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

Alan H. Bond
California Institute of Technology[1]

**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. Indexing at the top-level of chunking uses timing information.

**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 [Bond, 1996] [Bond, 1999b] [Bond, 1999a] [Bond, 2002] [Bond, 2003b] [Bond, 2003a], 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.

Figure 1 shows a simple brain model, and its correspondence to the neocortex. It consists of a set of interconnected modules which form a real-time control system.

[1]California Institute of Technology, Computer Science Department, Mailstop 256-80, Pasadena, California 91125. email: bond@cs.caltech.edu

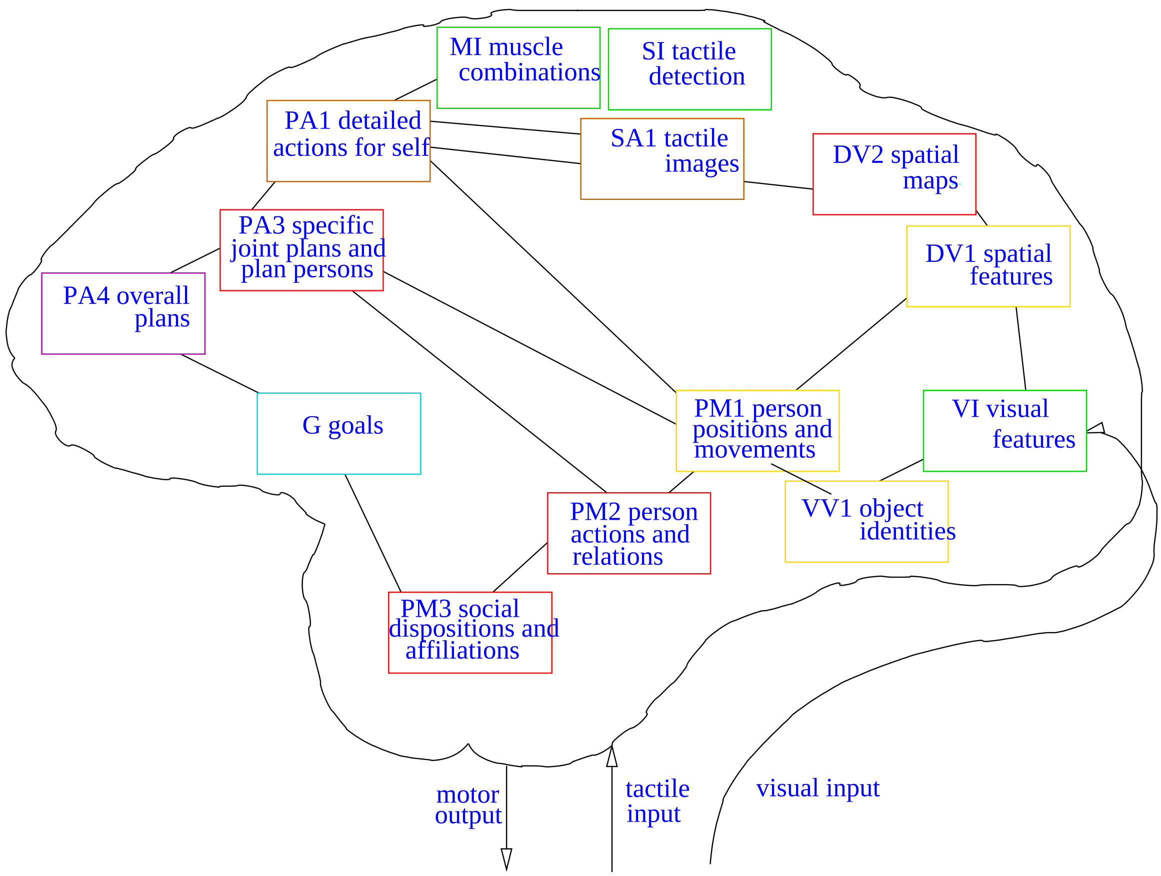


Figure 1: Our system model shown in correspondence with the neocortex

## 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.

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 pro-

cessing. 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 modules act continuously and in parallel; they do not wait for other modules or data.
10. Most of the learning occurs in specialized learning modules, which dynamically create knowledge for modules.

During the design, we have tried to adhere to our basic principles derived from biology. In particular, data items are of bounded size, since the channels have fixed bandwidth, and we also want rule processing to have bounded data width. In order to represent larger structures, such as maps for example, a set of data items is used, rather than a very large data item.

