Building a UTC from the Top Down

What Cognitive Architectures Might Offer Theories of Executive Function in Cognitive Neuroscience

Richard L. Lewis

Department of Psychology

University of Michigan

May 6, 2001

Acknowledgments

- Thad Polk (UM Psy), Eric Freedman (Psy-UM Flint), Patrick Simen (UM EECS), Chandra (OSU CIS)

Why cognitive neuroscience needs Soar

- Part of rapid growth in cog-neuroscience is increasing focus on higher-level cognition
  - In particular, many researchers aim to understand neural underpinnings of “executive processes”
- *Could this effort be more informed by past 40 years of modeling of higher level cognition?*
  - Does a well-developed cognitive architecture such as *Soar* have anything to offer?
  - Or general functional considerations such as those outlined in *Physical Symbol Systems?*

Not a new idea

- *“This (physical symbol systems) is a genuine prediction on the structure of the nervous system and should ultimately inform the attempt to understand how the nervous system functions. It does not appear to have done so, though from time to time the suggestion has even been made directly (Newell 1962)”*
- From Physical Symbol Systems, 1980

What are the “executive functions” & “central executive”?

- Actually have a fairly long (if curious) history
  - Baddeley & Hitch (1974), Norman & Shallice ‘86
- From Miyake & Shah (1999) volume:
  - Controller of “slave” systems (phonological loop, visual-spatial sketchpad), focus/switch attention *(Baddeley & Logie)*
  - Set of processes influenced by instruction/incentive; controls attentional focus *(Cowan)*
  - Controlled attention, active maintenance *(Engle)*
  - **Scheduling of subtasks *(EPIC, Kierias et al)*
  - “Processing interactions among subsystems” *(Barnard)*
  - **Limited sequential processing network that control submodules (8 separate functions) *(CAP2, Schneider)*
  - **Active maintenance of goals, context, for purpose of modulation/gating of posterior systems *(O’Reilly, Cohen, Braver)*

The conceptual landscape

Frontal deficits
Perseveration
Distractability
Utilization
Disinhibition
Sequencing, planning
Task switching
Source amnesia

Architectures
Soar ACT-R Epic
SS Turing machine . . .

Functional generalizations
Interpretation Symbols
Working memory LTM
Composability K-search
P-search Deliberation
. . .

Possible PFC functions
Working memory Attention
Goal management
Task switching Scheduling

Properties of PFC
Processes:
excitation, gating,
inhibition, attention
Organization:
domain specificity
function specificity
Posterior connectivity

Working hypothesis

- Theories of executive function can benefit from top-down guidance from quite general functional considerations
- In particular, an explicit theory of what constitutes a *functionally adequate control structure* (Newell, 1973)
  - Note: don't be misled by the simple association of ***control*** structure with ***controlled*** processing
  - Note: we are not searching for **THE central executive**
  - Also, I will later appeal to work in *Epic* to suggest that ***executive function*** should not be directly equated with the control structure

Plan for today

(1) **Lay out four claims about functional constraints on control structure in the human cognitive architecture**

