DRAFT MATERIAL: LIMITED DISTRIBUTION FOR COMMENT.
NOT FOR QUOTATION
File: c8af.mss
Current version: 31 August 89 14:15
Started Draft 3: 29 Jun 89

The 1987 William James Lectures
UNIFIED THEORIES OF COGNITION

CHAPTER 8
ALONG THE FRONTIERS

DRAFT 3.1

Allen Newell

31 August 1989

Departments of Computer Science and Psychology
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

Table of Contents

List of Figures

8. Along the Frontiers

We arrive at the last chapter. Where do matters stand on making the case for unified theories of cognition and what they promise for cognitive science? We have not argued the case abstractly, but have chosen to exhibit a candidate unified theory. This seemed the only way to communicate what such a thing would be and how having one might make a difference. Soar has been our exemplar of a unified theory of cognition throughout, to show what it might be like to have one.

It needs repeating once more. In this book I am not proposing Soar as *the* unified theory of cognition. Soar is, of course, an interesting candidate. With a number of colleagues, I am intent on pushing Soar as hard as I can to make it into a viable unified theory. But my concern here is that cognitive scientists consider working with *some* unified theory of cognition. Work with Act*, with Caps, with Soar, with a connectionist unified theory of cognition (CUTC).[214] Just work with some UTC. That is the message of this book and the end toward which all its chapters are aimed.

To continue restating basics, a unified theory of cognition is by definition a theory of mind. However, a theory of mind is not by definition a theory of the computational and informational processes that comprise the mind. Indeed, there is a significant amount of philosophic discourse that either worries about this or challenges it, although the field is not overloaded with specific well-developed alternatives (see Costall & Still, 1987). This general presumption is the basic premise of cognitive science and it is adopted without question by current candidates for unified theories of cognition. Soar shares this premise fully.

There is no alternative for a unified theory of cognition based on information processing, but to be a theory of the architecture. The architecture is where the strong organismic invariants are. However, it needs to be emphasized (one last time) that the architecture by itself does not determine behavior. Indeed, an architecture is precisely a device for making it possible for something else to determine behavior, to wit the knowledge encoded in its memories. Such knowledge is about many different things — the system's goals, the system's bodily state, the local task environment, the social environment, the general nature of the physical world, etc. This is the entire burden of the knowledge level and the symbol level that supports it. However, it is the shape of the architecture that gives shape to the psychological regularities at the cognitive level. In agreement with all this, Soar is an architecture and was presented as one in Chapter 4.

The hallmark of a unified theory is the range of central cognition and its surround that it addresses. The numbers of regularities addressed is only a surrogate for the coverage provided. Thus, central to the case to be made for unified theories is that a single theory can have a broad range. Nothing is more important to illustrate. Thus, once the architecture was out on the table, the remainder of the book, from the last of Chapter 4 through Chapter 7, was devoted to making plausible that a wide range was possible. Throughout this last part of the book, we have been accumulating a box score of the domains covered for our exemplar, Soar. Figure 7-21 at the end of the previous chapter provides the latest total. This may seem a little dramatic — or perhaps a little corny. Box scores don't belong to science, but to sports or politics. But some way was

[214] I was going to propose connectionist unified theories of everything, but that seemed a little too cute.

needed to emphasize how important coverage is. Once we have one or more unified theories — then what will count is whether a theory covers the particular collection of phenomena that some user of the theory needs. But until then, it is useful to be crass enough to emphasize sheer numbers — sheer coverage.

For Soar, what heads the list in Figure 7-21 is the prediction that humans can be intelligent. Soar's demonstration of this is as good as the state of the art in AI. Functionality is a cornerstone of a theory of cognition — to explain how it is possible for humans to be intelligent. Next is the demonstration that Soar exhibits the qualitative shape of human cognition, though the particular list of such global properties of cognitive behavior was ad hoc. Especially for a unified theory, it is important to attend to the in-the-large picture of the human that it presents, not just to how well it predicts detailed experimental results. For it should be possible to think generally and globally about of the nature of human nature by means of the theory. That should certainly be one of its important uses.

