Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Neurocomputing 65–66 (2005) 261–273
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Representing episodic memory in a system-level model of the brain

Alan H. Bond*

Computer Science Department, California Institute of Technology, 1201 California Boulevard Mailstop 256-80, Pasadena, CA 91125, USA

Available online 15 December 2004

Abstract

We discuss the problem of finding neuroscientific and psychologically plausible representations of the memories of events and episodes. In order to do this we need to take into account the neuroanatomical connectivity between the cortex and the hippocampal complex, and also the cognitive psychology of episodic memory. We then need to develop a model of the cortex and hippocampal complex and to find representations of events that are consistent with biological information-processing constraints. We conclude that events and episodes can be represented by certain codes which are stored in associative memories.

Keywords: Event; Episodic memory; Brain model; Hippocampus

1. Introduction

The problem studied in this paper concerns how to extend our existing system-level model of the brain [3,2], to provide for episodic memory. Our approach is to use the latest experimental evidence, from neuroanatomy and from cognitive psychology, and then biological principles for information processing at the system level, in designing and implementing any model. Fig. 1 shows our brain model, and its correspondence to the neocortex. It consists of a set of interconnected modules which form a real-time control system.

*Tel.: +1 626 3956430; fax: +1 626 7924257.
E-mail address: bond@cs.caltech.edu (A.H. Bond).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2004.10.018

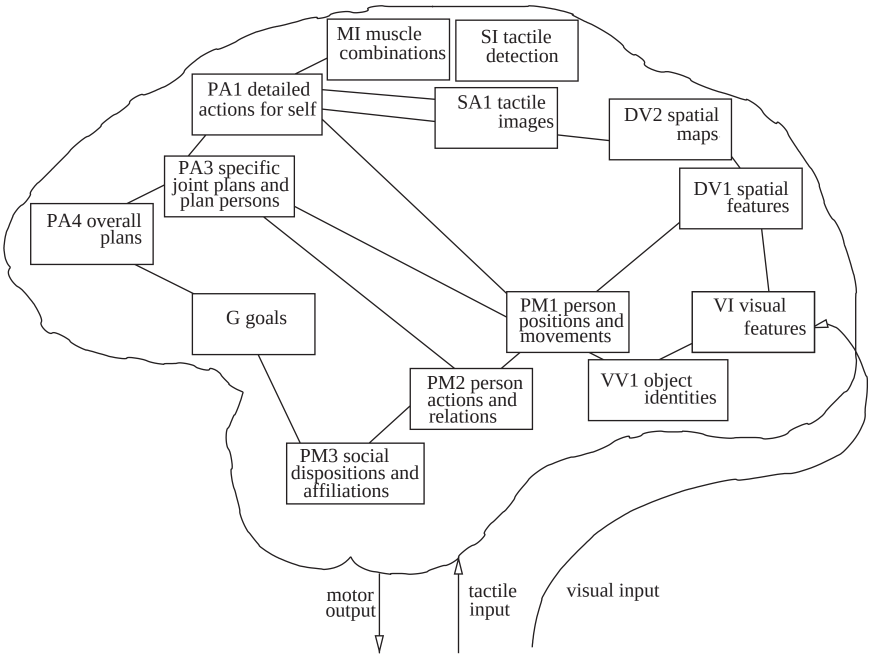

Fig. 1. Our system model shown in correspondence with the neocortex.

2. Biological information-processing principles

The basic principles of our design are derived from the biology of the neocortex:

(1) Each module stores and processes data of given types characteristic of that module; data items are of bounded size.
(2) To form systems, modules are connected in a fixed network with dedicated point-to-point channels.
(3) Modules are organized as a perception-action hierarchy.
(4) Modules process data received and/or stored locally by them. There is no central manager or controller.
(5) Modules are all rule based with a common rule execution process.
(6) All modules do similar amounts of processing and run at about the same speed.
(7) There is data parallelism in communication, storage and processing. Processing within a module is highly parallel. Parallel coded data is transmitted, stored, and triggers processing. Processing acts on parallel data to produce parallel data.
(8) The data items being transmitted, stored and processed can involve a lot of information; they can be complex.
(9) The set of modules acts continuously and in parallel.