- From each, derive more specific functional requirements

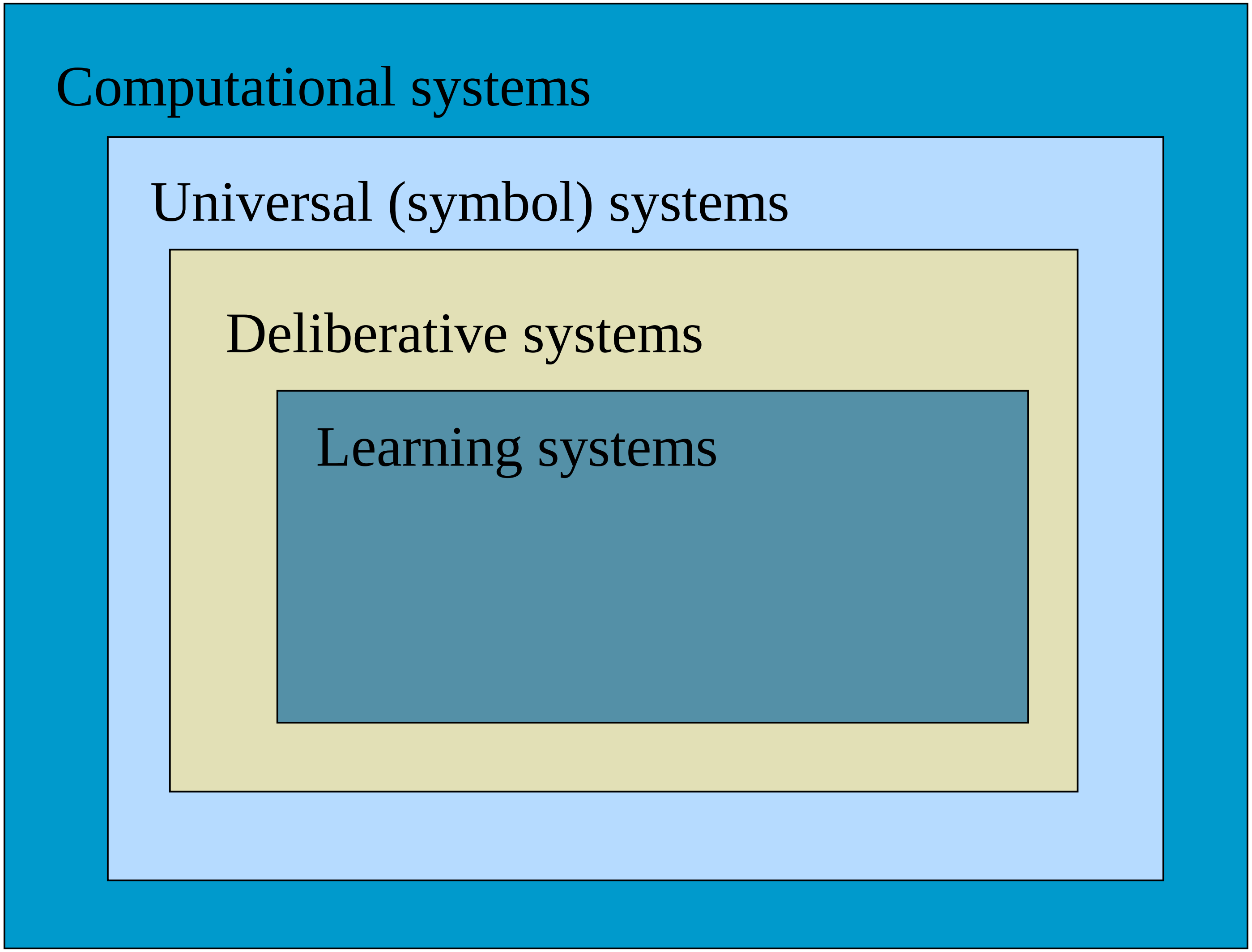
(2) **Lessons learned**

Two problems in bringing to bear high-level functional constraints

- Computer science has taught us that there is a large zoo of functionally equivalent mechanisms
- And, we must remember Newell's (1990) warnings about foundational material:

  "There is never much agreement on foundations. They are always intellectually less secure than what they support."

