not supported in practice. The ACT series of architectures fares better than Soar under a Lakatosian analysis because a greater proportion of theoretical progressions within the ACT series have been scientific. Nevertheless, the ACT series is vulnerable to accusations of pseudo-science because some progressions within the series have been pseudo-scientific and, while it has adhered to Lakatosian principles, that adherence has been implicit rather than explicit. The implication of this analysis is that explicit adherence by the developers of ACT to such principles is necessary to maintain the architecture's scientific credibility and theoretical growth. More generally any cognitive architecture must adhere strictly and explicitly to appropriate scientific demarcation criteria (such as those of Lakatos, 1970) if it is both to develop as a theory and to maintain scientific credibility as it does so.

2. Critical Features of a Lakatosian Research Programme

A key area of debate within the philosophy of science in the mid-twentieth century was the differentiation between science and pseudo-science. Philosophers were interested in providing demarcation criteria that would differentiate scientific thought, such as that generally agreed to underlie modern physics and chemistry, from pseudo-scientific thought such as that underlying astrology. Popperian falsification (Popper, 1935/1959) has come to be widely accepted (particularly outside of the philosophy of science community) as providing a resolution to this debate. According to this view, for a theory to be scientific it must, at least in principle, be falsifiable. Any theory that is not falsifiable is simply not scientific. Thus, modern physics and chemistry are scientific because they involve the generation of predictions and the experimental testing of these predictions. If a theory's predictions are found by empirical work to be wanting, then the theory has been falsified. Astrology is a pseudo-science because it does not offer testable predictions and hence does not offer ways in which its "theories" may be falsified.

While falsifiability may be a reasonable aspiration for a theory, Lakatos' analysis of the history of science (Lakatos, 1970) revealed that scientific inquiry did not simply involve the proposal, falsification, and then complete abandonment of scientific theories, as a naïve acceptance of Popperian falsification would suppose.[1] Rather, Lakatos (1970) observed that there is continuity between theories within what he referred to as a "research programme", with the empirical difficulties of one version of a theory generally being dealt with through modification of aspects of that theory, rather than through complete abandonment of the theory.

In order to structure these observations, Lakatos (1970) conceptualised scientific theories as consisting of two components: central assumptions and peripheral hypotheses. Central assumptions are ontological assumptions to which the theorist is committed (e.g., all cognition is goal-directed). They give a research programme (consisting of a series of related theories) its coherence. Peripheral hypotheses, on the other hand, form a "protective belt" around the "hard core" of central assumptions. They bear the brunt of experimental testing and when empirical anomalies are found, it is the peripheral hypotheses, rather than the central assumptions, that are adjusted in order to bring the theory into line with observation. Thus theory change within the Lakatosian perspective consists of refinement of peripheral hypotheses and gradual incorporation of such hypotheses into the hard core.

Given this conceptualisation, Lakatos (1970) was able to provide revised (and credible) criteria for differentiating between science and pseudo-science. Thus, according to Lakatos' criteria, change is said to be *theoretically progressive* if "each new theory has some excess empirical content over its predecessor" (Lakatos, 1970, p. 118) and *empirically progressive* if "some of this excess empirical content is also corroborated" (Lakatos, 1970, p. 118). A research programme is said to be *progressive* if it is both theoretically and empirically progressive, and *degenerating* otherwise. Finally, a research programme is said to be *scientific* if it is at least theoretically progressive, and *pseudo-scientific* otherwise. Lakatos (1970) provided several detailed historical case studies to illustrate and justify his definitions, including Newton's theory of gravitation, Prout's theory of the atomic weight of chemical elements, and Bohr's theory of light emission and the stability of atoms.

Lakatos' distinction between scientific and pseudo-scientific research programmes is particularly important within the context of demarcation criteria. Importantly, a research programme that is theoretically but not empirically progressive is degenerating but still scientific according to Lakatos, and a degenerating scientific research programme may prove to be progressive if excess empirical content is corroborated at some later date. Thus, according to Lakatos there is no "instant rationality" by which we may judge (and condemn) a research programme. Rather, rival research programmes compete and this competition is essential to scientific progress (cf. Lakatos, 1970, p. 155).

3. Soar as a Lakatosian Research Programme

If Soar is a Lakatosian research programme as Newell (1990) suggests, then it should be possible to identify, at each stage of Soar's development, a set of hard core assumptions and a set of peripheral assumptions, with the hard core growing as the theory develops. Once such assumptions have been identified, it should also be possible to classify transitions between successive versions of the theory as scientific or pseudo-scientific, and if scientific, as progressive or degenerating. This section aims to perform these identifications and classifications.

It must be acknowledged at the outset that there is an element of judgement in the historical assessment of theoretical revisions as theoretically or empirically progressive, and therefore as scientific or pseudo-scientific. This is inevitable given that Soar has not previously been stated in

[1] Popper's view of science was of course considerably more sophisticated. Falsification was proposed in contrast to induction, which was advocated by the then dominant logical positivist view of science. Popper was also concerned with the problem-solving character of science and the growth of human knowledge through conjecture and refutation. His concern was, however, more with what could or should prompt a theory to be revised or replaced, rather than how theories change in response to apparently falsifying evidence.

terms of hard core assumptions and peripheral hypotheses. The judgements expressed here therefore attempt to favour Soar wherever possible.[2]

3.1. A Brief History of Soar

According to Laird & Rosenbloom (1996), who provide a detailed history of the development of Soar, the goal of the initial Soar theory (*circa* 1982, and henceforth referred to as Soar1) was to provide a uniform mechanism for problem solving in which multiple "weak methods" (i.e., domain-general problem solving heuristics, such as hill-climbing and means-ends analysis) could be combined in a seamless fashion and as dictated by task requirements. A key concept in this endeavour was that of the problem space, in which problem solving involves a series of steps that transform an initial state into a goal state (e.g., Newell & Simon, 1972). Each problem solving step is conceptualised as the application of an operator to a representation of the current state. Problem solving then amounts to a cyclic process of determining the set of operators available given a state, and selecting and applying one of the available operators, with problem solving terminating when the representation of the current state corresponds to a goal state.

The Soar theory has never been defined as a set of assumptions (indeed, Newell, 1992, argues strongly against such an approach), but the twin assumptions of a uniform problem solving mechanism and problem solving as traversal of a problem space, together with the representational assumption that all knowledge is represented in a uniform symbolic/propositional form, would seem to correspond to the hard core of Soar1. Indeed these assumptions are still present in the most recent version of Soar (Soar8: Wray & Laird, 1998). Soar1 also included a number of additional assumptions which have the flavour of peripheral hypotheses: that problem solving consists of a sequence of decisions, with the objects of those decisions being problem spaces, states or operators; and that each decision is the product of a processing cycle consisting of two phases. In phase one (the *elaboration* phase), representations for alternative problem spaces, states and operators are created, and "votes" are cast for the alternatives. In phase two (the *decision* phase), the votes are totted up and the winner is selected and installed as the outcome of the decision. A production system was used to implement the mechanism by which representations were created and votes cast.