3. Episodic memory

3.1. Definition of episodic memory

To quote the originator of the concept, Tulving [11, pp. 134–136], "The term 'episode' as used in this volume may be regarded as a close synonym of 'event', although 'episode' usually carries with it the connotation of an event that occurs in an ongoing series of events. But since we hardly ever deal with events that are not part of some ongoing series, almost all events in which we are interested are also episodes."

One event we find useful to contemplate is knocking a glass vase of flowers off a bedside table onto the floor so that it smashes. In this case, it seems that one experiences this as a single event even though it has some temporal structure. The mental representation of this event probably would include the starting state—the vase on the bedside table, an intermediate state—the vase being knocked, or falling, and the final state—the vase broken on the floor with water all around. Also there would be a sound of the vase breaking and an emotional reaction of surprise and dismay. Thus one event, according to us, can have all of these components. Interestingly it seems that this can include a visual image in a relatively direct raw form, a "snapshot", or better "videoclip", in addition to more processed and abstracted visual perceptions of the spatial situation.

3.2. Episodic memory is formed in the hippocampal complex

We will assume that learning in neocortical areas is limited to priming and possibly to learning of structure and categories within the datatypes of that area. The learning of associations between data from different areas, the learning of events, episodes and semantic facts all need the specialized learning system of the hippocampal complex.

3.3. Hippocampal complex retains a cognitive map

In the *multiple trace theory* of Nadel et al. [7], episodic and semantic memory are treated somewhat differently. The hippocampus is always involved in storage and retrieval of episodic memories, independent of their age, and even if consolidated. They reached this conclusion from a detailed study of autobiographical memory in amnesics. Even with amnesia that started later in life, recollection of memories for events earlier in life is impaired. Semantic memory however through consolidation becomes more independent of the hippocampal complex with age.

3.4. Cognitive psychology of episodic memory

Friedman [5] has reviewed episodic memory including long term and short term. We will assume that his analysis applies to both short and long term, and that long-term episodic memory has a similar form to short-term episodic memory. Friedman reviewed nine classes of theory about episodic memory.

The main finding is that episodic memory does not consist of a continuous trace, like a video tape recorder, but rather consists of a sequence of discrete records representing discrete events. Further, the relations among events are not necessarily those of temporal adjacency but of other general semantic relations. It seems that there are some temporal ordering relations however.

To quote Friedman: ''Memory for time is not built on special temporal codes or a chronologically organized memory store. Instead, our chronological sense of the past is the product of an ongoing constructive process in which we draw on, interpret, and integrate information from:

(1) our stored knowledge of time patterns,
(2) and general knowledge about time,
(3) the contextual associations of particular memories,
(4) order codes linking related events,
(5) occasional direct associations between event and time names, and
(6) rudimentary clues to the ages of memories'' [5] (*my numbering*).

3.5. Extending our model to have episodic memory

We will start from our existing methodology and existing brain model. We seek to extend our brain model to have episodic memory. This will involve adding two new modules, one for the hippocampal complex and one for the context system. The context store also includes long-term episodes. The extended model should form episodic memories, use them in short-term problem-solving, and then consolidate them to the context module, where they will influence and provide episodic and semantic knowledge for the future perception and action activity of the brain model. Our proposed scheme is diagrammed in Fig. 2.

4. Main principles of our theory

1. There are events:
   (i) They correspond to neuroanatomy, i.e., connections from cortex and amygdala to the hippocampal complex,
   (ii) the components of events represent changes as well as the current state,
   (iii) the components of events are chunked within modules before being sent to the hippocampal complex.
2. The stream of events forms episodes:
   (i) An episode is a set of events plus structuring information, such as causal relations, temporal ordering, and perception–action structure.
   (ii) Episode beginnings and endings are created by various situations as well as changes in context.
   (iii) Episodes form sequences and hierarchies.

