Creative-IT Project Title: A computational model of creative problem solving
Principal Investigator: Alan H. Bond
Institute for Neuroscience, University of California at Los Angeles
Discussion Section of Final Report

I investigated the idea of multiple representations which interact and each solve a different dimenson of a given problem in parallel, and exchange information from their own perspective, which has to be recognized and translated into their own representation. I concluded that this was difficult and also that the Tower of Hanoi problem did not seem to lend itself to such methods. I decided to work in the last phase of the project on understanding matching processes which were inspired by the human brain, with a view to eventually implementing a system with such matching processes.

1 The Tower of Hanoi and creativity

Even as simple, and annoying, a problem as the Tower of Hanoi does involve creativity. There are two "advanced" strategies that need some creativity to discover.
(i) Seeing the problem as moving subtowers, or pyramids, rather than as moving single disks. This has been investigated and programmed by Dirk Luiz and Allen Newell in 1989 and also by a group at Nottingham University, UK, at that time lead by Frank Ritter. It involves having a stored set of useful templates and then recognizing their usefulness and applicability to the problem.
(ii) There is a little known strategy called the parity strategy. I have not seen a treatment of it but it is mentioned in passing in a paper by Herbert Simon. From the "seed idea" that the concept of parity could be useful, humans can rapidly discover the strategy, or at least I did. Basically, you label each disk as odd or even and then the strategy is (a) to move odd and even disks alternately, and (b) viewing the towers from above, to move odd disks always counterclockwise one peg and even disks clockwise one peg, or vica versa depending on the initial Tower of Hanoi problem. This strategy generates the optimal sequence of moves, needs no memory, and will work for any number of disks.

Even just the use of previously tried sequences in new contexts can be regarded as creativity. The behaviorist Tolman noted in the thirties that rats do this.

2 My brain-inspired computer architecture

My basic architecture is a set of processing modules, corresponding to brain areas, organized as a perception-action hierarchy. Data items, which are input from sensors, stored in modules, computed in modules, transmitted between modules and output to effectors, are all information chunks, and are represented in our abstract logical model as ground logical literals. The action of each module is to receive data items of certain types and to construct new data items of types characteristic of that module, which are then output and/or stored in the module.

Thus the dynamics of the model is *constructive*. This includes creativity since it can construct novel data items, however some rules are essentially reductive in that they produce a summary or evaluation from other data.

In a simple example of organizing modules in a perception-action hierarchy, shown in Figure 1, we have goals, plans, motor and state modules, and the basic dynamics is to generate or receive goals, to prioritize and select goals to attempt, to select a plan and to elaborate the state so it can be evaluated or provide feedback to the problem solving process.

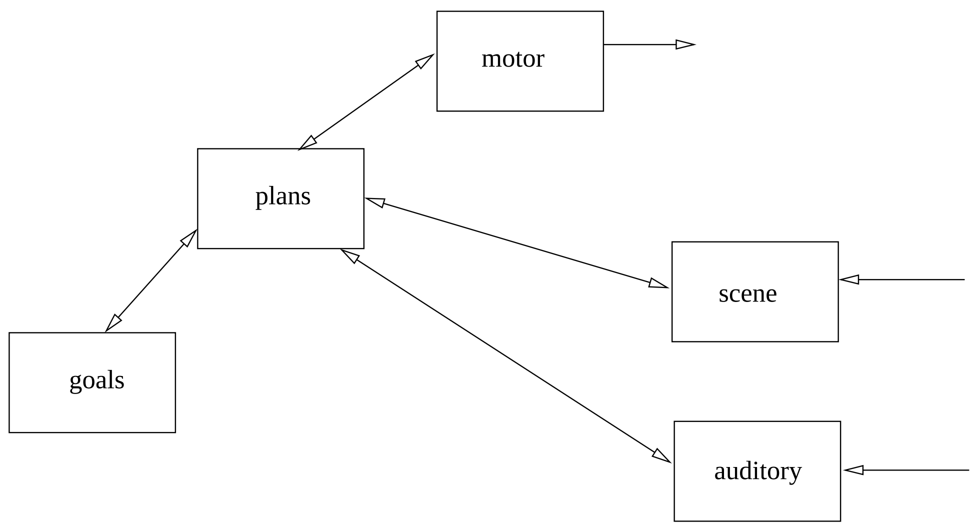

Figure 1: Simple example of a system

The modules run in a discrete time cycle, each module running to quiescence within every cycle, (actually we run to a fixed point of the logical model construction process). All of these modules run in parallel, and, typically, each module creates new data items during each cycle.