Soar2 built on Soar1 by incorporating a mechanism for automatic goal creation and revising the system of voting. The mechanism for automatic goal creation, universal subgoaling (Laird, 1984; cited in Laird & Rosenbloom, 1996), was the product of a new peripheral hypothesis: that all goals are created by the architecture in response to a blockage (or impasse) in problem solving, with the object of such a goal being to resolve the impasse. The revised voting system replaced votes with *preferences*, which indicated the absolute or relative worth of the objects of decisions (so a preference might state that operator_1 *is worst*, or that operator_2 *is better than* operator_3). This change to the architecture therefore involved a change to a peripheral hypothesis of Soar1. Soar2 also arguably saw the incorporation of one peripheral hypothesis of Soar1 into the theory's hard core: in Soar1 the processing cycle, yielding a sequence of decisions and consisting of alternate elaboration and decision phases, was simply a way of implementing problem space traversal. In Soar2 it was critical in the formulation of universal subgoaling.

Soar2 was able to solve a wide range of tasks using a range of methods by applying knowledge flexibly within a uniform framework. It was not, however, able to learn from its problem solving experience. Rosenbloom (1983; cited in Laird & Rosenbloom, 1996) developed a learning mechanism—chunking—that was compatible with Soar2, and this was incorporated into the architecture in what eventually became Soar3. Chunking is intimately related to subgoaling. An impasse in problem solving triggers the creation of a subgoal. The impasse is resolved (and the subgoal completed) when further problem solving within the subgoal leads to the creation of a

[2] Similar remarks apply to the judgements later in this paper of progressions in the ACT series as theoretically and/or empirically progressive.

preference that breaks the impasse. The chunk summarizes the processing within the subgoal in the form of a new production rule whose conditions are the conditions that led to the impasse and whose actions are the creation of the preference that resolved the impasse. Chunking was effectively an additional peripheral hypothesis in Soar3. Its dependence on the universal subgoaling hypothesis and the production system substrate suggests that universal subgoaling and the production system substrate had, by this time, become hard core assumptions.

By 1986 a community of researchers interested in Soar had begun to develop, and the next version of Soar, Soar4, was concerned with consolidating Soar's computational implementation for public release rather than advancing it theoretically. Numerous minor changes were made, perhaps the most notable of which involved refinements to the variety and interpretation of available preferences, suggesting that whilst universal subgoaling was a hard core assumption, the details of its implementation where not finalized. Chunking appears to have developed by this stage from a peripheral hypothesis into a hard core assumption.

In all versions of Soar prior to 1989 application of an operator led to the creation and subsequent selection of a new state. This was inefficient since the new state was typically created by copying many features of the prior state. It was also psychologically implausible as it meant that, if a sequence of states were created within a subgoal, then each of those states would remain accessible until the subgoal was resolved (Newell, 1990). Psychological evidence (in the form of the progressive deepening problem solving heuristic: Newell & Simon, 1972) suggests that people are only able to maintain a representation of one state for any goal at a time. Soar5 addressed this issue by introducing the concept of destructive state modification and the single state principle. In Soar5 there was a single state associated with each goal, and operator application involved modifying that state, rather than creating a new state. The use of a single modifiable state for the top-level goal also meant that Soar5 could more easily and reliably interact with the external environment than previous versions of the architecture.

The single state principle qualifies as a new peripheral hypothesis in Soar5, but its implementation was far from straightforward, and many significant modifications to Soar's low-level implementation details were made in order to support the principle. Thus, the interpretation of production rules was changed to support a system that combined truth maintenance (where a production's actions could be reversed if its conditions subsequently became false) with persistent state changes (where certain state modifications were irreversible), the use of preferences was extended from the distinguished problem solving elements (problem spaces, states and operators) to all attributes of these elements in Soar's working memory, and a further type of preference (for triggering reconsideration of selected problem spaces, states and operators) was added. The status of these additional changes is unclear. Cooper *et al.* (1996) demonstrate that they are not necessary consequences of the assumptions in Soar4 combined with the single state principle. A conservative position is that the additional assumptions should be treated as further peripheral hypotheses of Soar5.

1992 saw the release of Soar6, which, like Soar4, was the product of consolidation rather than theoretical advance. Soar6 addressed issues of concern to the Soar user community (scalability and maintainability of the software), and corrected some minor conceptual problems with the interaction between Soar5's truth maintenance system and its chunking mechanism, but it contained no new theoretical content. However, improved scalability of the software opened up new areas of investigation within the Soar community, including real-time control of flight simulators (Pearson *et al.*, 1993) and interactive natural language processing (Huffman & Laird, 1993).

One conceptual difficulty with versions of Soar prior to 1994 was a mismatch between the problem space level, whereby problem solving was described in terms of transforming an initial state into a goal state by the application of operators, and the symbol level, where problem space functions were implemented by production rules. The Soar architecture did not prevent users from specifying production rules at the symbol level that violated the spirit of the problem space level. This, and efficiency concerns, led to the development of the New New Problem Space Computational Model (NNPSCM: Laird & Huffman, 1994), a theoretical proposal which, together with constraints on the form of production rules at the symbol level, ensures a tight mapping between the problem space level and the symbol level.

Soar7 adopted the NNPSCM and constraints on the form of production rules. It retained the basic decision cycle of previous versions of Soar, but all decisions were concerned with operators: problem space and state decisions were eliminated because within the NNPSCM problem spaces and states are unique to a goal. The outcome of the Soar7 decision cycle was either selection and application of an operator to the current state, or an impasse and the subsequent creation of a sub-state. Adoption of the NNPSCM reflected a significant conceptual change in Soar, but the hard core assumptions of earlier versions remained, and a key peripheral hypothesis of Soar5—the single state principle—was incorporated into the hard core.

In 1998 Soar7 was superseded by Soar8. The main rationale for Soar8 was to overcome problems in ensuring consistency of state information during subgoaling (Wray & Laird, 1998). Essentially, these problems stemmed from interactions between the implementation of the single state principle and the NNPSCM, but the revisions actually amounted to significant simplifications. Preferences for all working memory elements apart from operators, for example, were eliminated, as were preferences to force reconsideration of a previous decision. The resulting system has more in common with Soar4 (augmented with the single state principle and a hierarchical truth maintenance system) than Soar5, Soar6 or Soar7.