The big picture



Claim #1: *A theory of executive function should be, in part, a theory of control structure*

- Having a control structure is a requirement for computational completeness
  - Can think of all computational systems as having *memories, processes, control*
  - Even most elementary automata (e.g. finite state machines) have a control structure
- The control structure determines what happens next in the computation
  - There are many kinds of control structures, from single serial streams to massively parallel; centralized and distributed

Examples of control structure

- Finite state machine

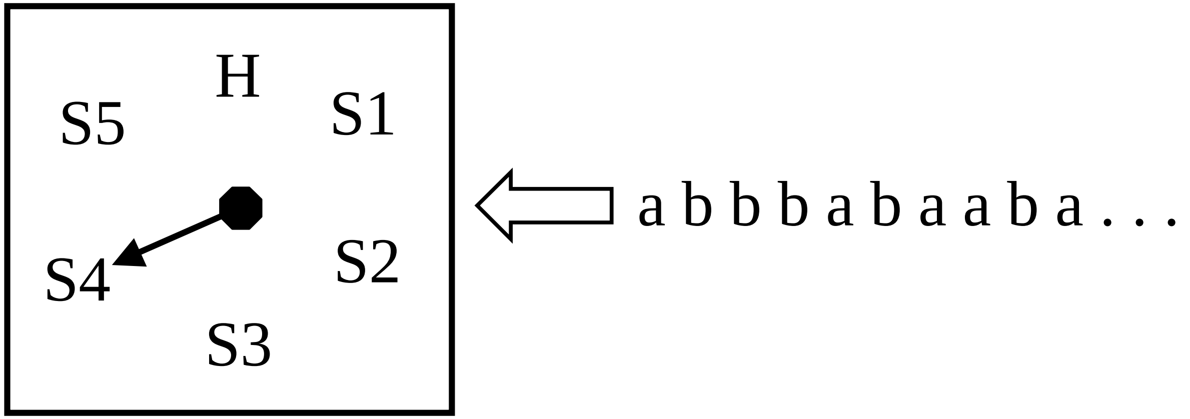
$\delta(S1,a)=S1$

$\delta(S1,b)=S2$

$\delta(S3,a)=S3$

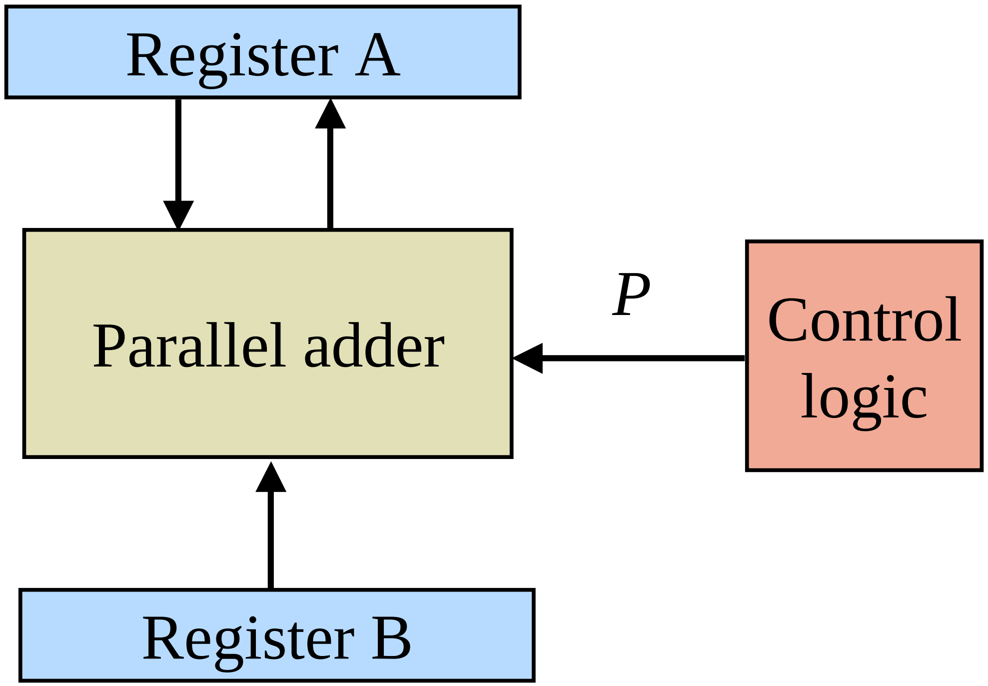
$\delta(S3,b)=S5$

. . . .



