

The Role of Consciousness in Memory

Stan Franklin¹, Bernard J. Baars², Uma Ramamurthy¹, Matthew Ventura³

Abstract:

Conscious events interact with memory systems in learning, rehearsal and retrieval (Ebbinghaus 1885/1964; Tulving 1985). Here we present hypotheses that arise from the IDA model (Franklin, Kelemen and McCauley 1998; Franklin 2001b) of global workspace theory (Baars 1988, 2002). Our primary tool for this exploration is a flexible cognitive cycle employed by the IDA model and hypothesized to be a basic element of human cognitive processing. Since cognitive cycles are hypothesized to occur five to ten times a second and include interaction between conscious contents and several of the memory systems, they provide the means for an exceptionally fine-grained analysis of various cognitive tasks. We apply this tool to the small effect size of subliminal learning compared to supraliminal learning, process dissociation, implicit learning, recognition vs. recall, and the availability heuristic in recall. The IDA model elucidates the role of consciousness in the updating of perceptual memory, transient episodic memory, and procedural memory. In most cases, memory is hypothesized to interact with conscious events for its normal functioning.

Consciousness and memory: Basic facts to be accounted for

Human memory seems to come in myriad forms: sensory, procedural, working, declarative, episodic, semantic, long-term memory, long-term working memory and perhaps others. How to make sense of all of this? And to add to the difficulty, these terms are used differently in different research traditions. Psychologists tend to use these terms to refer inferentially to systems that appear to hold memory traces and to the underlying knowledge that constitutes their contents. To computer scientists and to neuroscientists, memory refers only to the physical (not inferred) storage device. Further, in many cognitive studies consciousness is either taken for granted or labeled with its own set of synonyms such as explicit cognition, focal attention, and awareness. Yet the role of consciousness has concerned memory researchers since Ebbinghaus (1885/ 1964).

Because the question of the role of consciousness in memory is so broad, we attempt to approach it in a qualitative way, rather than trying to make precise, quantitative, but local, predictions (Broadbent 1958, Posner 1982).

The following methods have been used to dissociate consciousness and unconscious aspects of learning, memory storage and retrieval. They must therefore be accounted for by an adequate theory of the role of consciousness in memory.

1. **Subliminal vs. supraliminal learning.** Although subliminal acquisition of information appears to occur, the effect sizes are quite small compared to conscious learning. In a classic study, Standing (1973) showed that 10,000

¹Computer Science Department and the Institute for Intelligent Systems, The University of Memphis

² The Neuroscience Institute, San Diego and the Institute for Intelligent Systems, The University of Memphis

³ Psychology Department and the Institute for Intelligent Systems, The University of Memphis

distinct pictures could be learned with 96% recognition accuracy, after only 5 seconds of conscious exposure to each picture. No intention to learn was needed. Consciously learned educational material has been recalled after 50 years (Bahrick, 1984). No effects nearly as robust as this have been reported in the subliminal learning literature (Greenwald et al, 1996; Elliott & Dolan, 1998). Conscious access greatly facilitates most types of learning.

2. **Process dissociation** is one of the major efforts today to separate conscious and unconscious components in memory (Jacoby 1991). In a typical example, subjects are asked to avoid recalling items that they know from a previous list. Items that are nevertheless recalled are assumed to be influenced by unconscious source knowledge. In this fashion, conscious versus unconscious aspects of memorized lists can be dissociated.
3. **Implicit learning.** In this paradigm subjects are always asked to attend to (and therefore become conscious of) a set of stimuli. The organizing regularities among the attended stimuli --- such as grammars that generated those stimuli --- can be shown to be learned, even though people are not generally aware of them. Nevertheless, because the stimulus set is always conscious, conscious access appears to be a necessary condition even for implicit learning of inferred rules and regularities. (Reber et al, 1991)
4. **Recall vs. recognition.** There is considerable evidence that people are conscious of retrieved memories in recall, but not necessarily in recognition tasks. (e.g., Gardiner et al, 1998). For early memory researchers like Ebbinghaus, indeed, the term “recall” meant retrieval to consciousness. The feeling of knowing that characterizes recognition is a “fringe conscious” phenomenon, that is, an event that has high accuracy but low reported conscious content (Baars, 2002). In numerous experiments these differences result in striking dissociations between subjective reports in “remember” vs. “know” types of retrieval.
5. **Working memory.** In cognitive working memory, the active operations of input, rehearsal, recall and report are conscious (Baddeley 1993). The contents of working memory prior to retrieval are not. A companion article by Baars & Franklin describes the way IDA, a large-scale model of GW theory, accounts for this evidence. (Baars & Franklin, 2003).
6. **Availability heuristic.** In problem solving, people tend to select the first solution that becomes conscious. This heuristic can lead to predictable patterns of errors (Kahneman, Slovic and Tversky 1982; Fiske and Taylor 1991). The IDA model can account for these patterns qualitatively.
7. **Automaticity.** Predictable skills tend to become more automatic, that is less conscious, with practice. When such automatized skills repeatedly fail to accomplish their objectives, consciousness tends to return. The IDA model’s account of these processes is the subject of a separate article (Negatu, McCauley and Franklin in preparation).