One last development of Soar merits special mention. Chong & Laird (1997) describe EPIC-Soar, an offshoot of the Soar project obtained by interfacing the central production system component of Soar7 with the peripheral perceptual and motor subsystems of EPIC (Kieras & Meyer, 1997; Meyer & Kieras, 1997). While EPIC is a complete architecture in its own right, its semi-modular peripheral components incorporate assumptions regarding the detailed timing of perceptual and motor subsystems, and EPIC-Soar was developed in order to allow Soar to model the executive control processes required by a complex dual-task situation (the Wickens task: Martin-Emerson & Wickens, 1992). At the time of writing, EPIC-Soar and Soar8 are separate systems, but the systems are compatible and their future merger would seem likely.

3.2. Progressive and Degenerating Steps in the Development of Soar

The preceding analysis of the development of Soar in terms of hard core assumptions and peripheral hypotheses is plausible but somewhat artificial. Soar has not previously been stated in terms of a set of assumptions independent of an implementation. There is good reason for this. The above assumptions are abstract and even vague. They would not support the kind of detailed computational work that has been carried out using Soar over the last two decades. However, and as argued above, if Newell's (1990) use of Lakatos as a defence against the Popperian falsificationists is to have force, then an analysis such as the above must be provided. Equally it must be possible to demonstrate that theoretical developments have been scientific rather than pseudo-scientific.

Recall the Lakatosian definition of a progressive step in scientific theory development: A step is progressive if the resulting theory makes new predictions (i.e., it is theoretically progressive), at least some of which are corroborated (i.e., it is also empirical progressive). Given appropriate task knowledge, the principal feature of Soar2, automatic subgoaling, makes predictions about the situations in which subgoals are created. It is therefore theoretically progressive (and scientific). For it to be empirically progressive it must be possible to detect when the human cognitive system creates subgoals. There is no background observational theory that allows us to do this, so the step cannot be empirically progressive.

The step from Soar2 to Soar3 is also theoretically progressive, because the inclusion of chunking as a learning mechanism makes predictions about changes in task performance with experience. Some of these predictions (e.g., the power law speed up of reaction time in the Seibel (1963) 1023-choice reaction task: see Rosenbloom & Newell, 1986) have received empirical corroboration, so this step is also empirically progressive. Furthermore, the predictions from chunking within a particular task depend in part on the decomposition of a task into subgoals, so learning data also provide indirect empirical support for the earlier Soar1 to Soar2 transition.

The next major step, the addition of destructive state modification and the single state principle to Soar4 to yield Soar5, is also theoretically progressive because it makes predictions about how states evolve during the execution of a task and in response to interaction with external environments. The minutia of these predictions cannot be tested without a theory of how the content of states might be determined.[3] However the progressive deepening problem solving strategy may be seen as corroborating evidence for the single state principle. A charitable interpretation of the transition to Soar5 is therefore that it is both theoretically and empirically progressive. This interpretation is charitable for at least two reasons. First, progressive deepening is not a true prediction of Soar5—it is an anomaly of Soar4 specifically addressed by modifications in Soar5. In discussing scientific explanation and theory growth, Lakatos argues that "a given fact is explained scientifically only if a new fact is also explained with it" (Lakatos, 1970, p. 119). Given this, the transition to Soar5 would only qualify as scientific if it made predictions beyond progressive deepening. Second, there are many differences between Soar4 and Soar5 that are independent of the assumptions and hypotheses leading to progressive deepening, such as the extensive revisions to the preference semantics and the introduction of impasses on non-context state elements. These assumptions do not appear to lead to predictions beyond those of Soar4. Cooper & Shallice (1995) provide a more detailed examination of the Soar4 to Soar5 progression, and come to a less charitable conclusion.

Subsequent major theoretical developments (Soar6 to Soar7 and Soar7 to Soar8) have modified the predictions of Soar5 about the evolution of states (and led to a simpler theory), but have not led to additional predictions. They have therefore been neither theoretically nor empirically progressive. In addition it remains the case that the predictions of Soar concerning the evolution of states cannot be tested with current means. Consequently developments since Soar5 have been neither theoretically nor empirically progressive. This situation could change, however, if some additional theory or approach were developed that would allow the content of mental states to be determined (and thereby compared with Soar predictions).

One theoretical development remains to be considered: EPIC-Soar. Here the assessment is more positive. EPIC-Soar is a theoretically progressive development of Soar7 because it makes additional predictions (e.g., in the Wickens task: see Chong & Laird, 1997). It is also empirically progressive because some of those predictions have been corroborated. The transition from Soar7 to EPIC-Soar is therefore clearly scientific according to Lakatosian criteria.

3.3. Conclusion

Newell (1990) suggested that the evolution of Soar was Lakatosian. The above analysis suggests that this is true, but the development of the theory has not been consistently progressive. Indeed, if one applies Lakatosian criteria in the most liberal of fashions, only two transitions have been genuinely progressive (i.e., both theoretically and empirically progressive). (See Figure 1, left.) One of the remaining transitions is scientific but degenerating, since predictions resulting from it could not be directly corroborated. More worryingly, however, many of the most recent progressions are pseudo-scientific in Lakatosian terms. This may in part reflect the current preoccupations of the Soar community, which, with the exception of the work of Chong & Laird (1997), are dominated by technical and engineering applications (e.g., Jones *et al.*, 1999) rather than cognitive scientific concerns relating to the human cognitive architecture. However, the lack of empirical grounding in the majority of recent theoretical progressions undermines the scientific credibility of the Soar research programme.

[3] Protocol analysis (Ericsson & Simon, 1984) provides the only current theory for determining the content of mental states during problem solving, but the theory does not yield the kind of detailed data necessary to support or falsify the predictions of Soar5.