Next, we showed that Soar was a theory of immediate responses, those that take only ~1 s. Since Soar's original development has been at the problem-solving level, this was an important demonstration that Soar also applies at the micro level — that the details of its architecture are to be taken seriously. We extended Soar to simple discrete motor-perceptual skills, namely typing. This reinforced the work on immediate-response behavior. It was as close as we could get to motor-perceptual skills, since the total cognitive system (P-E-C-D-M) is still underdeveloped.

Soar provides a theory of acquisition of skills through practice. We showed that Soar still exhibited the power law of practice. Through a rather circuitous route, we describe the Soar theory of recognition and recall of verbal material. This aspect of Soar, arising from the solution to the data chunking problem, provided a demonstration of how an architecture can be a source of unanticipated consequences. We touched on the Soar theory of short-term memory that was implicit in Soar, although it has not been well developed. We did show that it provided a theory of where the plethora of short-term memories come from. Finally, we showed how Soar provided a detailed theory of problem solving, a theory of logical reasoning, and a theory of how instructions are converted into the self-organization for doing new immediate response tasks. These tasks were another temporal level or two above immediate responses, and demonstrated the temporal reach of the theory.

These are the range of things which this exemplar unified theory has addressed. It has done this with varying degrees of success, varying degrees of depth and varying degrees of coverage. Mostly these limitations reflect where we have had time and energy to push Soar development. Still, it is one system — one architecture — that does all of these tasks and does them in fair accord with human behavior on the same tasks.

It is important that Soar be seen to be a theory and not a framework of some kind. For instance, it is not a loose verbal framework, which guides the form of invented explanations, but does not itself provide them. Nor is it the sort of computational harness systems that has begun to appear in Human Factors, in which specific psychological theories can be plugged in as modules, with the framework providing some sort of integration. On the contrary, Soar is a specific theory by means of which one calculates, simulates and reasons from its fixed structure. When it seems not to be that way — when it seems to be too underdetermined — the difficulty is almost invariably that the other sources that determine behavior are not sufficiently known (goals, knowledge, the structure of the task environment). That is a genuine problem for all the human sciences, even more so than for other sciences, just because humans have so much knowledge, i.e., bring so much of the rest of the world and its history to any current situation.

Soar is hardly perfect. No scientific theory is, of course. But with Soar just an exemplar, there is no need to conceal its vulnerability throughout. Moreover, we have spent no effort probing the limits of Soar's explanatory adequacy. Instead, we have been driven by the need to show that Soar has the earmarks of a unified theory of cognition.[215] Certainly, the positive description of Soar does not gainsay what still needs to be done to analyze complete sets of regularities in detail and to find where Soar breaks down.

In particular, remembering the ~~3000, there is long way to go in obtaining coverage of the aspects of human cognition about which much is already known. To help make this last point, Figure 8-1 lists a potpourri of areas that have not been done by Soar. It would be easy to generate twice as many. We have put them in alphabetical order — what other order is there for a potpourri?

Some of these items may seem peripheral. But some are exceedingly important in probing whether Soar has the right fundamental character. Consider a sampling (the starred items), to appreciate the variety of what is missing. Start with *consciousness*. Soar provides a theory of *awareness*. Soar is aware of something if its deliberate behavior can be made to depend on it. In this general sense, awareness is a operationally defined and fundamental, and is a much used notion throughout cognitive psychology. But consciousness can be taken to imply more than awareness in this sense, namely, the phenomenally subjective. It can mean the process, mechanism, state (or whatever) that establishes when and what a human would claim to be conscious of, both concurrently or retrospectively.[217] Soar does not touch the phenomena of consciousness, thus delineated. Neither does much else in cognitive psychology. This can dictate pushing this issue into the future, until additional regularities and phenomena accumulate, but the challenge remains.

Contingencies of reinforcement can be made to stand for a large and well-established body of regularities that link human behavior and animal learning behavior. First, speaking specifically, Soar must explain why (and to what extent) humans obey the various laws established in operant conditioning about how the frequency of specific responses can be brought under the control of specific environmental events. That these regularities have been developed in an area outside cognitive psychology (and which mostly predate cognitive psychology) is not relevant to their status as regularities and with the need for Soar to cover them. Speaking more broadly about the psychology of animal behavior, unification is certainly required. It is already underway from the animal-behavior side (Rescorla88-8), but the integration needs to be pursued from the human-cognitive side as well.