Examples of control structures

- Logic circuit implementing $P{:}A \leftarrow A+B$



The functional requirement for a control structure

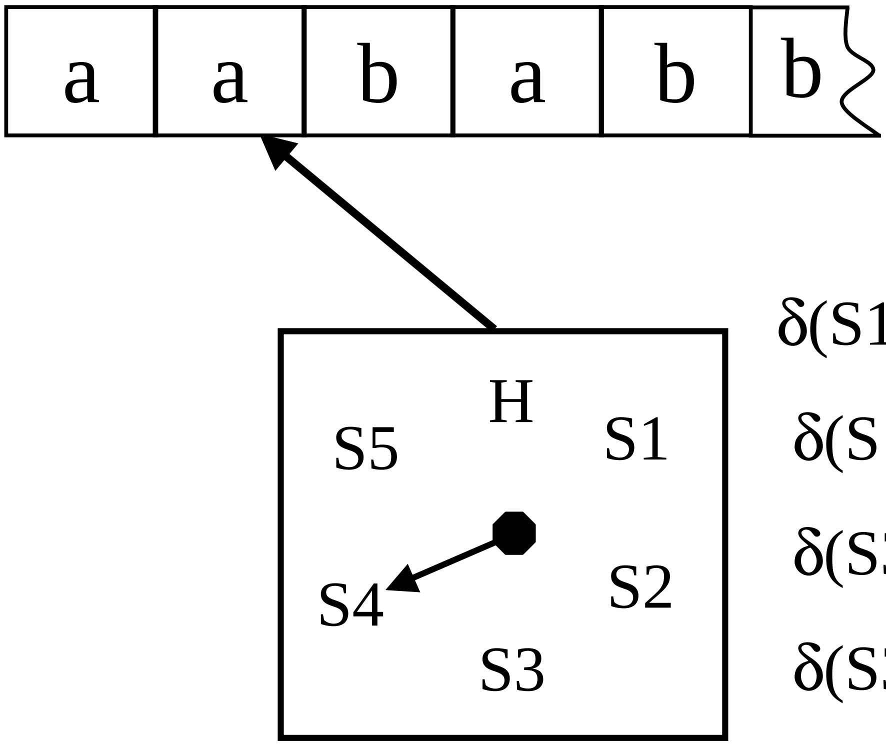
- **Control structures must specify when processes are executed.**

- Many, many ways to do this:
  - Control signals
  - Finite state controls
  - Pure parallelism
  - ……

Claim #2: *A theory of executive function must include a control structure that supports universal computation*

- Turing-universal systems the only ones we know that are sufficient to account for extreme flexibility, variety of human behavior
  - I.e., *programmable* systems
  - That humans can program themselves is a feature psychologists take advantage of in almost every experimental situation
- But what does this tell us about *control structure*?

A Turing Machine



δ(S1,a)=(S1,write a)

δ(S1,b)=(S2, move-right)

δ(S3,a)=(S3, write b)

δ(S3,b)=(S5, move-left)

A functional requirement for universal control

- **Control structures must specify how process execution is made dependent on data.**

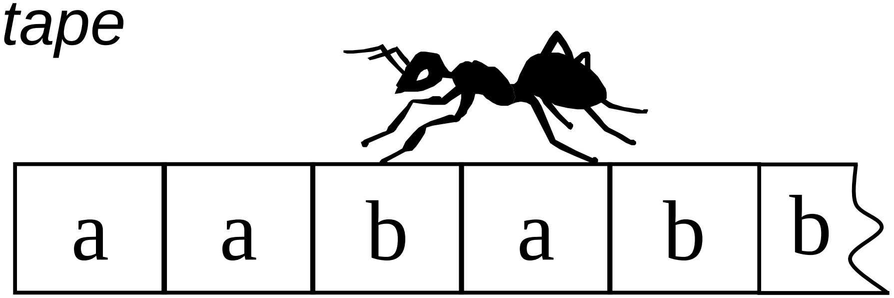
- Many, many ways to do this:
  - Conditional statements
  - Finite state control
  - Minimization functions in recursive function theory.. . . .