## 3 Different kinds of memory

Psychological thinking is in the main independent of considerations of how memory is achieved or implemented in the brain. We will break memory into (i) short term and of limited capacity, (ii) working memory which is of limited capacity and time, (iii) short-term episodic memory, of large capacity although limited indexability, for events in the last 15 minutes or so, (iv) long-term episodic, (v) semantic, and (vi) procedural.

A mainstream view, and our idea of how this maps onto the brain is as follows:
(i) sensory buffers - areas in unimodal hierarchies
(ii) working memory - areas in prefrontal cortex, mainly in area 46 or the principal sulcus.
(iii) short-term episodic memory - in the hippocampal complex
(iv) long-term memory - originates in the hippocampal complex and is generated by consolidation in cortex, long-term episodic mainly in ventral prefrontal areas, and semantic memory mainly in temporal areas
(v) procedural memory in the basal ganglia and in various cortical areas.

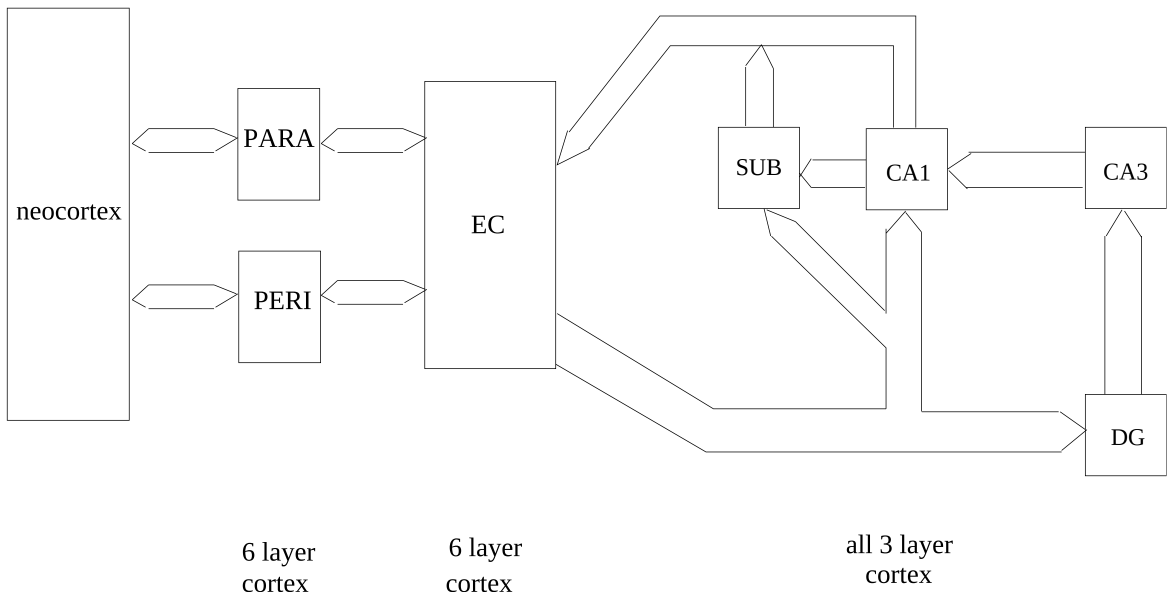


Figure 2: The hippocampal formation, block diagram

# 4 Neuroanatomy for episodic memory

Cohen and Eichenbaum's 1993 book [Cohen and Eichenbaum, 1993] gives a very clear and comprehensive treatment of the hippocampus and its function. Detailed neuroanatomy and connectivity for rhesus monkeys can be found in [Kobayashi and Amaral, 1999]. Treatments of the human neuroanatomy can be found in [Paxinos, 1990] and [Parent, 1996].

Following McClelland's review [McClelland et al., 1995], 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 formation.

Figure 2 shows the connectivity of the components of the hippocampal complex as a block diagram. We will often refer to the hippocampal complex, or hippocampal formation, as the hippocampus.

The standard idea of the hippocampal complex is that two things occur. First, it creates associations between a stimulus and a context. This is conceived as concerning a particular moment in time rather than over an extended interval of time. A context is the larger percept including perceived distal objects, features in peripheral vision, etc. We assume this also includes all other modalities, the gustatory, olfactory, visceral, somatosensory and auditory backgrounds. Second, subsequently over a period of time, this episodic memory is copied, or re-represented, to reside in the neocortex. This is *consolidation* of long-term memory. After consolidation, the memory no longer resides in the hippocampus, but before consolidation is complete it resides in the hippocampus and can be retrieved from there as required. The experimental basis for this idea is from amnesics with hippocampal damage, and from temporary retrograde amnesia caused by trauma (retrograde means concerning memory for events that occurred before the trauma occurred, as opposed to anterograde). The long term memory for general episodes which are over extended intervals of time is referred to as autobiographical memory and also is created in the hippocampus. Semantic memory for facts is treated the same as episodic memory. Procedural memory, the memory of how to perform action sequences, and skills, tends to be independent of hippocampal involvement.