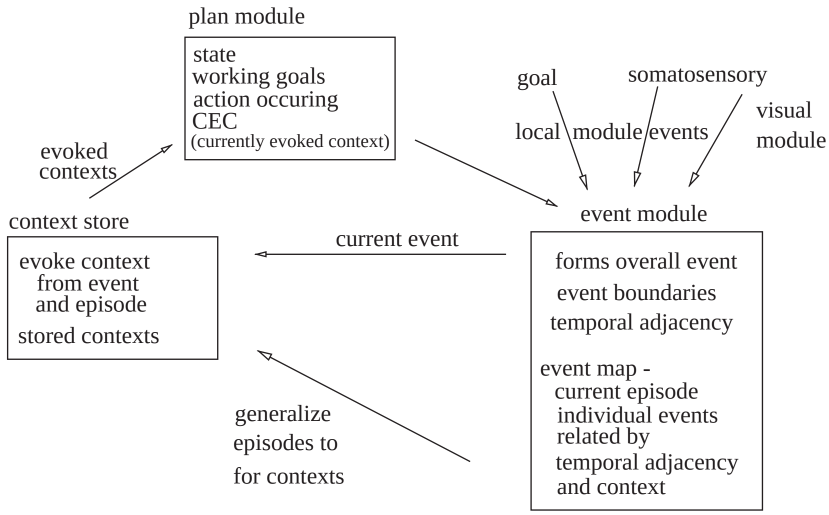

Fig. 2. The proposed basic action of our episodic memory system.

(iv) The number of events or episodes in one episode is limited, to about 4 or 5, so episodes form hierarchies with a maximal branching factor of 5.

3. The short-term episode store plays various roles in brain functioning:
   (i) Answering questions about the recent episodes.
   (ii) Reinstating events/parts of events by merging with the current state.
   (iii) The form of access to this short-term episodic memory is the same as for long-term episodic memory.
   (iv) It checks for novelty, familiarity and repeated events. This is involuntary and may reinstate previous events.
4. Episodes consolidate into long-term memory:
   (i) Long-term autobiographical memory is distributed over cortical modules with a cognitive map in the hippocampal complex.
   (ii) Contexts, comprising plan descriptions, are formed, generalized and updated and stored in the context module.
   (iii) Semantic memory also emerges, and is less distributed, being mainly stored in temporal areas.

5. A representation for events and episodes

Our concept is to use the form of episodes reported by psychologists, as explained above, and the information available for events, from the neuroanatomical connections to the hippocampal complex, to develop a representation for an event and an episode, and from this a context. At the same time, there will be constraints on the size of descriptions, they should not grow to arbitrarily large size. All accessing and retrieval has to use an associative-type memory.

5.1. Instantaneous events

The first question we must ask is what exactly is an instantaneous event, i.e., what is input to the hippocampus at each instant? The anatomical connections to and from the hippocampal complex, for the rhesus monkey, have been described by Kobayashi and Amaral [6], and are summarized in Fig. 3, taken from their paper.

5.2. Events as changes

We concluded that the information represented in a primitive event has to be a change of mental state. This is rather like one rule firing, however it corresponds to the action of many rules firing, possibly in many different modules. It can contain something about the state prior to the change, the state after the change, and something about how the change occurred. We can indicate this by a notation: event = state → change-in-state, where a change is usually some state component which has changed, for example, $e1 = s1 \rightarrow obj1$.