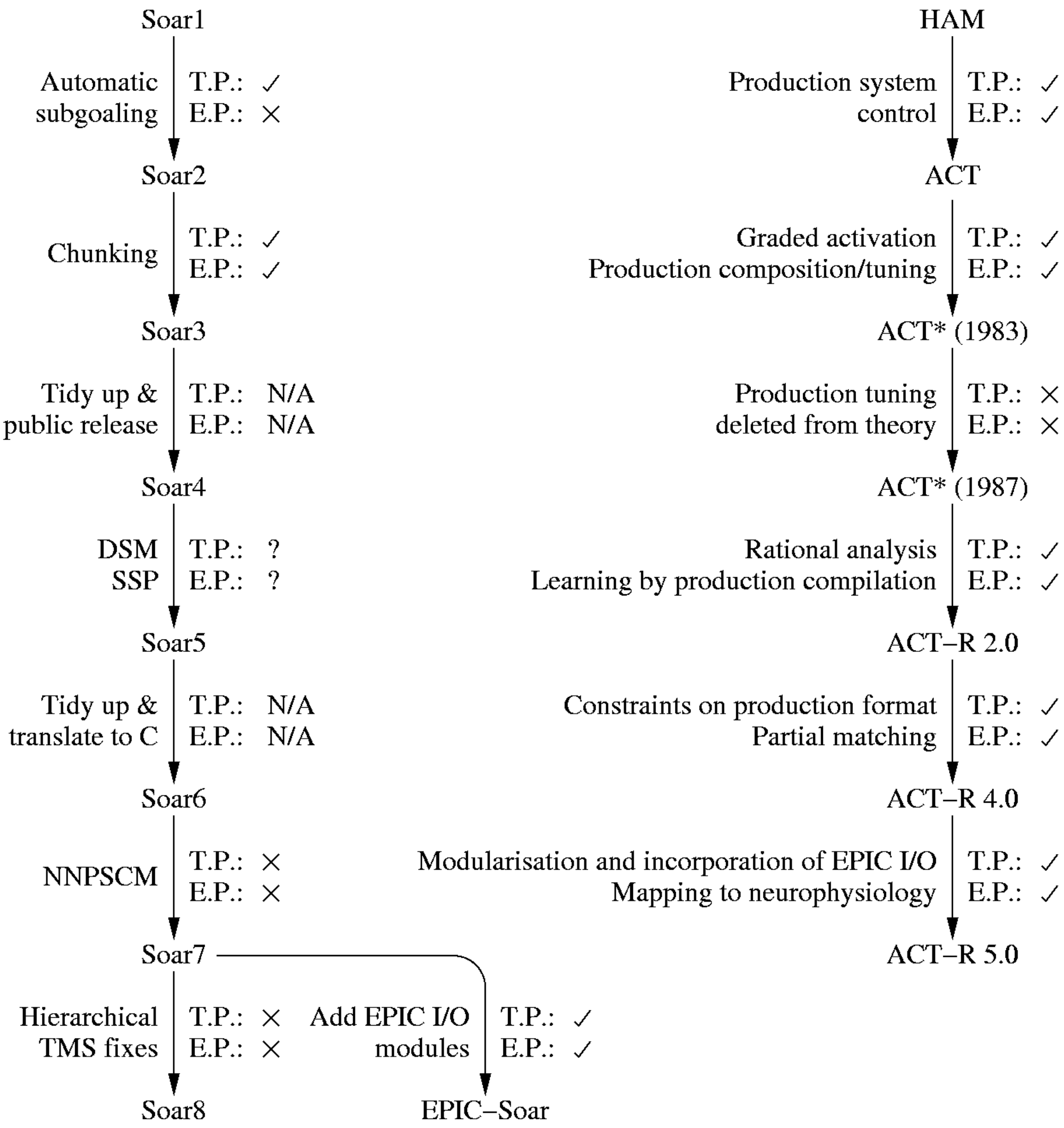

Figure 1: The development of Soar and the ACT series of architectures. Major differences between successive theories are shown to the left of each arrow. Whether the progression was theoretical or empirically progressive (T.P. and E.P. respectively) is shown to the right of each arrow.

4. The ACT Series as a Lakatosian Research Programme

While there are numerous theoretical proposals concerning the structure of the human cognitive architecture, only one, the ACT-R architecture of Anderson and colleagues (see, e.g., Anderson & Lebiere, 1998; Anderson *et al.*, 2004) has a discernable history comparable in length to that of Soar (cf. Anderson & Bower, 1973; Anderson, 1976; Anderson, 1983; Anderson, 1993; Anderson & Lebiere, 1998; Anderson *et al.*, 2004). Like Soar, ACT-R is not directly falsifiable. Behaviour on any task is argued to be the product of ACT-R operating with task knowledge, so any differences between observed behaviour and that of an ACT-R model may be due to errors in the specification of task knowledge. ACT-R is therefore subject to the falsifiability problem, and adherence to Lakatosian principles is an appropriate response to this problem.

4.1. From HAM to ACT-R: A Brief History

Work on the ACT series of architectures dates back to Anderson & Bower's HAM theory of Human Associative Memory (Anderson & Bower, 1973). HAM had numerous strengths in, for example, predicting response times in some memory experiments, but it also made some false predictions, and the theory was limited in scope to the memory domain. Anderson (1976) developed ACT—Adaptive Control of Thought—in an effort to address simultaneously some of the empirical weaknesses of HAM and phenomena from domains such as inference, language comprehension and language generation. The formulation of ACT was constrained by what Anderson (1976) referred to as "preconceived notions". These may be identified with ACT's hard core. These assumptions were: that cognition is the outcome of parallel processing; that there is a separation between declarative and procedural knowledge; that declarative knowledge is non-erasable, non-ambiguous, referential in character, and takes the form of propositions comprising predicates applied to subjects; and that procedural knowledge consists of simple modular units, the application of which is data driven and the acquisition of which is by doing (see Anderson, 1976, pp. 115–122).

Anderson (1976) was clear that these preconceived notions did not constitute a complete theory. He continued by fleshing out the notions into 11 assumptions that formed the basis of the implementation of ACT.[4] Thus, declarative knowledge was assumed to be represented in a network of linked nodes similar to the original HAM theory, while procedural knowledge was assumed to be represented by condition-action production rules with conditions that might match declarative memory and actions that specified a sequence of changes that might be made to that memory. These fleshing out assumptions are peripheral assumptions in the Lakatosian sense. Other peripheral assumptions related to concepts such as activation of declarative memory nodes (which in the 1976 ACT theory was an all-or-nothing concept), spreading of activation within declarative memory along weighted links, periodic dampening of activation within declarative memory to prevent activation growing without bound, strength of productions within procedural memory and algorithms for selecting, applying and strengthening such productions.

ACT successfully addressed some of the weaknesses of HAM concerning memory phenomena. It also demonstrated how a single set of principles could be applied to phenomena from a range of cognitive domains (memory, induction, language processing, etc.). However once again the theory had several acknowledged weaknesses. The all-or-nothing character of activation, for example, did not sit well with evidence from a range of tasks pointing to continuous levels of activation, and the theory was silent on the issue of procedural learning. Some of these weaknesses were addressed with the development of ACT*, which Anderson (1983) presented as an instantiation of the ACT framework. ACT* shared the preconceived notions of ACT (as described above), but differed in the way those notions were fleshed out. This reinforces the status of Anderson's "preconceived notions" as hard core assumptions. At the same time, most peripheral hypotheses of ACT were modified in the transition to ACT*.