Nadel, Samsonovich, Ryan and Moscovitch [Nadel et al., 2000] have challenged this standard idea. In their *multiple trace theory (MTT)*, 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 hippocampus with age.

# 5 Episodic memory

**The definition of episodic memory.** To quote the originator of the concept, Endel Tulving [Tulving, 1983], pp. 134-6:
"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."

**Underlying mechanisms of episodic memory.** We are interested first of all in short-term episodic memory, which we defined to be the memory of events that occurred in the last 15 minutes or so. After or concurrently with this there will also be long-term episodic memory which is autobiographical memory and which is mediated by a different neural mechanism than short-term episodic memory. The characteristic time for LTP and SCP mechanisms is about 20 minutes, and also modification due to RNA transcription using induced transcription factors takes a similar amount of time. The process of copying or re-remembering from short term to long term corresponds to consolidation.

The short term memory mechanism is very fast and is "elastic" in the sense that the energy involved is quickly reusable. The short-term episodic-memory traces decay rapidly over a period of a few minutes, and the energy is recovered and used to create new episodic memory traces.

**The cognitive psychology of episodic memory.** Friedman[Friedman, 1993] 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." [Friedman, 1993] (*my numbering*).

**Temporal indexing.** The representation of events in autobiographical memory has been discussed by Lawrence Barsalou [Barsalou, 1988]. His main conclusion was that the top level of indexing is based on time.

**Rhythms and clocks.** There could well be variables with diurnal and other rhythms [Treisman, 1963] which would be input to the hippocampal formation. This can be thought of as a set of values of clock phases for a set of clocks. These would be folded into the stored episodes. The current phases could then be used in reminding, and would be available on recall of event components and event reinstatement.

# 6 Our overall approach to memory

We plan to eventually model several different mechanisms of memory and learning which serve complementary roles. Let us briefly indicate a possible overall scheme for this:
(i) memory for objects and persons is stored in cortex and is subject to facilitation/priming from successful use. Objects and persons are represented by distributed representations made up of components in different cortical areas. Each cortical area learns component data of the data types processed and stored by that area,
(ii) memory for associations among different data types from different areas is formed in parahippocampal and perirhinal areas, where it continues to function for some time before being consolidated into cortical areas as records, which contain associations to other areas involved,
(iii) memory for objects in context is formed in the hippocampus, and also consolidated to cortical areas,
(iv) memory for temporal sequences of events is also formed in the hippocampus,
(v) these association and event memories are available to the rest of the brain as working memory, in the short run,
(vi) the influence of reward in memories is received by the hippocampus from the amygdala, and associations between events and reward are sent to the amygdala from the hippocampus,
(vii) Consolidation results in long-term memories which are distributed over the cortex, different components being stored as data of different types in corresponding cortical areas. These components can cross evoke each other, and there is an index or map, storing inter-item relations, in the hippocampus.
(viii) Access to long-term memories is via prefrontal retrieval actions.
(ix) Episodes will be stored autobiographically, but in addition, generalized episodes call context will be formed. Contexts contain sets of generalized events which are plans.
(x) Particular contexts are evoked from the context store and are used by planning

modules to create action.
(xi) procedural memory is laid down separatly in the basal ganglia but its long-term use is controlled by cortical areas, notably prefrontal.

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 system 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.

We have previously argued for episodic/context memories being stored in ventral prefrontal areas [Bond, 2003b]. For example, Shallice et al. [Shallice et al., 1994], using PET, isolated acquisition from retrieval of verbal episodic memories. Retrieval was associated with activity in right areas 10, 46 and 12 of prefrontal cortex (12 being strongest) and the bilateral precuneus (31). Left anterior cingulate (32) was active in both acquisition and retrieval.

# 7 Main principles of our theory

1. There are events:
(i) they correspond to neuroanatomy, i.e., connections from cortex and amygdala to the hippocampal formation
(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 hippocampus.

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
(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 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.

# 8 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.

**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 [Kobayashi and Amaral, 1999], and are summarized in Figure 3, taken from their paper.