In our new design, multiple goals and plans can run in parallel with modules, only limited by shared resources. However, there is typically a dominant goal and plan which take priority. Problem solving sequencing occurs by a plan conditionally incrementing to different actions after state evaluations.

In a perception-action hierarchy, see Fig 2, there are "vertically" related modules which relate data types which are either generalizations or specializations, there are "horizontally" related modules which connect between different data types or modalities, and there are relations to combined, or polymodal, data types. These relationships, mediated by the exchange of data items, provide a basis for original constructions.

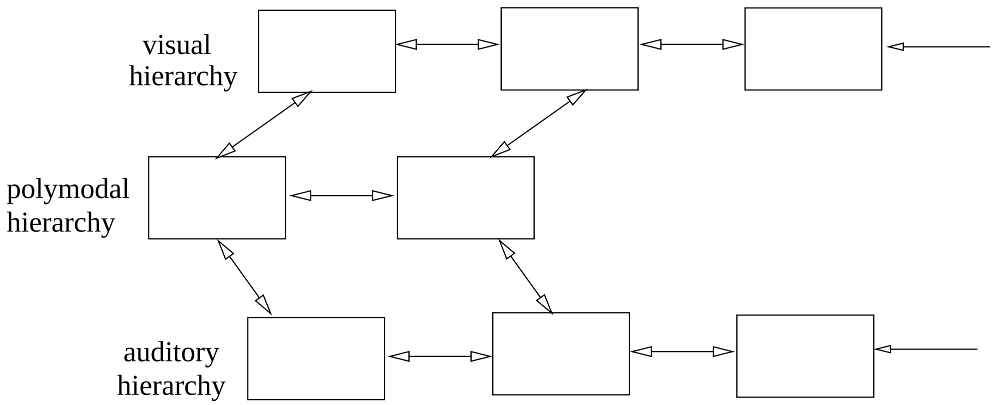

Figure 2: Perception-action hierarchy

There could be other ways of controlling the matching process, which we can think of as *attention mechanisms* and which may be quite natural in a brain-like architecture. For example, (i) emphasizing one particular datum and keeping it constant or more firm while other data are varied, and (ii) focussing on one module as the most important driver or controller of the multimodule match. Also, the system could dynamically vary its fit criterion, initially using lax criteria and later tightening them to give a small number of matches, or a unique match.

3 Events, episodic memory, and knowledge

My approach to learning new knowledge is that it is obtained by abstracting from mental events. This would also mean that all knowledge apart from innate knowledge would be based on this idea.

Events are gathered together, and accumulated and abstracted, in the hippocampal complex. A mental event is made from the set of inputs from all the different modules which are connected to the hippocampus. Basically, this includes most of the modules of the perception-action hierarchy, with the exception of some frontal areas, see Figure 3.

The modules send data items to the hippocampus which constructs a data item representing the mental event. The entire module event of each module is not sent, just an abstraction which identifies and characterizes the modular event. The complete mental event is recognized and characterized and a key computed, and this key is sent back to all the modules.