The area of *concepts* and *concept learning* was originally defined in terms of category, defined

---

[215] EdNote: These paragraphs clearly hark back to Figure 1-6; should we be more explicit about re-evoking it?

[217] This claim, of course, is filtered through culturally established communication, with the meaning *conscious* established by the same social process as for any other term. This certainly complicates matters, but it doesn't get rid of the phenomena of consciousness.

Arithmetic (simple and multidigit)
Cognitive styles
* Consciousness (as opposed to awareness)
* Contingencies of reinforcement (animal learning)
* Concept learning - prototypes
Counting and comparing (also, subitizing and estimating)
Daydreaming
* Decisions under risk and uncertainty
Discourse comprehension
* Dreaming
Dual tasks
Emotion and affect
* Imagery
Individual differences
Lexical access
Memory for frequency of occurance
Memory for items versus memory for order
Metaphor and analogy
Motivation - Re oomph, not directionality
Phonemic restoration effect
Play
* Priming
Probability matching
Same-difference judgements
Story understanding
Symbolic comparisons
Values and morals

Figure 8-1: A potpourri of what has not been done.[216]

by means of a predicate (Bruner, Goodnow & Austin, 1956). A person has a concept if he or she could determine whether the predicate was true of a presented object or situation. The area was enriched in the 1970s by being extended to prototypes, where membership was defined by the resemblance to a central category (Rosch & Lloyd, 1978). Hence, an object was a better or worse exemplar of a concept. The extension to prototypes has been taken to imply a fundamental revision of our views of the nature of mind (Lakoff, 1987). At issue is whether the mental world of humans is constructed propositionally, hence with all concepts categorical, or whether in some other way — the same underlying concern that motivates the study of syllogisms. Soar must give an account of prototypes and show how they arise when the situations call for them, just as predicates arise when the situations call for them. Here is a place where it would occasion surprise (at least on my part) if additional architectural assumptions were needed to get these different types of concepts, and others as well.

Decisions under risk and uncertainty (Pitz & Sachs, 1984) is the domain of behavioral decision theory, which is based on the concept of subjective expected utility (SEU). Namely,

[216]FigNote: Changes: See verbatim, judgement = judgment.

game theory and its psychological derivative, decision theory, basically provide[249] — this can become common ground for all such efforts.

8.5. The Role of Applications

Applications form an important part of the frontier of any theory. Of course, they are not quite the same as the scientific frontier, at least not in the usual way of describing the scientific enterprise. But they are equally diverse in character. Further, the commerce across an application frontier goes in both directions, not just from the science to the application.

The first important point about applications was already touched on in Chapter 1. A unified theory of cognition is the key to successful applied cognitive science. Real tasks engage many aspects of cognition, not just a single aspect. Consequently, it is necessary for an applied cognitive psychology to be able to treat all the aspects involved — not just memory, not just decision making, not just perception, but all of them, in an integrated enough way to produce the answers in the applied domain. This proposition says that unified theories of cognition are exactly what are needed to make progress in building an applied cognitive psychology.

To illustrate this, consider human computer interaction. Recall from the Chapter 1 (Figures 1-9 and 1-10) that one of my personal precursors of the current attempt at unified theories of cognition was the work that Stu Card, Tom Moran and I did on what we called the Model Human Processor (MHP), attempting to get a first cut at a broad-spectrum theory of the user (Card, Moran & Newell, 1983). We can take Soar as the foundation for Model Human Processor II (MHP-II). During the course of this book, we've gotten Soar to cover a number of activities that are integral to interaction across the computer interface. Figure 8-13 lists them: Stimulus response capability, typing and keying behavior, immediate instruction, cognitive skill acquisition, improvement with practice, and reading and comprehension. We did not explicitly cover the routine cognitive skills involved in using command language systems such as text editors, which occupied a central role in the earlier work as the GOMS model and the Keystroke-level model. But that sort of routine cognitive skill, which consists of a succession of short (~~10 s) unit tasks, with clear task phases (get the next task, locate the place to modify, make the modification, verify if necessary) fits easily within the multiple problem spaces of Soar, with operators corresponding to the larger tasks (*edit-manuscript*), requiring implementation with subspaces of more specialized operators (*get-next-edit*, *make-edit*, *verify* and *get-next-page*), and so on. Indeed, the stimulus-response compatibility methods of Chapter 5 were the extension of these techniques down to a finer level of detail.