But wait…not *all* Turing Machines are universal!

- The *class* of TMs as a whole spans the entire class of computable functions
- But we want *one* machine that can do it all
- Only very special machines (The Universal Machines) are constructed such that they satisfy this constraint
  - By using part of the tape as a *specification* of a Turing Machine to be simulated

Problem #1: The *Finite-State Ant*

- Control structure the same across hierarchy of automata
  - Finite state machine? *A finite state control*
  - Push down automaton? *A finite state control + stack*
  - Turing machine? *A finite state control + infinite tape*
  - Universal Turing Machine? *A finite state control + infinite tape*



One step toward the resolution

- Observe that what is crucial is not simply that the control structure is *finite state*, but that it is a very *particular finite state control* that is wired up to perform a very special function

Problem #2: Zoo of equivalent mechanisms

- We know a huge range of formalisms accomplish the same thing
  - Post-production systems, recursive function theory, TM, RAMs, etc. etc.
- And even within Turing Machines, huge range of possible specific variants on UTMs
- Can anything general be said—any useful abstractions over these specifics?

The solution to the problems

- *Physical Symbol Systems* are exactly an explicit attempt to abstract the invariant functional requirements of universality

  - **Two critical functions: *designation* and *interpretation***
    - **<u>Interpretation</u>: *act of accepting as input an expression that designates process, then performing the process***
  - SS (Newell's exemplar symbol system) *builds interpretation into the control*
    - Though this is not strictly necessary

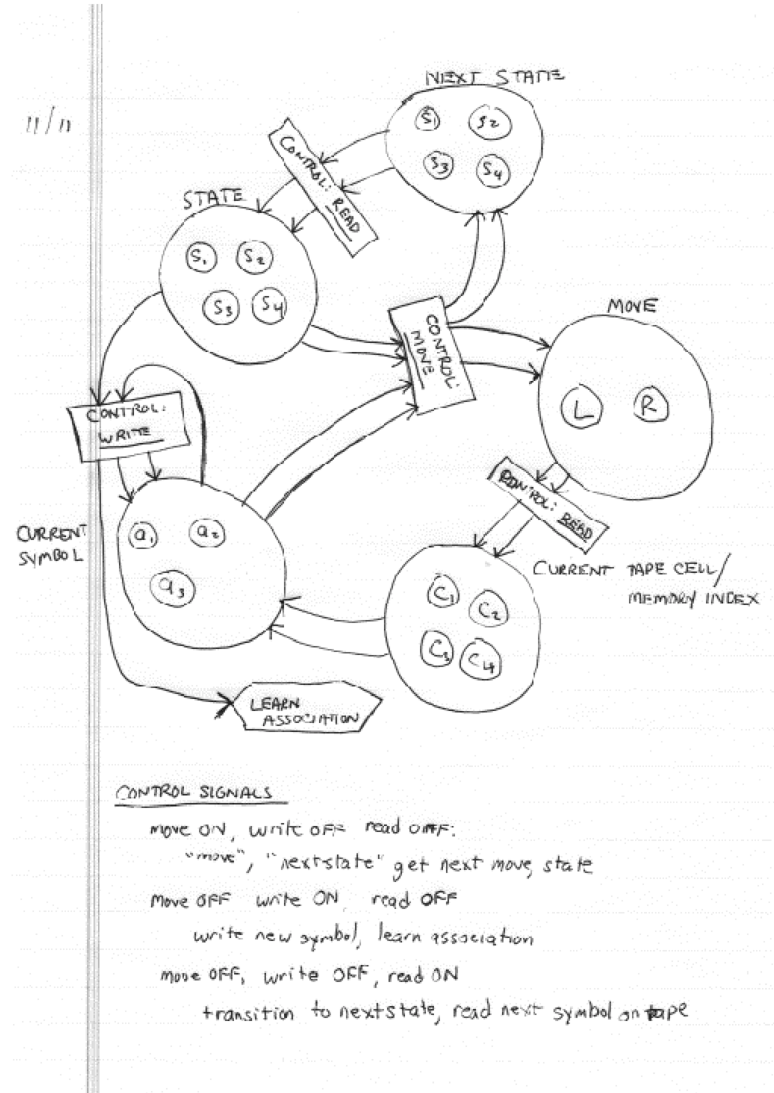
A functional requirement for interpretation