Thus what is input is basically the perception by the person of their own plans and actions, as well as input from the anterior cingulate. There is also subcortical input, mainly from the amygdala.

Thus what is remembered in an event, input to the learning module, is that part of the

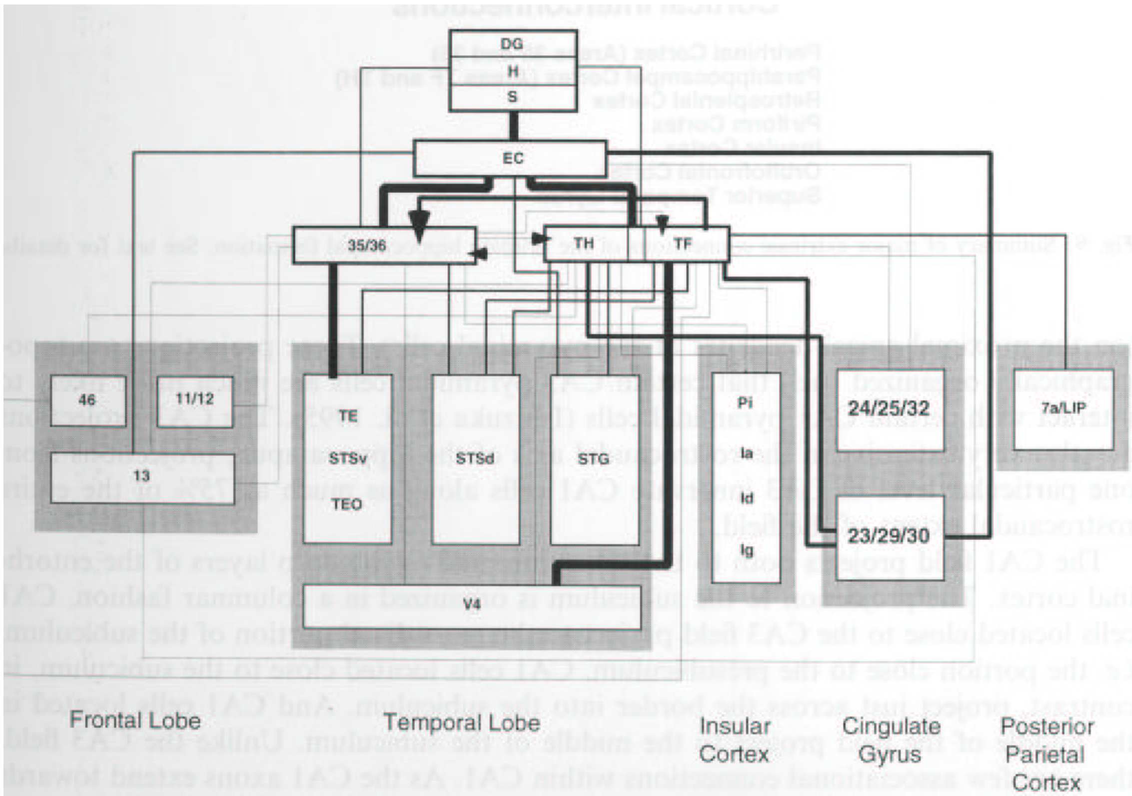


Figure 3: Cortical connections to hippocampal complex, for rhesus monkey, from Kobayashi and Amaral [Kobayashi and Amaral, 1999]

behavioral, or mental, state which includes:
(i) percepts of external events, objects and people
(ii) percepts of the self as an external person
(iii) percepts of the internal activities of the self
(iv) motivational states and goals of the self, and
(v) percepts of others' intentions.

**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 hippocampus. One example is the chunking of phonemes into words by a phonological buffer.

**The concept of event.** 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 me, 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.

We concluded that the information represented in a primitive event has to be a change of state or perception. 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 = initial_state $->$ final_state, for example, e1 = s1 $->$ s2.

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.

**Event descriptions.** An event is a description like event(evdes,[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. There is a relatively unique descriptor of the event which we denote by evdes, which is a summary expression which characterizes the event, and which will serve to allow it to be accessed in associative memory. evdes is typically an expression but it is of limited size and certainly much smaller than the full description of the event. cec and pcec are also summary descriptors of these evoked contexts. We will call this summary expression the *characterization* of the event.

**Uniqueness of reference to events.** (i) With a simple model, it is too difficult to obtain a unique episode characterization, whereas with a complex model, and therefore coding space, it is much easier.