Anderson (1983) stated ACT* as a set of 14 assumptions, paralleling the original 11 assumptions of ACT (Anderson, 1976). Only two key peripheral hypotheses remained intact—that a production system component operated on the system's declarative knowledge representation, and that the productions took the form of condition-action pairs with conditions matching features of declarative memory and actions creating or strengthening declarative memory elements. These are the only peripheral hypotheses of ACT that might reasonably be considered additional hard core assumptions of the ACT series *circa* 1983. Indeed, these "peripheral" hypotheses have been a central feature of all

[4] In fact, it was the implementation of ACTE, the fifth instantiation of the ACT theory, that was reported in Anderson's 1976 book.

instantiations of the ACT theory. If they were not hard core assumptions of the original 1976 theory, then they have become so since.

The key innovations in the other peripheral hypotheses of ACT* relate to the introduction of continuous activation levels for declarative memory elements, the use of distinguished goal elements and a goal stack to control or direct processing, and principles for the acquisition and modification of procedural knowledge units (i.e., productions). The first of these revises the discrete activation hypothesis of ACT (a hypothesis which can be thought of as an initial approximation), the second systematizes serial control, which was unconstrained in ACT, and the third addresses procedural learning, an acknowledged weakness of ACT. Further peripheral hypotheses concern the structure of declarative memory and the procedure for selecting productions for application.

It is worth noting that each of the three key innovations discussed in the preceding paragraph is independent in the sense that a consistent theory could have been produced from ACT by the addition of any subset of the three. Indeed, Anderson (1987) presents a slightly revised version of ACT* that includes a simplified theory of the acquisition of procedural knowledge. The 1983 ACT* theory proposed that procedural knowledge was acquired through a variety of architectural mechanisms including generalization (in which a condition of two otherwise identical productions could be removed allowing the productions to be merged) and discrimination (in which a condition might be added to a production which had proved to be overly general). Anderson (1987) rejects generalization and discrimination on the grounds that a) there is no clear empirical evidence to support automatic generalization or discrimination, and b) similar effects may be obtained using other production acquisition mechanisms within ACT* combined with suitable productions encoding reflective learning strategies.

A major step in the development of the ACT series came with Anderson's adoption of rational analysis. Anderson (1990) argued that much if not all of human cognition could be understood in terms of an optimal adaptation to the environment. The relation between rational analysis and the ACT series was initially unclear, but by 1993 Anderson had incorporated his principles of rational analysis into the ACT framework, yielding ACT-R (or more precisely, ACT-R 2.0: Anderson, 1993). ACT-R 2.0 builds on the hard core of ACT*. Thus, it retains the assumption of a procedural/declarative distinction, with the procedural component taking the form of a production system working on the declarative component. ACT-R 2.0 also retains the concepts of quantitative activation of units in declarative memory, quantitative production strength in the procedural memory, and goal-directed processing—assumptions that appear to have migrated to the hard core of the theory.

One set of key differences between ACT-R 2.0 and ACT* concerns the equations that govern declarative knowledge activation and production strength. In ACT* the equations for the former are based on spreading of activation from related knowledge units coupled with temporal decay and for the latter on previous successful applications. In ACT-R 2.0 the equations are justified by the principles of rational analysis and derived from statistical and temporal features of the use of declarative knowledge units and productions. Rational analysis is also employed in ACT-R 2.0's mechanism for production selection, which depends on the expected probability that a production will achieve the current goal, the value of the goal, and the expected cost of achieving the goal using the production. One further difference of note between ACT* and ACT-R 2.0 concerns the mechanisms available for production acquisition. ACT-R 2.0 does away with production composition. Instead, productions are acquired through a process of analogy with declarative knowledge and the subsequent compilation of this knowledge into procedural knowledge.

ACT-R 4.0 (Anderson & Lebiere, 1998) refines several aspects of ACT-R 2.0. For example, it imposes limitations on the structure of chunks, the form of productions, and the execution time of production actions. These limitations mean that ACT-R 4.0 productions are more "fine-grained" than those of earlier versions of the theory. A wide range of further innovations are introduced in ACT-R 4.0, including "partial matching" (whereby a production that almost matches working memory may be incorrectly selected, allowing for error), a modified mechanism for production selection (which in ACT-R 4.0 is based initially on matching the top-most goal on the goal stack), a prohibition on simultaneous matching of multiple production instantiations, scoping rules that affect the estimation of production utility parameters, and revisions to the mechanism of production compilation that enables declarative learning (Anderson & Lebiere, 1998, pp. 431–439). Many of the changes reflect modifications to peripheral hypotheses of the research programme. They emphasize that the details of, for example, production selection and learning are still fluid, and while rational analysis is clearly a hard core assumption of ACT-R 4.0, the rational analysis assumption combined with the hard core of ACT* does not uniquely determine numerous aspects of the theory.

ACT-R continues to evolve. The most recent version, ACT-R 5.0 (Anderson *et al.*, 2004), is more modular than previous members of the ACT family. At the same time it attempts to address an even wider variety of phenomena. ACT-R 5.0 shares the hard core assumptions of ACT-R 4.0, but includes several additional peripheral hypotheses. Specifically, ACT-R 5.0 includes modules for intention setting, declarative memory recall, visual processing and motor control. These modules interact through a series of dedicated buffers with a central production system component that is based closely on ACT-R 4.0. ACT-R 5.0 is an attempt to take seriously issues arising from interaction with the environment, and the perceptual and motor modules draw heavily on the corresponding modules in the EPIC architecture (Meyer & Kieras, 1997; Kieras & Meyer, 1997), in a manner not dissimilar to that of EPIC-Soar. One of the key new domains to which the resulting architecture has been applied is that of psychological refractory period (PRP) effects, and Byrne & Anderson (2001) have demonstrated good fits between the behaviour of ACT-R 5.0 PRP models and human behaviour in both simple reaction time PRP tasks and more complex cognitive PRP-like tasks.