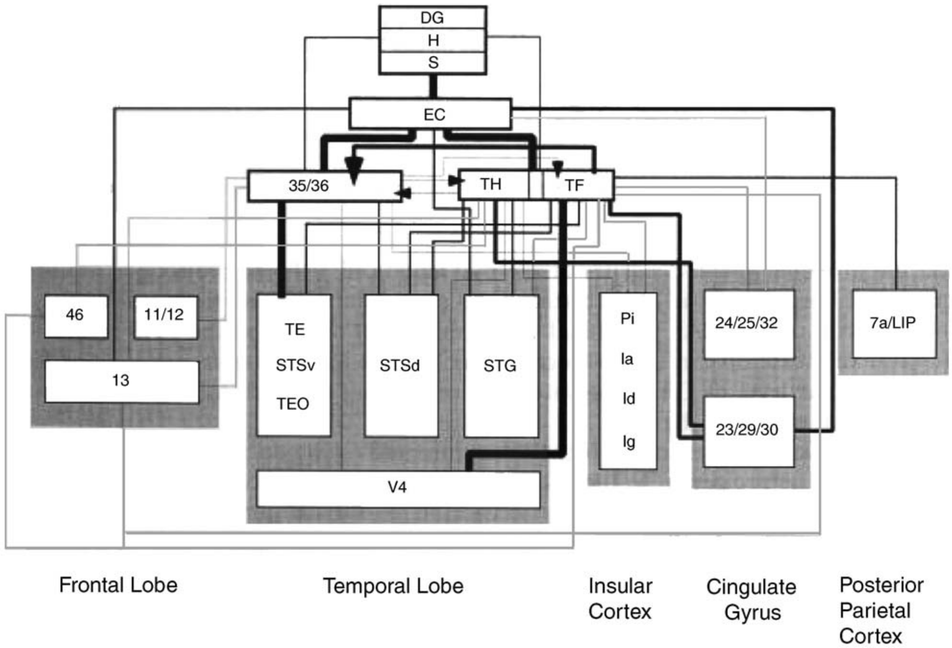

Fig. 3. Cortical connections to hippocampal complex, for rhesus monkey, from Kobayashi and Amaral [6].

A state will consist of the information reaching the hippocampal complex, which will include:

(1) Percepts such as visual, auditory, etc percepts from areas which contain processed information. For example the largest connections from the visual hierarchy to the hippocampal complex are from V4, which contains the 2.5D sketch and other 3D information.
(2) Information from frontal areas such as 9 and 46 giving the state of execution of the current plan and working memories in the planning modules.
(3) Information from anterior cingulate giving current goals.
(4) Information from posterior cingulate and parahippocampal areas giving the external spatial framework of the event.
(5) Information from the amygdala, giving the state of subcortical motivation, including subcortical goals, actions and evaluations.

5.3. Segmentation and chunking within each module

It is reasonable to assume that each module has its own temporal scale and performs segmentation of its stream of data to produce chunks, and it is these chunks that are sent to the hippocampal complex. One example is the chunking of phonemes into words by a phonological buffer.

5.4. Event descriptions

An event is a description like event(evkey,[sp1,obj1,obj2,obj3,goal1,wgoal1,-cec1,pcec1]) which a more or less fixed structure, which can be stored in the hippocampal store. The square brackets indicate a fixed number of slots rather than a list of variable length. Each component is a change received from some cortical module; usually only one, or a few, of these have changed. evkey is the event *characterization*, or *key*, which will be constructed by the hippocampal complex. It is a brief summary of the event, which includes the important features that change, maybe one or two very important features, and maybe any novel features or changes.

5.5. Uniqueness of reference to events

With a simple model, it is too difficult to obtain a unique event or episode characterization, whereas with a complex model, and therefore complex coding space, it is much easier. According to Friedman, time coding seems to definitely be present in human episodic memory, it is just not totally dominating. According to Barsalou and coworkers [1], indexing by time is the top-level organizing feature of autobiographical memory. Hence we can use a time code as part of the episode characterization, and this will provide uniqueness of reference and access. If we then scale the model to larger and richer descriptions, time coding will become less important but the characterization will continue to be unique.