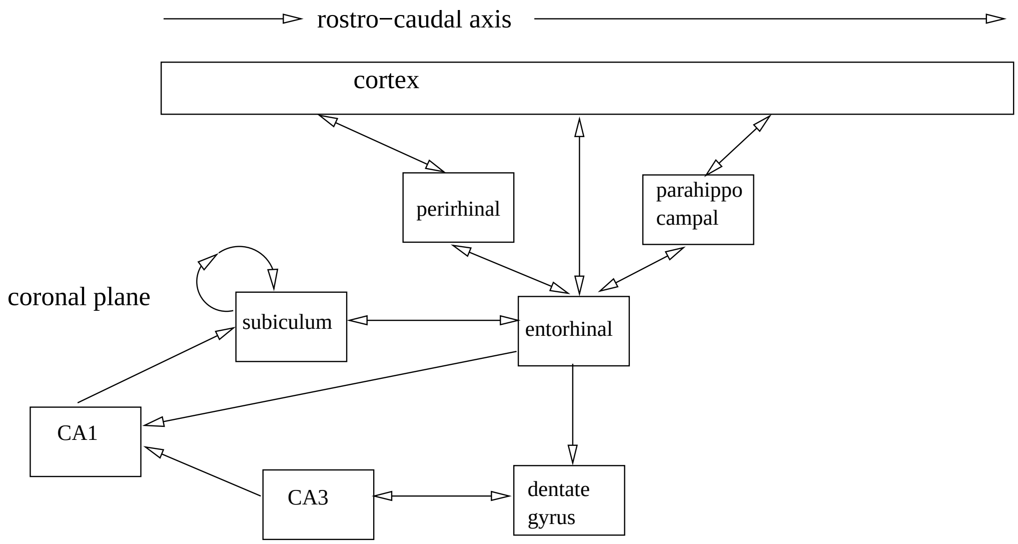

Figure 3: The hippocampal complex

Thus we end up with a distributed detailed representation of the event and modules can use the key to communicate with other modules. This produces an *autobiographical memory*, any event of which can be evoked by the modules. In addition, the sequence of events in the hippocampus is abstracted to produce knowledge items which are sent for storage in the frontal areas, and for subsequent use by planning modules. A knowledge item is a set of data items, including construction rules and also integrity constraints, which can be evoked together. This is similar to, but more general than, a problem space.

My idea for episodic memory is there could be two components which are associatively linked together, one being the memory of the spatial situation within which the event takes place and the other being the physical objects in the space, their relations and their movements. The spatial map component doesn't change as fast as the object component, so it is associated with several object component sets from several different events. Anatomically, the spatial map component is formed in the dentate gyrus, in interaction with CA3, and is stored long term in the retrosplenial area and posterior cingulate with an intermediate area which is the parahippocampal cortex. The dentate gyrus checks the incoming current spatial situation and if it has changed significantly it starts another spatial map, otherwise the current map stays the same. The CA1 and CA3 areas take the incoming event and update its object component. Incidentally, in most versions of event representations in Schank's theory, he does differentiate between events with the same spatial situation which he calls "scenes", and events involving a change of spatial situation.

My idea of a multiple representation is where each module in a perception-action hierarchy has its own (partial, incomplete, biased) representation of the event. These modules have connections which allow them to send data which influences their neighbors, see Figure 4.

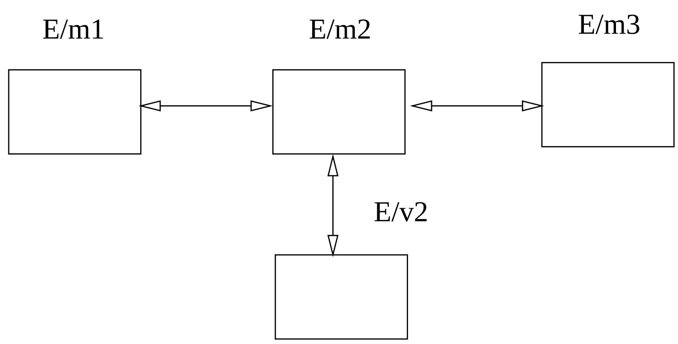

Figure 4: Modular representation of events

Thus when the system tries to match two events, usually the current event against a set of memories of events, or against an abstraction of a set of events corresponding to a plan, then each module tries to match its component and it sends out messages giving information on its match candidates. The parallel process continues until a best overall match is found. By allowing each aspect to be matched independently, the system can find new ideas, i.e., novel matches, more easily than for a single representation.

4 Brain architecture and matching at the neural level

In order to develop a matching process that was biologically inspired, I concluded that to develop ad hoc matching processes for each problem would be unlikely to produce a general approach to matching among different representations. In addition, I had proposed, and am interested in, developing mechanisms which are biologically-inspired, that is, based on the action of the human brain.

I concluded that I needed to elaborate the processing of brain modules into a more detailed processing structure based on neurons. This promised to be useful because a lot of experimental data on the brain is expressed in terms of neurons and neurotransmitters. Conversely, the action of a layer of neurons is probably limited to matching of smooth surfaces in neuron space, so that as a basis we may be able to use such well-behaved matching processes. Then the overall matching process at more aggregated and abstract levels would be determined by the neural architecture of modules.

After some study, I concluded that I could use a two step approach, first representing modules by systems made up of interconnected layers - the layer level, and second representing layers as neural nets - the neural level of representation. It seemed to me that at least the cortex and the hippocampal complex could be represented as sets of interconnected layers. By a layer I mean a two dimensional array of of neurons, usually a set of interconnected pyramidal cells and a set of interneurons also connecting them.