Some other areas are also important in HCI, which Soar has not yet covered. One is visual search and inspection. Lots of good regularities exist here, with not only very good data but also good quantitative models (Teichner & Krebs, 1974, Teichner & Mocharnuk, 1979). No attempt has yet been made to extend Soar to this domain, in part because it depends critically on the perceptual structure for Soar, which is not yet in sufficient shape for such an exploration. However, many of the existing visual search models do not require any more of perception than transcription typing requires of the motor system. This is an excellent area in which to extend

---

[249]EdNote: Revise! — The work in limits to rationality from behavioral decision theory (Dawes, Kahneman, Lichtenstein, Nisbitt, Slovic, Tversky) all imply a more complex picture.

1. Stimulus-response compatibility (immediate response)
2. Typing and keying behavior
3. Performance of unit-tasks (~~ 10 s)
4. Immediate instruction
5. Acquisition of cognitive skill
6. Reading and comprehension

....

7. Visual search and inspection - Regularities exist
8. Use of external temporary memory - Not studied

Figure 8-13: UTC's for applications.[250]

Soar with every expectation of success.

Another important area is the use of external temporary (or scratch) memory in rapid feedback situations. Here there is not much data in the literature, and there may be interesting surprises. There are actually two areas here. One is the aspect coming to be termed *situatedness*, in which the external environment is as much a source of knowledge for what the human does as long-term memory (SituatednessXX-8). The second is the external short-term temporary memory, in which the human builds a rapid (~~ 1s) feedback loop between himself and an easily modified and accessed medium. The former is beginning to receive a lot of attention; but this latter still has not evoked the experimental work necessary to lay out the basic regularities.

The collection listed in the figure is almost a complete set of activities that occur at the display interface in modern interactive computing. Soar does not quite cover the set, and even in those activities where something has been done with Soar, there is still much to do. Yet, it is a reasonable research goal to set for Soar that it cover the entire collection. Such coverage would indeed create a new version of a Model Human Processor that would be able to deal in an integrated way with the tasks at the interface.[251]

The most important thing to say about applications is in the other direction — not about what UTC's can do for application, but what applications can do for unified cognitive theories. Applications provide important ingredients for the overall basic scientific enterprise. Of course, they provide the successes that convince the polity that the science is worth supporting. That goes without saying in the current era. My point is that applications provide critical *internal* ingredients. They establish what is worth predicting. They establish what is sufficient accuracy. They establish when a regularity is worth remembering. They establish when a theory should not be discarded.

All that seems perhaps like quite a bit. But if a theory is doing good things in terms of applications, then, even if it seems uninteresting for other reasons, it will be kept and used. Applications say it is worth remembering Fitts Law, but applications are silent on the phenomena

---

[250]FigNote: #3 becomes just "Performance of unit tasks."

[251]EdNote: Give an example to make concrete and convincing.

of the Release from PI. The latter is certainly a regularity.[252] It was quite interesting during the early 1970s, in connection with interference theories of short-term memory. Fitts Law is tied in deeply with many applications. Release from PI is not tied in with any applications, as far as I know. Without applications as a guide, there is only current theory. And current theory exists in the minds of scientists not in the world of real effects. Thus, as the theory goes, so goes interest in its facts. There is a theme in current philosophy of science that emphasizes that facts are theory bound — which even goes so far as to deny the status of fact to a regularity if it is not part of a theoretical view. And so we tend to forget about release from PI when interference theory goes out of fashion. And we tend to ignore Fitts Law when motor behavior is not fashionable in current cognitive theory. But some of us don't forget Fitts Law, because applications keep us interested in it.[253]

This is not an argument against basic science. Just in case it is not obvious from my discussion in Chapter 5, I hereby go on record as being in favor of the ~~3000 regularities. Many, indeed most, of these are arcane and not related to application. They gain their significance only when theory finds them important. We can't just attend to all ~~3000 regularities on the hope of relevance for the future. And applications have a wisdom that the current fashions of theory do not, and indeed probably should not.

An area of application provides an external reason for working with certain kinds of theories and therefore for grounding the theories. It creates sustained communities that nourish a theory for a right reason — because it serves them. Interestingly, pure concern for science doesn't have that character. The scientist is free to forget some phenomena and just to go on to other phenomena. As individual scientists, we properly exercise that freedom all the time. But applications provide grounding because the grounding is there in the larger motivation for the application.