The use of characterizations is an important issue since if we can use them, then we do not need to use any symbols for reference purposes, and therefore we do not have to subscribe to the physical symbol hypothesis [8]. On the other hand, we can if we wish use symbols for reference purposes in some modules. This is what we have tried to achieve since we think it correctly captures the representations used in the brain. We use chunks represented as logical descriptions, and reference is by associative matching of a chunk into an associative store of logical descriptions. Unlike symbols, chunks are not arbitrary tokens; they can be computed.

5.6. Episodes

For a sequence of events within one governing context, the system forms an episode representation as a sequence of events. Our idea, taken from the work of Wickelgren, is that the size of this sequence is limited to a small number, such as 3, 4 or 5, of events. In this case, it doesn't matter much whether we think of events as having "next" or "prior" relations, or whether we think of the episode as having a small number of slots containing event representations. This is an expression, of the form episode(epkey,[evkey1,evkey2,evkey3,evkey4]), where epkey is the key of the episode. epkey is a relatively unique description of the episode, which is a summary expression which characterizes the event, and which will also serve to allow it to be accessed in associative memory. It is of bounded size and certainly much smaller than the full description of the event. Episodes can also be composed of other episodes: episode(epkey,[epkey1,epkey2,epkey3,epkey4]).

When a further event occurs which would overload this representation, or if a change occurs that is not understood, or if a standard type of event occurs which involves a change of spatial context, a signal from an interacting partner, an emergency state, and so on, then a new episode is started. If driven by rules from a currently evoked plan then the chunking and structure of this plan influences the formation of episodes. Episode descriptions will form into shallow hierarchies, as in Fig. 4. The arrows in the diagram show associations which are not explicitly stored pointers but are instead a property of the set of descriptions. Thus what actually exists in the hippocampal complex are only descriptions such as (evkey1,[me1,me2,...]) and (epkey1,[evkey1,evkey2,...]). The system can retrieve or activate any of these descriptions by an associative match to any, or any combination, of their component fields. In order to recall a given event from a cue, the current top episode will have to queried to determine the relevant component episode and iteratively down to the event that matches the cue. Thus events are not immediately available, but a higher-level description, in terms of subepisodes, is. Hence long-term episodic memory has hierarchical structure due to episode nesting required by bandwidth constraints.

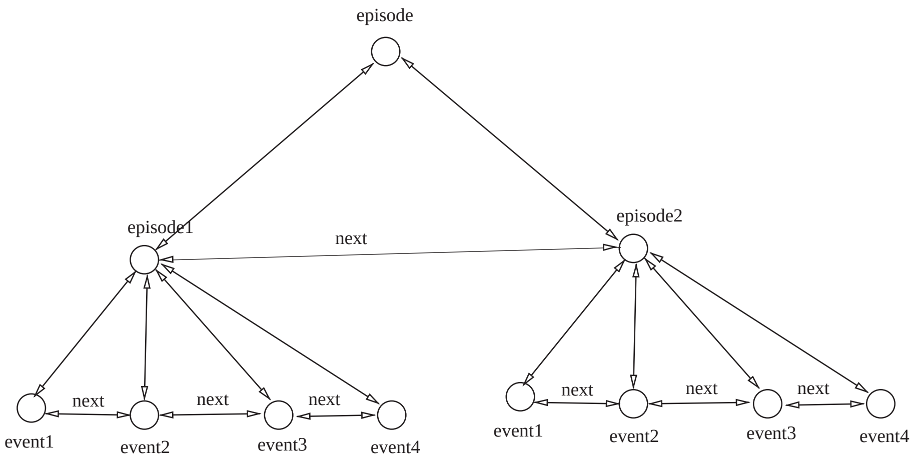

Fig. 4. An episode as a sequence of episodes.

6. Episodes in thinking

6.1. Episodic structure of problem solving

Episodes are to some extent constructed from observations, and this is how they are discussed in the psychological memory literature, however they can also be constructed from thinking, that is, sequences of mental operations.