Major differences between ACT-R 4.0 and the central component of ACT-R 5.0 concern the structuring of goals and the mechanism for production acquisition. Anderson appears never to have been comfortable with an architecturally specified goal stack of unlimited depth. With respect to ACT*, for example, he suggested that "it might be reasonable to assume that three to five goals are available […] The rest of the goals, however, will have to be retrieved from long-term memory if they are to be had at all." (Anderson, 1983, p. 161: see also Anderson, 1993, p. 136, for comments concerning the goal stack in ACT-R 2.0). ACT-R 5.0 assumes a goal buffer capable of holding a single goal, with super-ordinate goals reconstructed from declarative or procedural memory where necessary (following suggestions by Altmann & Trafton, 2002). Similarly, the mechanisms for production acquisition within the ACT series have been a constant source of flux. ACT-R 5.0 employs a restricted form of production composition, in which two productions that fire in sequence can under some conditions be merged into a single production (Taatgen & Anderson, 2002). The mechanism is similar to the production composition mechanism of ACT*, but is dependent on the modular structure of ACT-R 5.0 and absent from earlier versions of ACT-R. ACT-R 5.0 also attempts to address neuro-imaging data by including a mapping of architectural components to brain regions.

4.2. Progressive and Degenerating Steps in the ACT Series

As in the case of Soar, adopting the Lakatosian perspective on theoretical change allows steps in the development of the ACT series to be categorized as theoretically progressive, empirically progressive, degenerating, or pseudo-scientific. The first step, HAM to ACT, is theoretically and empirically progressive. It is theoretically progressive because HAM makes no predictions beyond the domain of memory. ACT has broader scope, so (when supplemented with suitable productions) it makes predictions above and beyond those made by HAM. It is empirically progressive because some of the additional predictions (e.g., relating to interactions between memory and induction: Anderson, 1976, pp. 355–375) have been corroborated.

The step from ACT to ACT* (*circa* 1983) is also theoretically and empirically progressive. ACT makes no predictions about priming effects on reaction time, goal directed behaviour or procedural learning. ACT* makes predictions in all three of these areas. Continuous activation levels, for example, allow accounts to be given of priming effects (Anderson, 1983, pp. 96–106) and of effects in letter and word perception (Anderson, 1983, pp. 152–155). Assumptions about goal structuring allow accounts to be given of complex goal-oriented behaviour, such as problem solving and planning (Anderson, 1983, pp. 156–169). Finally, the mechanism of production compilation allows one to provide an ACT* account of procedural learning in domains such as high-school geometry (Anderson, 1983, pp. 215–260). The predictions in each of these three areas have been corroborated, so again theoretical progress is matched by empirical progress.

The preceding positive assessment of the transition to ACT* depends upon an acceptance of the ACT* predictions as genuine predictions, and not *post-hoc* rationalizations. A more critical review could argue that pre-existing empirical anomalies drove the reshaping of ACT, and consequently that the step to ACT* was not empirically progressive. (Recall Lakatos' assertion that a given fact is explained scientifically only if another fact is explained with it.) The scientific credibility of the progression to ACT* would also have been strengthened if predictions could have been made based on interactions between the new peripheral hypotheses of ACT*. More critically, the apparent independence of the hypotheses, and the lack of any predictions resulting from their interaction, partially undermines the status of ACT* as a genuinely integrative theory of cognition.

The 1987 revision of ACT* (which involved reducing the number of ways in which procedural knowledge could be acquired) was a theoretical simplification, arguably reflecting significant insight, but it led to no new predictions. From a strict Lakatosian perspective the 1987 revision is therefore neither theoretically nor empirically progressive, and so represents a pseudo-scientific step. However, as in the case of simplifying progressions in the Soar research programme, this assessment seems overly harsh. The progression did not introduce new assumptions without motivation. Rather, it eliminated some assumptions once it was clear that learning data could be accounted for with only a subset of the original assumptions.

Anderson (1990) argues that rational analysis predicts many behavioural phenomena in domains such as memory recall, causal inference and categorization. ACT-R 2.0, with its adoption of the principles of rational analysis, inherits many of the (corroborated) predictions of rational analysis. These predictions go beyond those offered by ACT*. ACT-R 2.0 thus appears to represent a progressive development of ACT*. However, the transition from ACT* to ACT-R 2.0 must be interpreted with caution. First, Anderson (1990) suggests that many of the predictions that follow from rational analysis are independent of cognitive architecture (and hence independent of the ACT framework). This implies that some of the progressive aspects of ACT-R 2.0 could be achieved by merging rational analysis with some other architecture (e.g., Soar). Second, it is unclear if ACT-R 2.0 inherits all of the empirical content of ACT* (and hence whether ACT-R 2.0 is empirically progressive). For example, some corroborated empirical content of ACT* relates to learning, yet the learning mechanisms embodied in ACT-R 2.0 differ substantially from those in ACT*.

The transition from ACT-R 2.0 to ACT-R 4.0 is more clearly theoretically and empirically progressive. It is theoretically progressive because the introduction of constraints on declarative and procedural memory, and more critically the adoption of a default timing estimate for production actions, ensures that ACT-R 4.0 makes predictions that go beyond those of ACT-R 2.0. ACT-R 4.0 has also been applied in many more tasks than ACT-R 2.0 (ranging from arithmetic and acquiring numerical competence to dual tasking/interleaving), and in these tasks many of the ACT-R 4.0 models have been genuinely predictive. The transition is empirically progressive because many predictions of those models have been corroborated.

The most recent transition in the ACT series is embodied in ACT-R 5.0. The integration of perception and action and the mapping of ACT-R modules and processes to brain structures and regions ensure that ACT-R 5.0 makes new predictions. Some of these predictions have already been corroborated (see Byrne & Anderson, 2001; Anderson *et al.*, 2004). Changes to the mechanisms governing goals introduced in ACT-R 5.0 (i.e., the replacement of a goal stack with a goal buffer and an activation-based process of goal reconstruction) are also theoretically and empirically progressive because these changes yield predictions about goal interference, goal encoding time and goal decay (Altmann & Trafton, 2002), all of which have received empirical corroboration. From this perspective, the transition to ACT-R 5.0 appears to be both theoretically and empirically progressive. However, it is unclear if empirical coverage has been entirely cumulative. For example, are the impressive ACT-R 4.0 simulation results of the acquisition of numerical competence described by Lebiere & Anderson

(1998) retained in ACT-R 5.0, given that versions 4.0 and 5.0 of ACT-R employ substantially different approaches to goal maintenance and production learning? If not, theory development is not cumulative and the degree to which the most recent transition is progressive must be reassessed. If so, ACT-R's learning mechanism would appear to play a non-critical role in generating predictions about the time-course of learning—a counter-intuitive result that leads one to wonder how ACT-R's account of learning might be tested.