(ii) 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 [Barsalou, 1988] [Barsalou, 1992] [Lancaster and Barsalou, 1997], indexing by time is the top-level organizing feature of autobiographical memory.

(iii) Hence we will use a time code as part of the episode characterization, and this will provide uniqueness of reference and access.

If we then were to scale the model to larger and richer descriptions, time coding will become less important but the characterization will continue to be unique.

To give an abstract example, an event is a description which is a fixed structure of components corresponding to the different chunks received from modules. Then the characterization would include the important features that change, maybe one or two very important features, and maybe any novel features or changes.

The use of characterizations is a key issue since if we can use them, then we do not need to use any symbols for reference purposes. We do not have to subscribe to the physical symbol hypothesis [Newell, 1980]. 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.

**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 [Wickelgren, 1981], 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 again is an expression, of the form episode(epdes1,[evdes1,evdes2,evdes3,evdes4]),

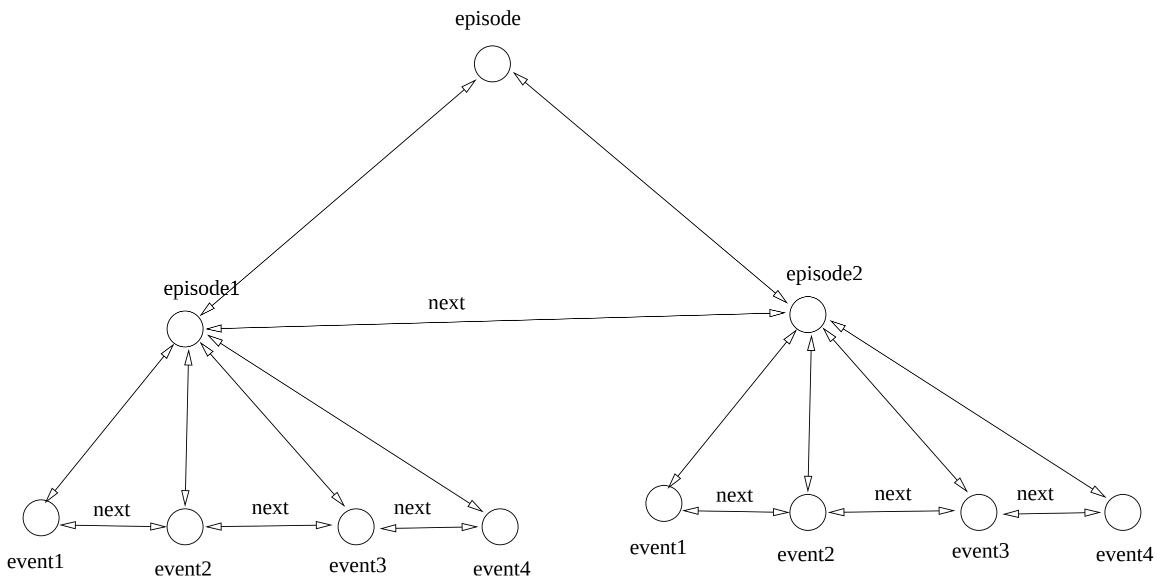


Figure 4: An episode as a sequence of episodes

where epdes is a characterization. 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 context then the chunking and structure of this context determines the formation of episodes.

Hence, in general, episodes will form into shallow hierarchies, as in Figure 4.

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.

Episodes and events will initially be stored in a short-term episodic-memory module, corresponding to the hippocampal complex. It will have connections from the cortex, and also the amygdala if one is included in the model.

# 9 Consolidation and long term memory

We assume that the episodic store can only remember episodes for a certain limited time, such as 15 minutes or so. During this time, another mechanism is able to make another representation which is longer lived. Most of this representation will be in the cortex and not in the episodic memory.

We imagine that in this long-term memory representation each event will be decomposed and redistributed back to the modules from which it originally came. This allows the different data types to be handled by the appropriate modules. However in addition the event component will carry an event descriptor and episode descriptors. These descriptors will allow the future reconstruction of the event from the event descriptor. By using a single event descriptor sent to all modules, all the components of that past event will be evoked simultaneously, recreating and reliving the past event. By attention control, some part only of a past event may be evoked, and also modified or combined with other evoked event memories. In addition, following Nadel and Moscovitch, there is a residual map of episodes and their relations, but without any concrete event information. This episode map is stored in the episodic memory by another longer term memory mechanism.

This approach allows the same query to access episode information from the episodic module in the shorter term and from the cortex in the longer term.