Depending on the layer, the pyramidal cells may be interconnected with their axons not only producing output from the layer but also reconnecting back into the set of pyramidal cells. There may be more than one type of interneuron, the different types often differ in the geometry of their connections, i.e., whether they synapse on the dendrites, cell body or axon of the target pyramidal cell. Many types of interneuron are inhibitory but some, e.g., stellate cells, are excitatory.

At the layer level of representation, a layer is represented by a set of rules which are iterated to quiescence. They will probably be simpler than the rules used at the modular level. The representation of the cortex and hippocampal complex by layers is an approximation. Although most of the connectivity is included, there are some connections which are not. The layer representation of cortical modules is diagrammed in Figure 5, and the layer representation of the hippocampal complex is diagrammed in Figure 3.

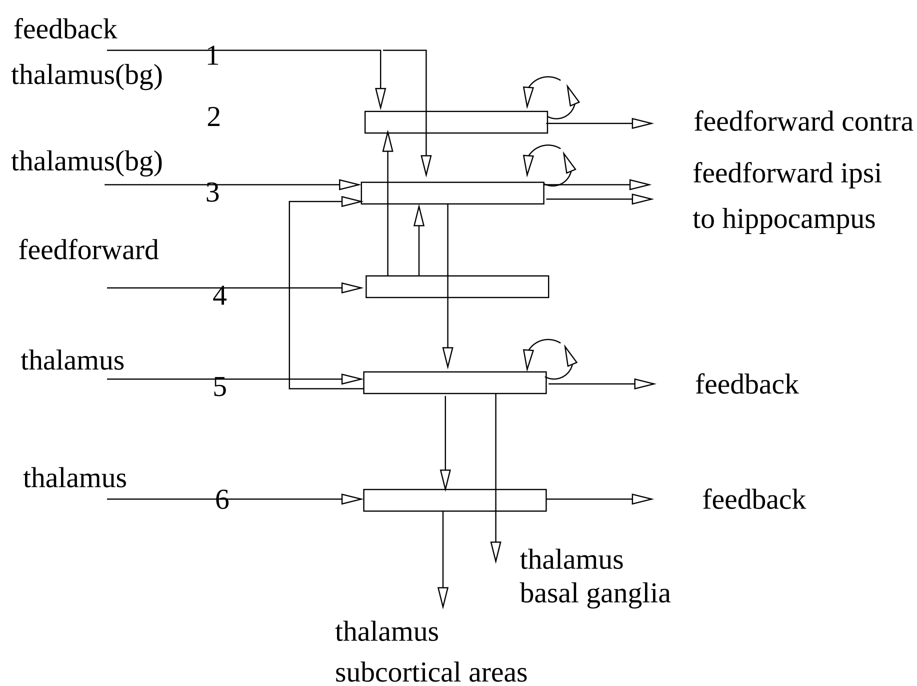

Figure 5: Layer representation of the neocortex

The layers of the hippocampal complex are the dentate gyrus, CA3, CA1 and the subiculum. The layers of the cortex are 2, 3, 5, 6, with pyramidal cells, plus 4 which has only interneurons but we will regard as a layer. The connections between layers can be the axons of interneurons; for example, interneurons in layers 2 and 3 are also connected with interneurons in layers 5 and 6. The entorhinal cortex has four layers, 2 and 3 sending input to the hippocampus and 5 and 6 receiving input from the hippocampus.

This work also includes a further year during which I had no funding.

I would like to complete this note with some personal observations on the *creative life*, by which I mean a lifestyle which is needed for creativity.

5 The creative life

Abstract. It is time to take creativity seriously.

Motivation. I am interested in creativity as a subject to understand, since I am quite a creative person. I would like to understand the dynamics of creativity so I can understand my own experience better. I think there is a total conflict between the needs of the creative mind and the current systems of research management and teaching.

Creative focus. For example, I find that I need to maintain a focus on one subject for a long time, days if not weeks. It may take a day or more to get into a focussed state. You can do other things during this time but can only have the one proper focus. By focus I mean paying attention, but mainly this is my unconscious paying attention. Consciously, I just experience knowing that I am occupied with some unknown thing. It seems that there can only be one idea or theme or problem that is paid attention to at this level. Others still are processed but at a lower level of attention, i.e. not deeply or creatively. Other ideas may need to be "parked" until cognitive resources are available.

Happiness. It is also necessary for me to be happy, I cannot do this if I am depressed. I also cannot be distracted by too many worries so I need to keep my life in some state of tidiness so this doesn't happen. However, sometimes when I am unhappy or upset or uncomfortable or even a little drunk (when I was still drinking alcohol) then I can get some lateral original ideas. But this is different from creative problem solving.