4.3. Conclusion

Given Anderson's (1976) theoretical approach, and the clarity with which ACT's original hard core assumptions were specified, it is perhaps not surprising that the development of ACT may be viewed as a Lakatosian research programme. The constant attention paid to empirical effects throughout the development of ACT has also ensured that the research programme has by and large been both theoretically and empirically progressive. (See Figure 1, right.) The picture is not entirely rosy, however, as peripheral assumptions have been subject to considerable turbulence throughout the ACT series. The role of goals and the nature of the goal stack, for example, have been revised with almost every version of the theory. Similarly production selection and more critically learning have also come in for repeated revision. This turbulence means that corroborated predictions of early version of ACT may not hold within ACT-R 5.0.

In fact, the recent changes to goal structures are particularly interesting from a Lakatosian perspective because they reflect the replacement of a simplifying assumption that allowed substantial theoretical development (the goal stack) with a more complex mechanism (activation-based goal reconstruction). Given the concern expressed by Anderson about the treatment of goals in earlier theories within the ACT research programme, recent work on goals may be viewed as an instance of elaborating a long-standing simplifying assumption. This form of theory development parallels, for example, the replacement in Newton's work on gravity of simplifying assumptions about point-mass celestial bodies orbiting around the centre of the heavier body (which allowed significant theoretical progress in some areas of the theory) with more realistic assumptions and principles concerning extended bodies orbiting around a common gravitational centre (which were only feasible once Newton had sufficiently developed the calculus). Thus, and on the positive side, there is reason to believe that recent work may have resolved long-standing questions about the nature of goals and the role of the goal stack.

This positive assessment cannot easily be applied to other sources of flux such as those relating to production selection and learning, for in these cases theory change has involved modification to mechanisms that are already quite complex (and have not previously been identified as theoretically problematic), rather than the elaboration of simplifying assumptions. In addition, the only non-ambiguously pseudo-scientific progression in the ACT research programme (that reflected in the 1987 version of ACT*), was concerned entirely with a revision to the learning mechanism. There is therefore reason to believe that future revisions of the ACT architecture are likely to witness further flux before peripheral hypotheses concerning learning and production selection are incorporated into the hard core.

5. General Discussion and Conclusion

The preceding case studies demonstrate that while neither Soar nor the ACT series have explicitly adopted Lakatosian principles, such principles can be applied to the development of both cognitive architectures. In this sense, Newell (1990) was justified in appealing to Lakatosian principles in his attempt to defend the scientific credibility of Soar: Soar may indeed be described as a Lakatosian research programme. However, the Lakatosian analysis of Soar reveals that it has at times adopted pseudo-scientific criteria for theory development. It is therefore not a scientific research programme according to the principles of Lakatos (1970). This is not necessarily a critical failing of Soar as an engineering enterprise, but given Newell's appeal to Lakatosian principles it is a critical failing of Soar as a scientific theory of the human cognitive architecture.

The Lakatosian analysis of the ACT series demonstrates that it too may appeal to Lakatos in order to claim scientific credibility. The analysis supports the contention that the ACT series is more scientific than Soar (and this contention is reflected in the number of refereed journal articles within the psychological literature reporting empirical support for each architecture: over 100 for ACT and none for Soar). It might appear then that if Soar is to regain its scientific credibility it should attempt to emulate the methodology adopted by the ACT series. However, the ACT series is not a perfect role model: some areas of the theory appear to be in constant flux (particularly learning, a critical area of the theory). More critically, it remains to be demonstrated that Lakatosian principles are appropriate for the development of cognitive architectures. As noted above, Newell's original motivation for invoking Lakatos was to defend Soar against claims that it was unfalsifiable. Newell's purpose was to argue that Soar was a scientific theory, rather than merely a framework or pseudo-scientific belief system. Lakatosian principles are not necessarily required to make this argument—other credible demarcation criteria would serve just as well.

An alternative position is to deny that demarcation is an issue. Laudan (1983) does just this, arguing that science is "not all cut from the same epistemic cloth" (Laudan, 1983, p. 124), and that the demarcation problem is spurious. Laudan's view is particularly relevant in the current context as his notion of a research tradition provides what appears to be a viable alternative to Lakatos' conception of a research programme. Research traditions (Laudan, 1977) lack hard core assumptions, but are defined by metaphysical and methodological commitments (i.e., by the entities of their theories and the methods of inquiry employed within the tradition). By doing away with hard core assumptions, Laudan acknowledges that science (or at least research and theory development) is often not strictly cumulative as Lakatos would have us believe. In addition, Laudan notes that "the succession of specific theories … involves the elimination as well as the addition of assumptions" (Laudan, 1977, p. 77). Thus, the elimination of assumptions seen in research on both Soar and the ACT series is, in Laudan's view, common within research traditions, and hence not something that should count against the status (scientific or otherwise) of Soar or the ACT series.

Abandoning demarcation, however, is a significant step that comes at a cost. For example, within a Lakatosian framework pseudo-scientific progressions are to be avoided because they undermine the scientific status of a theory. The above analysis demonstrates in addition that, at least in the cases of Soar and the ACT series, they also correspond to theoretical progressions that, with hindsight, have not stood the test of time. What, then, is required of a cognitive architecture if we accept demarcation and adopt an explicitly Lakatosian approach to theory development? Essentially, each revision of the theory must address five questions:

1. What are the hard core assumptions and peripheral hypotheses of this version of the theory, and how do these relate to those of the previous version of the theory?

2. What empirical anomalies arising from the previous version of the theory are addressed by the revised theory (i.e., what existing empirical findings does this version of the theory account for that the previous version did not)?

3. Are all empirical phenomena that were addressed by the previous version of the theory also addressed by the new version?

4. What empirical phenomena does this version of the theory predict that the previous version did not?

5. What empirical anomalies remain (i.e.. what phenomena remain to be accounted for)?

The answer to the last of these questions cannot necessarily be stated at the time the revised theory is proposed, but the principles require that empirical anomalies are catalogued as the theory is explored, for it is these that provide the impetus for theory development.

Our analysis has been based on the development of just two cognitive architectures: Soar and ACT-R. While it would be instructive to analyse the development of other architectures (e.g., EPIC, 4-CAPS, etc.), none have a published history comparable to Soar or the ACT series. The lessons learned from this analysis nevertheless may be applied to other cognitive architectures. Thus, while the above analysis has concentrated on the issue of science versus pseudo-science through the *descriptive* application of Lakatosian principles, the same principles may be applied *prescriptively* to help guide theoretical progress and ensure positive theoretical growth. We therefore suggest that addressing the above five questions with each revision of a cognitive architecture will provide the architecture with more consistent, positive, theoretical growth than is currently achieved through the existing practice of *ad hoc* theory development.