Let me draw the lesson for Soar, just to make the point concrete. I expect Soar to work for HCI and to be used, so to speak, every day in every way — at least eventually and I will work to make that happen (Newell & Card, 1985). Grant that for the moment. Now suppose, just for example, that some data from neuroscience on localization of function, turns out to be radically different from the apparent way Soar is. Thus, some strong disconfirming evidence shows up. Soar will survive that just fine, as long as there is a domain for which Soar is being used, for which the neuroscience data isn't particularly relevant. For Soar that area in HCI applications. It will hold work on Soar in place, until someone finally comes up with a theory which takes into account both the functional neuroanatomy data and the higher-level phenomena about how

---

[252]Here is the basic phenomenon, for those not acquainted with it. A basic paradigm to demonstrate that short-term memory for verbal material is only ~10 s, gives a verbal item (say three letters), then fills a fixed interval (say 15 s) with distracting activity (counting backwards by 7s) to prevent rehearsal, then asks for recall of the item. There is almost perfect recall the first time this is performed in an experimental sequence of trials, but then recall falls off with additional trials. It is said there has been a *build up of proactive inhibition*. If a somewhat different item is given after awhile, say three digits, then this is remembered easily, although if triplets of digits are continued, they too become hard to remember after a few trials. It is said there has been a *release for proactive inhibition (PI)*. The standard explanations ascribe the phenomena to having to discriminate the new items from the old ones, so that when a different item is given it is easy to discriminate.

[253]EdNote: Really cryptic.

humans interact with computers.[254] Without that — having found a domain which clearly shows that Soar is wrong — the general consensus can lead cognitive scientists simply to walk away Soar. In short, I believe that having domains of applications are extremely important to having good theories. Unified theories of cognition are no exception.

8.6. Final Remarks

We have finished our tour of the frontier, although we have hardly traveled it all. The remaining question is not one of science, but one of strategy. How are we to evolve unified theories of cognition? How are we to get from *here*, where I have claimed they are in prospect, to *there*, where we will all be working within their firm and friendly confines? What is needed are recommendations, and I record mine in Figure 8-14. There are seven of them.

1. Have many UTCs
   - Unification cannot be forced
   - They can even be quite similar - Act* & Soar
   - Exploring the architecture requires owning it
2. Develop consortia for working with a UTC
   - UTCs require many person-years
3. Be synthetic
   - Incorporate (not replace) local theories
4. Modify UTCs - even in apparantly radical ways
   - Coupling between parts is loose (but nonzero)
5. Create data bases of results and adopt a benchmark philosophy
   - Each new version must run on the benchmarks
6. Make models easy to use, easy to make inferences from
7. Acquire one or more application domains to provide support

Figure 8-14: Recommendations for how to evolve to UTCs.

(1) There should be many unified theories of cognition — at least for awhile. I've reiterated this theme many times throughout the book. It is worth doing it once more. You cannot force unification. We are a field not-yet-unified and we do not know what it feels like. Everyone is king, or a least can think it so, as long as no one is. Any attempt to settle on a unified theory by any means other than open competition in the marketplace of ideas — to use the cliche — will only end in squabbling. View this book simply as a device for my making the case for Soar, then cynical dismissal follows. I get my day in the sun, courtesy of the Harvard Psychology Department; other days belong to others, from other petty principalities. I am, of course, intrigued by Soar; I am even in favor of it. After all, if not I (and my co-creators), who else should be? But that is not the point.

[254]Of course, if Soar also turns out to be no good for the applications, or less good then new theories that come along, then it simply goes away, as it should.

Thus, there must be many unified theories. They can even be quite similar, like Act* and Soar, which have an immense communality in their base. The case of Act* and Soar makes clear one important reason why there must be many unified theories. Anyone who wants to work with architectures, must control one. He or she must be free to reformulate it. Ultimately, when a single unified theory of cognition takes its place, and its acquisition is behind us, if I may speak that way, then the architectural theorists will work differentially off of the base theory. And the rest of us can go about our business of using and developing the base theory — of working on puzzles within the paradigm, to use the Kuhnian way of speaking. But at the moment there is no accepted base theory with anything like the required coherence and coverage.