It was observed by De Groot [4], and developed by Newell and Simon [9,10], that, in solving chess problems, people think by generating and examining possible situations in a special pattern which one can call episodic. Possible situations, starting from a given position, form a tree. Instead of searching this tree in a breadth-first or depth-first manner, for example, people search a series of narrow paths, and after each they return to the beginning to consider what will be their next path. Episodic structure, as shown in Fig. 5, is a sequence of states, with valuations, or conclusions, at the end of each subsequence. There is very little branching.

6.2. Anzai and Simon protocol

We will illustrate our ideas using the example of learning by doing in the Tower of Hanoi problem. We give the first few lines of the verbalization protocol, from Yuichiro Anzai and Herbert A. Simon, "The Theory of Learning by Doing", Psychological Review, vol. 86, pp. 124–140, 1979. We can analyze this fragment as a sequence of 3 episodes, comprising one larger episode which is one attempt at the problem.

```
1.   I'm not sure, but first I'll take 1 from A and place it on B.
2.   And I'll take 2 from A and place it on C.
```

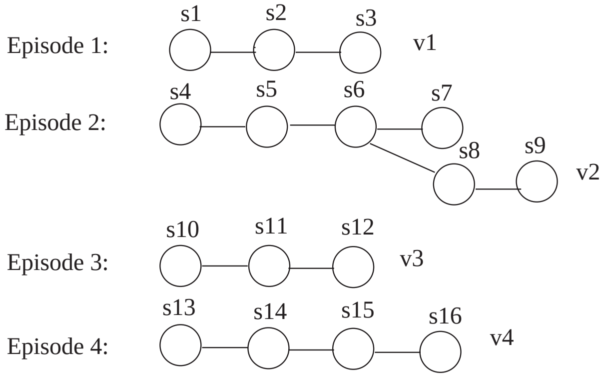

Fig. 5. Episodic structure of thinking.

3. And then, I take 1 from B and place it on C.
(If you can, tell me why you placed it there)
4. Because there was no place else to go, I had to place 1 from B to C.
5. Then, next, I placed 3 from A to B.
6. Well . . . , first I had to place 1 to B, because I had to move all disks to C. I wasn’t too sure though.
7. I thought that it would be a problem if I placed 1 on C rather than B.
8. Now I want to place 2 on top of 3, so I’ll place 1 on A.
9. Then I’ll take 2 from C, and place it on B.
10. And I’ll take 1 and . . . place it from A to B.
11. So then, 4 will go from A to C.
12. And then . ., um . . ., oh . . ., um . . .,
13. I should have placed 5 on C. But that will take time. I’ll take 1 . . (If you want to, you can start over again. If you are going to do that, tell me why.)
14. But I’ll stay with this a little more . . .
15. I’ll take 1 from B and place it on A.
16. Then I’ll take 2 from B to C.
17. Oh, this won’t do . . .
18. I’ll take 2 and place it from C to B again.
19. And then, I’ll take 1, from A . . .
20. Oh no! If I do it this way, it won’t work!
21. I’ll return it.
22. OK?
23. I’ll start over.
(Go ahead)