- **Requires a combinatorial medium in which to express the machine spec**

- A variety of possibilities:
  - A passive medium that can be executed
  - Or an active medium that can be evoked at a controllable time

Sidebar: Don't be misled by the simplicity of a finite state control

- I tried designing a set of attractor neural nets to simulate a Turing Machine
- It's quite difficult–there are tricky timing issues
- Problem is getting proper updating, sequencing to work



Claim #3: *A theory of executive function must include a control structure that supports deliberation*

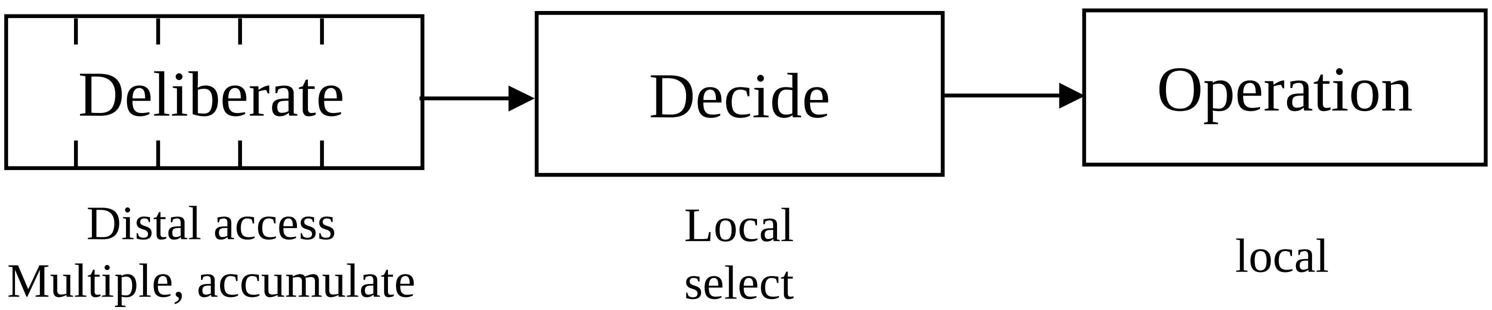
- Elementary deliberation is bringing distal knowledge to bear to choose one operation rather than others (Newell 1990)
- Actually, we can see a hint of this in the Universal Turing Machine
  - At the level of the operations of the *simulated machine*, the UTM is engaging in a search on the tape for relevant information (extracted from the spec) about what to do at each step
  - (However, the knowledge source and operations are fixed in advance)

Universal Ant problem again

- We of course know that we can construct deliberative control structures out of primitive mechanisms such as the FS control in a TM
  - But again, note that this is a *particular specialized organization* of symbol systems–not all symbol systems are deliberative, so having these primitive mechanisms alone is not enough
  - Thus we can try to describe some basic functional requirements that hold across deliberative control structures

Functional decomposition of deliberation

- Newell 1990 gives following stages



- Each stage is necessary—indicates minimum that must happen

Implications of deliberation

- There must be some way for the system to ***recognize*** **or *detect* which operations represent alternatives**

- There must be **a representation of each candidate operation that *does not itself elicit the operation***
  - This representation must be what provides **access to the distal knowledge** (I.e., it is symbolic)
  - Execution must be **delayed long enough for distal access**
  - There must be a **detection of when "long enough"** has passed