(2) We ought to develop consortia — substantial collections of cognitive scientists who are committed to putting together a particular unified theory. Unified theories of cognition are not the product of a single person. They take a lot of work. Remember all those ~~3000 regularities! There is no critical subset of 7 regularities, so that the rest can be ignored. Architectures are complex and somewhat loosely coupled, and so it takes a lot of regularities to hammer one into shape. It may seem an odd way to express it, but there can be no Newton for cognitive science, although perhaps a Newton Inc.

I estimate there are already at least fifty or sixty person-years in Soar circa mid 1988, and I suspect it will not take too long to raise the investment into the hundreds. And Soar is at the very beginnings of its development as a UTC. You can't put such a unified theory together without a lot of people engaging in a lot of work. Isolated, individual investigators simply cannot do the amount and kind of research that is required. I trust it is evident that consortia are cooperative efforts among free-standing scientists, so coordination is implied, not subordination. Still, it will be a shift of style for many cognitive psychologists. Ed Note: How to get into Soar!

(3) Be synthetic. Incorporate existing local theories. Indeed, incorporate other groups' unified theories! This is another theme I've been saying throughout the book. It is not necessarily a good thing for Soar to propose theories that are radically different than those already floating in the literature. Better to see Soar as an attempt to express these other theories in a way that will allow them to combine with other parts of cognition. What holds for Soar holds for other unified theories as well. The aim is to bring theories together. So operate cooperatively, not competitively.

(4) Unified theories of cognition can be modified, even strongly and in radical ways. They are not monolithic, though the whole does bring constraints to bear that make modification a stirring, even exciting proposition, and one for the young and energetic. Let me illustrate the notion of radical change by a conversation I had with the Thinking Machine folk (who make the Connection Machine). We were talking about whether it was possible to abandon the production system entirely and replace it with something like the memory-based reasoning work that they are currently doing on the Connection Machine, and which seems to fit its structure very well (Stanfill & Waltz, 1986). We could identify clearly that the role of the production system in Soar was as an associative recognition memory with a certain modularity for chunking. Maybe, if we just preserved those two properties, we could wheel out the productions and wheel in a memory-based reasoning system. It appears by now (circa 1988) that this path may not be so easy. Still the approach seems absolutely right.[255]

[255] EdNote: Maybe Soar5 instead?

(5) Create data bases of results. Develop a benchmark mentality, in which one doesn't walk away from old data, but preserves it in operational form. Then expect any new theory to run on the old benchmarks — that is, to provide postdictions and explanations of them. Simply disbelieve, or at least disown, the Kuhnian rhetoric that new paradigms speak past each other, so that progress does not happen at the paradigm level, only the choice of what swamp to wander in. Be an engineer in this regard, rather than a scientist. Establish data banks of the best of the ~~3000.

Play a cooperative game of *anything you can do, I can do better*. Your attitude should be — I have this UTC and I've got to try to make it work on all the phenomena that Sam's UTC works on. Theory replication must be as important as experimental replication. In truth, we should try to make it a good deal more important. Thus, we will gradually build up to more and more integrated theories. In this respect, do not believe that to show a UTC wrong in some particulars is immediately fatal. Big investments demand that the abandonment be a deliberate business, done only with when the preponderance of evidence shows it.

(6) UTCs must be made easy to use and easy to make inferences from. This becomes especially important for theories of very wide scope and for theories that are, in important ways, embedded in simulations. We have learned in computer science the immense effort required to produce a successful programming language or a new operating system or a new statistical package. It is much more than the lone investigator with the good idea and the initial implementation — as many lone programmers have found out to their sorrow. It requires attention to the user interface, to user documentation, to training workshops, to system support.

(7) Finally, I repeat the exhortation to acquire domains of application, which can support and invigorate the theory intellectually. Building application domains, by the way, creates a community with a large investment in ease of use, and hence with a willingness to expend the effort to create the tools to make it happen.

-------o-------

So we end where we began, but with understanding.

> Psychology has arrived at the possibility of unified theories of cognition — Theories that gain their power by having a single system of mechanisms that operate together to produce results over the full range of cognition.
>
> I do not say they are here. But they are within reach and we should strive to attain them.