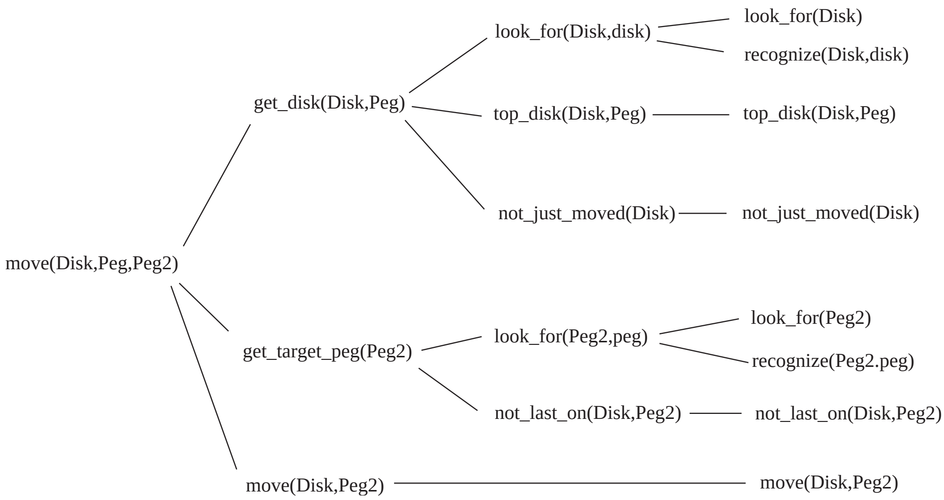

Fig. 6. An example of a plan, for the selective search strategy.

6.3. *Form of the ss plan*

We can now show in Fig. 6 the plan for the selective search strategy. This is the approximate strategy being used by the subject during the first attempt at solving the Tower of Hanoi problem. It is the repeated execution of this plan which generates the ss strategy.

7. The Tower of Hanoi example

In this section, we will give a very condensed example of the formation of episodic memory during the solution of the Tower of Hanoi problem, referring to the protocol given above.

In our formulation, initial state of the external world contains instances of on (Disk,Peg): on (1,1), on (2,1), on (3,1), etc. and the initial state of the goal module of the brain contains instances of goal(Disk,Peg): goal(1,3), goal(2,3), goal(3,3), etc.

When the system runs, it perceives the external world and this creates descriptions. Using the notation module name:module event description, we might have perc1: on(1,1), on(2,1), on(3,1); perc2: top(1,1), where perc1 and perc2 are perceptual modules, using ";"s as separators.

This may evoke context ss in the context module, which will transfer the plan ss to the plan module. The plan module generates move(1,2), which is sent to an action module.

The modular events are thus: perc1: on(1,1); perc2:top(1,1); context:context(ss); plan: plan(ss), move(1,2).

The event is: Ev1:(plan:on(1,1) → move(1,2)), [perc1:on(1,1); perc2:top(1,1); context:context(ss); plan:plan(ss), move(1,2)], using the notation event name: event key, list of modular events separated by ";"s, (however, the names of events and episodes used here are only for our ease of explanation, they are not part of the actual description). Here ss is a shorthand for the complete description of the ss strategy, which contains the ss plan as well as some other descriptions.

At the same time, an evaluation plan is launched in parallel, as part of context ss. It will detect current goal status, success and failure, progress, etc.

The system perceives the change in position: on(1,2), so the next modular events are perc1: on(1,1) → on(1,2); perc2: top(1,1) → top(2,1), top(1,2). This also terminates the previous plan step which was continuously generating move(1,2) until the result on(1,2).

This cycle continues until we reach the external state on(3,2), on(2,3), on(1,3) and eval assesses OK, need reassessment.

So we have the first episode: Ep1: [evkey1: plan:on(1,1) → move(1,2); evkey2: plan:on(2,1)→ move(2,3); evkey3: plan:on(3,1) → move(3,2)], using the notation episode name: episode key, [list of event keys separated by ";"s]. The key of Ep1, (on(3,1) → on(3,2)), is used as an index in retrieving this episode from episodic memory.

We will assume that the associative store provides access by *relative association*, by which we mean that there is an existing currently retrieved item and then the key will find an item stored relative to it. We can represent this in our formalism by giving the complete global key, but in more realistic implementations, only the local key would be needed.

The modular events are stored in their respective modules associated with the key of the event of which they are a component. The original modular event corresponds to a short term memory description, d, and this is augmented by the hippocampal complex to the description (event key,d) which will become long-term memory in the same module, although it will also be available as short-term memory before the LTM process has had time to store it, i.e., for about 15 min. For example, perc1: the STM modular event on(1,1)→ on(1,2), gives the (STM then) LTM modular event (plan:on(1,1) → move(1,2))(on(1,1) → on(1,2)).