Acknowledgements

This paper grew out of a presentation made in a symposium titled "What would convince me that my theory was wrong?" at the 5th International Conference on Cognitive Modelling. While little of that presentation remains here, I am grateful to John Anderson for organising the symposium and inviting me to speak, and to him, the other presenters (Werner Tack and Michael Freed) and the audience for raising issues that shaped this paper. I am also grateful to Robert Wray for helpful discussions on the recent history of Soar, and to William Bechtel and two anonymous reviewers for their comments on an earlier draft of this paper.

References

Altmann, E.M. & Trafton, J.G. (2002). Memory for goals: An activation based model. *Cognitive Science*, *26*, 39–83.

Anderson, J.R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J.R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, *94*, 192–210.

Anderson, J.R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J.R. (1993). *Rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J.R., Bothwell, D., Byrne, M.D., Douglass, S., Lebiere, C. & Qin, Y. (2004). An integrated theory of mind. *Psychological Review*, *11*, 1036–1060.

Anderson, J.R. & Bower, G. (1973). *Human associative memory*. Washington, DC: Winston & Sons.

Anderson, J.R. & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, M.D. & Anderson, J.R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review*, *108*, 847–869.

Chong, R.S. & Laird, J.E. (1997). Identifying dual-task executive process knowledge using EPIC-Soar. In M.G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual conference of the cognitive science society* (pp.107–112). Mahwah, NJ: Lawrence Erlbaum Associates.

Cooper, R. P., Fox, J. Farringdon, J. & Shallice, T. (1996). A systematic methodology for cognitive modelling. *Artificial Intelligence*, *85*, 3–44.

Cooper, R.P. & Shallice, T. (1995). Soar and the case for Unified Theories of Cognition. *Cognition*, *55*, 115–149.

Huffman, S.B. & Laird, J.E. (1993). Learning procedures from interactive natural language instruction. In P. Utgoff (Ed.), *Machine learning: Proceedings of the twelfth national conference on Artificial Intelligence* (pp. 506–512). Amhurst, MA: American Association for Artificial Intelligence.

Hunt, E. & Luce, R.D. (1992). Soar as a world view, not a theory. *Behavioural and Brain Sciences*, *15*, 447–448.

Jones, R.M., Laird, J.E., Nielsen P.E., Coulter, K., Kenny, P., & Koss, F. (1999). Automated intelligent pilots for combat flight simulation, *AI Magazine*, *20*, 27–42.

Just, M.A., Carpenter, P.A., & Varma, S. (1999). Computational modelling of high-level cognition and brain function. *Human Brain Mapping, 8*, 128–136.

Kieras, D.E. & Meyer, D.E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of the psychological refractory-period phenomenon. *Psychological Review, 104*, 749–791.

Laird, J.E. (1984). *Universal Subgoaling*. Unpublished doctoral dissertation, Carnegie-Mellon University, Pittsburgh, PA.

Laird, J.E. & Huffman, S.B. (1994). *NNPSCM.* Note to Soar-group email list. Available at http://www.nottingham.ac.uk/pub/soar/papers/nnpscm17.2.94.txt. February 17, 1994.

Laird, J.E., Newell, A. & Rosenbloom, P.S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence*, *33*, 1–64.

Laird, J.E. & Rosenbloom, P.S. (1996). The evolution of the Soar cognitive architecture. In D.M. Steier & T.M. Mitchell (Eds.), *Mind matters: A tribute to Allen Newell* (pp. 1–50). Mahwah, NJ: Lawrence Erlbaum Associates.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge, UK: Cambridge University Press.

Laudan, L. (1977). *Progress and its problems*. Berkeley, CA: University of California Press.

Laudan, L. (1983). The demise of the demarcation problem. In R.S. Cohen & L. Laudan (Eds.), *Physics, philosophy and psychoanalysis* (pp. 111–127). Dordrecht, The Netherlands: D. Reidel Publishing Company.

Lebiere, C. & Anderson, J.R. (1998). Cognitive arithmetic. In J.R. Anderson & C. Lebiere (Eds.), *The atomic components of thought* (pp. 297–342). Mahwah, NJ: Lawrence Erlbaum Associates.

Martin-Emerson, R. & Wickens, C.D. (1992). The vertical visual field and implications for the head-up display. In *Proceedings of the 36th annual symposium of the human factors society* (pp. 1408–1413). Santa Monica, CA: Human Factors Society.

Meyer, D.E. & Kieras, D.E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104*, 3–65.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newell, A. (1992). Soar as a unified theory of cognition: Issues and explanations. *Behavioural and Brain Science*s, *15*, 464–492.

Newell, A. & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Newman, S.D., Carpenter, P.A., Varma, S., & Just, M.A. (2003). Frontal and parietal participation in problem solving in the Tower of London: fMRI and computational modelling of planning and high-level perception. *Neuropsychologia*, *41*, 1668–1682.

Pearson, D.J., Huffman, S.B., Willis, M.B., Laird, J.E. & Jones, R.M. (1993). Intelligent multilevel control in a highly reactive domain. In F. Groen, S. Hirose & C. Thorpe (Eds.), *Proceedings of the 3rd international conference on intelligent autonomous systems* (pp. 449–458). IOS: Washington, DC.

Popper, K. (1935/1959). *The logic of scientific discovery*. New York, NY: Basic Books.

Rosenbloom, P.S. (1983). *The Chunking of Goal Hierarchies: A Model of Practice and Stimulus-Response Compatibility*. Unpublished doctoral dissertation, Carnegie-Mellon University, Pittsburgh, PA.

Rosenbloom, P.S. & Newell, A. (1986). The chunking of goal hierarchies: A generalized model of practice. In R.S. Michalski. J. Carbonell, & T. Mitchell (Eds.), *Machine learning: An artificial intelligence approach, II* (pp. 247–288). Los Altos, CA: Morgan Kaufman.

Seibel, R. (1963). Discrimination reaction time for a 1023-alternative task. *Journal of Experimental Psychology*, *66*, 215–226.

Taatgen, N.A. & Anderson, J.R. (2002). Why do children learn to say “Broke”? A model of learning the past tense without feedback. *Cognition*, *86*, 123–155.

Vere, S.A. (1992). A cognitive process shell. *Behavioural and Brain Sciences*, *15*, 460–461.

Wray, R.E. & Laird, J.E. (1998). Maintaining consistency in hierarchical reasoning. In J. Mostow, C. Rich, & B. Buchanan (Eds.), *Proceedings of the fifteenth national conference on Artificial Intelligence* (pp. 928–935). Menlo Park, CA: American Association for Artificial Intelligence.