Rumination and "creative pause". I also find there seems to be a clear time when there is not point in continuing a conscious/unconscious problem solving session. Usually that is it for the day. This happens when I reach a point which is unclear and which requires "cogitation" or "rumination". The next morning enough unconscious progress may have been made to revisit the problem. Usually, some small progress has been made and then after a short time we will hit another situation where we need to pause again. Sometimes this rumination process can take a very long time, of the order of months. You have to understand this and not get anxious about it. A more extreme form is a "block"

I have not had much experience of this, although I have occasionally had pauses lasting a year or more where I could not make any progress on a certain problem.

Creativity violation. If you continue before understanding the issue, you can often get a solution but it will lack the integrity you really need. This can be dangerous in the sense that everything flowing from this solution may be a waste of time and may have to be eventually undone.

Stimulants and depressants. The use of stimulants like coffee and depressants like alcohol is a tricky area and quite idiosyncratic. I know there is a history of artists using drugs to achieve greater creativity.

Physical stress. I have noticed that conditions of physical stress such as illness or missing night of sleep can result in good ideas and insights.

Artists and creativity. I think a lot can be learned from artists who are much more attuned to creativity. They have ways of living, scheduling their work, maintaining mental states, handling states of uncertainty and crisis, making changes to work, helping each other by praise and criticism, and so on.

My experience with artists. During my thirties, I spent my time mainly with artists who were mostly painters. I was impressed with how open minded and supportive they were with each other and with me. They achieved an atmosphere of great tolerance, but included honest criticism. One type of event is of course the opening where one artist displays his work and invites everybody. This is a very testing time for the artist and also can lead to a boost to his morale, and solidarity for the group of artist friends. I don't know of any equivalent for scientists.

Standards and lack of compromise. Good artists also know about not compromising, rejecting work that is evaluated highly by others. We all know the scenario where an artist come off the stage with an enraptured audience throwing bouquets and crying "encore", and says their performance was "shit" (for example Jean Moreau in "Eve"). I recently heard on the radio a recording by a British pianist named Clifford Curzon, who had specified that the recording could not be played until after his death. Apparently there are several such recordings and his contract with the recording company allowed him to do this.

The need for creativity. I personally need creativity in others and in my environment. Sometimes I cannot stand to hear another record unless I have never heard it before. Going to an art gallery can often help. Edward De Bono said that if he got stuck on a problem he would sometimes go to Woolworths, because there is so much stuff and different gadgets and ideas that it freed his mind up.

The stark contrast between the creative life and present day "normal" behavior. Normal lifestyles involve:
(i) multitasking so we need to switch attention and solve several different problems in parallel.
(ii) time deadlines forcing the violation of the creative process.
(iii) decomposing problems into PhD sized subproblems.
(iv) lack of training in creative thought and its management.

Creative inertia. Getting into the context for solving a creative problem in my experience can take a time of the order of days and sometimes weeks. The minimum is for me to spend a whole day just "doing nothing", but this is serious nothing and needs to be tended appropriately. It may be necessary to return to consciousness and reassert the statement of the goal. Once in the context and doing unconscious creative problem solving it is very difficult to change direction to solve a different problem. It is like an oil tanker trying to change direction.

Forced expression. There seems to be a natural process of completion of a creative act, and a natural time when expression starts. Working to a deadline forces expression from an incomplete creative process. We try to avoid this by decomposing problems into doable, and completable, subproblems for which we can estimate the resources and time that will be needed for completion of the task. This tends to reduce an interesting creative problem to something mundane and uninteresting.

Failure to express. Conversely, when a creative process has completed and needs expression and this doesn't take place for some reason then the moment may be lost and the results become unretrievable. It can result in a real mess where one cannot get back into the process and cannot start a new related process either.

My own experience of the creative life. During the last six months, I have had no funding but have had enough support to survive, so I have lived something approximating the creative life. I find myself (i) pursuing questions in a lot more depth, (ii) tackling much harder problems, ones that I would normally have excluded as too hard, and (iii) I find I am understanding better, and, surprisingly, solving problems more quickly.

Policy ideas. (i) Sabbaticals should provide for CL, and (ii) research budgets and plans could specify CL for the PI for some parts of the project.

Science as art. My basic conclusion is that creative scientists should follow a lifestyle very similar to creative artists.