Claim #4: *A theory of executive function must include a control structure that supports learning*

- Deliberation must work for ***novel (composed) alternatives*** that have not yet been considered together
  - So it can't always be a pre-wired competition
- There must be some way to **compose the knowledge elicited by the (potentially novel) alternatives into something that permits the control structure to effect the choice**

Further implications of novel deliberation

- Control structure must support a shift from deliberate processes to automatic
  - A continuum can be defined in terms of degree of distal access before decision—for both *Generation* of alternatives and *Selection* among them

- Then raises new issue: **Must be able to control automatic processes in novel situations**
  - Particularly, get ***automatic processes to participate in deliberation***, so deliberate can be freely composed from automatic

Implementing control of automatic processes

- **In general, may require explicit *suppression* or *inhibition* functions**
- But note that Soar has solved this problem by recovery-from-error regime—forcing impasses with new alternatives
- And by living with automaticity of elaboration
  - Can't inhibit elaboration productions

Some mechanisms of deliberation

- **Soar**
  - The elegant recognize-decide-act least-commitment control structure we all know and love
  - Any operation tagged as context-slot object is deliberated on—a serial control flow (solves the problem of detecting which operations to deliberate on)
  - Common fixed language: *preferences* (solves problem of how to translate knowledge into something usable by control structure)
  - *Quiescence* (solves problem of delaying execution until right moment)

More mechanisms of deliberation

- **ACT-R**
  - All productions forced to compete with each other (A race in ACT-R 3.0; expected gain competition in 4.0)
  - Common fixed language: Time (ACT-R 3.0) or expected gain (ACT-R 4.0)
  - Though at this level, restricted knowledge source for selection (the current parameters associated with the production) — so perhaps this is *not* an architectural mechanism for deliberation

Mechanisms of deliberation, cont.

- **Epic**
  - No forced selection/deliberation: no architectural support for deliberation (**purely parallel**)
  - All programmed in advance—*all* deliberation mechanism is in the productions themselves
  - Not clear how this works with novel operations
  - At one level, can take Epic as a content theory about what we used to call ***control annotations*** in annotated models

More mechanisms of deliberation

- Polk et al Tower of London **attractor model**
  - Forces selection by arranging for each operator to *compete in the same attractor network*
  - Distal knowledge comes in form of activation from biasing goal representations (other attractors)
    - Not clear will work with novel operations
    - And, knowledge source fixed in advance

Where does this leave "Executive Processes"?

- Seems we're still a far cry from our laundry list of executive functions
  - Process scheduling, task-switching, goal management
  - Seems likely that many of these are realized as strategies (over a continuum of domain-generality/specificity), not architectural mechanisms
  - *Epic* can be seen as occupying extreme point on this issue (though we raised some concerns)
- **Executive control = fixed architecture + strategies**

Summary

- Theory of executive function must be, in part, theory of control structure
- Must support universal computation
- Must support deliberation
- Must support learning

Each of these constraints yields specific functional requirements on control structure architecture

Lessons learned from this little exercise

- This little exercise proved difficult
- It's impossible to lay out precise requirements for *mechanism*—only general functional requirements
  - Concrete architectures such as Soar help considerably in thinking through the issues
  - In fact, they may be more help than the general analysis
- Constraints associated with ***novel deliberation + learning*** seems to have the most bite
  - Suggests focusing on ***instruction taking*** will be useful strategy—programming the models in advance with task specifics may be missing the point
  - In general, should use each specific functional requirement to motivate a benchmark task (as simple as possible)

Lessons learned, cont.

- Important to keep in mind that executive functions (like any behavioral function) = *architecture* + *content*
- My hypothesis about focusing on control structure proved a little difficult to maintain
  - Notice I kept slipping in other constraints (e.g., compositionality)

- STOP