In general, we assume that queries are generated from plans in the planning module. In addition, there may be more routine plans not in the planning module, which are evoked as routine steps by the planning module but executed in other areas such as parietal areas. We will discuss this issue later. These routine steps may also generate queries to be sent to, and responded by, episodic memory.

It is our belief that short-term episodic memory is where the main working memory is to be found. There is other working memory in cortical modules but this is very much limited to a few items only in each module, and of the data types handled by that module. This frontal working memory corresponds to registers on a computer, which are a small number of implicit stores used for immediate working data.

## 10 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 codes which are stored in associative memories, and are retrievable by association. In addition, such codes are consolidated into cortical areas into longer term memory mechanisms. 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 by timing information and at lower levels by characterization codes which are formed in the hippocampus.

## References


[Barsalou, 1988] Barsalou, L. W. (1988). The content and organization of autobiographical memories. pages 193–243. in [Neisser and Winograd, 1988].

[Barsalou, 1992] Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In Kittay, E. and Lehrer, A., editors, *Frames, fields, and contrasts: New essays in semantic and lexical organization*, pages 21–74. Lawrence Erlbaum Associates.

[Bloom et al., 1999] Bloom, F. E., Bjorklund, A., and Hokfelt, T. (1999). *Handbook of Chemical Neuroanatomy, Volume 15 The Primate Nervous System, Part III.* Elsevier Science Publishers B.V.

[Bond, 1996] Bond, A. H. (1996). A Computational Architecture for Social Agents. In *Proceedings of Intelligent Systems: A Semiotic Perspective, An International Multidisciplinary Conference, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, Oct 20-23*.

[Bond, 1999a] Bond, A. H. (1999a). A System Model of the Primate Neocortex. *Neurocomputing*, 26-27:617–623.

[Bond, 1999b] Bond, A. H. (1999b). Describing Behavioral States using a System Model of the Primate Brain. *American Journal of Primatology*, 49:315–388.

[Bond, 2002] Bond, A. H. (2002). Problem-solving behavior in a system model of the primate neocortex. *Neurocomputing*, 44-46C:735–742.

[Bond, 2003a] Bond, A. H. (2003a). A Computational Model for the Primate Brain based on its Functional Architecture. accepted for publication in Journal of Theoretical Biology.

[Bond, 2003b] Bond, A. H. (2003b). An Information-processing Analysis of the Functional Architecture of the Primate Neocortex. accepted for publication in Journal of Theoretical Biology.

[Cohen and Eichenbaum, 1993] Cohen, N. J. and Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. M.I.T. Press, Cambridge, Massachusetts.

[Friedman, 1993] Friedman, W. J. (1993). Memory for the Time of Past Events. *Psychological Review*, 113:44–66.

[Kobayashi and Amaral, 1999] Kobayashi, Y. and Amaral, D. G. (1999). Chemical neuroanatomy of the hippocampal formation and the perirhinal and parahippocampal cortices. pages 285–401. in [Bloom et al., 1999].

[Lancaster and Barsalou, 1997] Lancaster, J. S. and Barsalou, L. W. (1997). Multiple organisations of events in memory. *Memory*, 5:569–599.

[McClelland et al., 1995] McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102:419–457.

[Nadel et al., 2000] Nadel, L., Samsonovich, A., Ryan, L., and Moscovitch, M. (2000). Multiple Trace Theory of Human Memory: Computational, Neuroimaging, and Neuropsychological Results. *Hippocampus*, 10:352–368.

[Neisser and Winograd, 1988] Neisser, U. and Winograd, E. (1988). *Remembering reconsidered: Ecological and traditional approaches to the study of memory*. Cambridge University Press, Cambridge, England.

[Newell, 1980] Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4:135–183.

[Parent, 1996] Parent, A. (1996). *Carpenter's Human Neuroanatomy, Ninth Edition*. Williams and Wilkins, Baltimore.

[Paxinos, 1990] Paxinos, G. (1990). *The Human Nervous System*. Academic Press, London.

[Shallice et al., 1994] Shallice, T., Fletcher, P., Frith, C. D., Grasby, P., Frackowiak, R. S., and Dolan, R. J. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature*, 368:633–635.

[Treisman, 1963] Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the internal clock. *Psychologica lMonographs*, 77(13, Whole no. 576).

[Tulving, 1983] Tulving, E. (1983). *Elements of episodic memory.* Oxford University Press, Oxford.

[Wickelgren, 1981] Wickelgren, W. A. (1981). Human learning and memory. *Annual Review of Psychology*, 32:21–52.