There is then a sequence of such episodes: Ep2: goal(2,3), move(1,1), move(2,2), move(1,2), move(4,3), evaluation - incorrect, need 5 → 3, where the episode key is on(4,1) → on(4,2), and Ep3: move(1,1), move(2,3), no, move(2,2), move(1,-), evaluation - won't work.

At the end of this sequence, the episodic memory in the hippocampal complex contains a *cognitive map*, formed from the episode hierarchy, which is made up of the following hierarchy using episode keys: ((on(3,1) → on(4,2)), [on(3,1) → on(3,2), on(4,1) → on(4,2), on(1,1) → on(2,2)]).

The sequence of these episodes could form what we call a *tail structure*, i.e., the next episode could be taken to always be the last subepisode of the current episode.

The system also forms what we call *header structures*, which contain a description of the basic overall mental state relative to which all these events take place: header(on(1,1), on(2,1), on(3,1),...goal(1,3), goal(2,3), goal(3,3),...). This has associations with the episode representations. We speculate that the dentate gyrus has the right properties to form header structures.

Note that in the example protocol fragment there is in addition a verbalization process where the subject is asked, i.e., given a goal, to explain verbally how they have been thinking. This probably involves accessing the cognitive map in the hippocampal complex and short term memory in modules, as well as the construction of sentences.

8. Summary and conclusion

We have analyzed the information involved in memories of events and episodes, and have derived representations satisfying neuroanatomical and psychological findings, as well as biological information-processing principles. This representation can be used in abstract system-level models, and also used as a neural code for neural net models.

We have argued that events and episodes are represented by certain characterizing descriptions called keys, which are formed in the hippocampus, are stored in associative memories, and are retrievable by association. In addition, such keys are consolidated into cortical areas as components of longer term memory descriptions. Episode information links module event information, and allows for cross-evocation of event components of different modalities. Indexing is achieved at the top-level of chunking using timing information and at lower levels using keys.

References

[1] L.W. Barsalou, The content and organization of autobiographical memories, in: U. Neisser, E. Winograd (Eds.), Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory, Cambridge University Press, Cambridge, England, 1988, pp. 193–243.

[2] A.H. Bond, A computational model for the primate neocortex based on its functional architecture, J. Theoret. Biol. 227 (2004) 81–102.

[3] A.H. Bond, An information-processing analysis of the functional architecture of the primate neocortex, J. Theoret. Biol. 227 (2004) 51–79.

[4] A.D. de Groot, Het Denken van den Schaker, North-Holland, Amsterdam, 1946 (Translation into English and published as Thought and Choice in Chess, Mouton, 1965.).

[5] W.J. Friedman, Memory for the time of past events, Psychol. Rev. 113 (1993) 44–66.

[6] Y. Kobayashi, D.G. Amaral, Chemical neuroanatomy of the hippocampal formation and the perirhinal and parahippocampal cortices, in: F.E. Bloom, A. Bjorklund, T. Hokfelt (Eds.), Handbook of Chemical Neuroanatomy, vol. 15, The Primate Nervous System, Part III, Elsevier, Amsterdam, 1999, pp. 285–401.

[7] L. Nadel, A. Samsonovich, L. Ryan, M. Moscovitch, Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results, Hippocampus 10 (2000) 352–368.

[8] A. Newell, Physical symbol systems, Cog. Sci. 4 (1980) 135–183.

[9] A. Newell, H. Simon, An example of human chess playing the light of chess playing programs, in: N. Wiener, J.P. Schade, (Eds.), Progress in Biocybernetics, vol. 2, Elsevier Science Publishers B.V., Amsterdam, 1965, pp. 19–75.

[10] A. Newell, H. Simon, Human Problem Solving, Prentice Hall, Englewood Cliffs, NJ, 1972.

[11] E. Tulving, Elements of Episodic Memory, Oxford University Press, Oxford, 1983.