Novel Hypotheses

With its finer-grained model of these processes, the IDA model (Franklin 2000a, 2001; Franklin & Graesser 2001; Franklin 2002) of global workspace theory (Baars 1988, 1993, 1997, 2002) offers hypotheses that suggest a simple account of several forms of human memory and their relationships with conscious events. This account addresses each of the issues raised in the previous section (see below). With this article, we hope to entice experimentalists to test our hypotheses. (See also Baars and Franklin 2003) Here are the major hypotheses:

- 1. The Cognitive Cycle:** Recall William James' claim that the stream of conscious thought consists of momentary "flights" and somewhat longer "perches" of dwelling on a particular conscious event. Such findings have been reported by neuroscientists (Halgren et al 2002, Fuster et al, 2000; Lehmann et al 1998; Freeman, 2003b)⁴. Much of human cognition functions by means of continual interactions between conscious content, the various memory systems and decision-making. We call these interactions, as modeled in IDA, cognitive cycles. While these cycles can overlap, producing parallel actions, they must preserve the seriality of consciousness. The IDA model suggests therefore that consciousness occurs as a sequence of discrete, coherent episodes separated by quite short periods of no conscious content (see also VanRullen and Koch 2003).
- 2. Transient Episodic Memory:** Humans have a content-addressable, associative, transient episodic memory with a decay rate measured in hours (Conway 2001). In our theory, a conscious event is stored in transient episodic memory by a broadcast from a global workspace. A corollary to this hypothesis says that conscious contents can only be encoded (consolidated) in long-term declarative memory via transient episodic memory.
- 3. Perceptual Memory:** A perceptual memory, distinct from semantic memory but storing some of the same contents, exists in humans, and plays a central role in the assigning of interpretations to incoming stimuli. The conscious broadcast begins and updates the process of learning to recognize and to categorize, both employing perceptual memory.
- 4. Procedural Memory:** Procedural skills are shaped by reinforcement learning operating through consciousness over more than one cognitive cycle.
- 5. Consciousness:** Consciousness is implemented computationally by way of a broadcast of contents from a global workspace, which receives input from the senses and from memory (Baars 2002).
- 6. Conscious Learning:** Significant learning takes place via the interaction of consciousness with the various memory systems (e.g. Standing 1973; Baddeley 1993). The effect size of subliminal learning is therefore small compared to the learning of conscious events, but significant implicit learning can occur by way of unconscious inferences based on conscious patterns of input (Reber et al,

⁴ To say that conscious moments may be separated by short periods of no conscious contents does not mean, of course, that people momentarily fall asleep between conscious episodes.

1991). All memory systems rely on consciousness for their updating, either in the course of a single cycle or over multiple cycles.

7. **Voluntary and Automatic Attention:** In the GW/IDA model, attention is defined as the process of bringing contents to consciousness. Automatic attention may occur unconsciously and without effort during a single cognitive cycle (Logan 1992). Attention may also occur voluntarily and effortfully in a conscious, goal-directed way, over multiple cycles.
8. **Voluntary and Automatic Memory Retrievals:** Associations from transient episodic and declarative memory are retrieved automatically and unconsciously during each cognitive cycle. Voluntary retrieval from these memory systems may occur over multiple cycles using conscious goals.

Global Workspace Theory as a Functional Interpretation of Consciousness

Global workspace theory is a cognitive architecture with an explicit role for consciousness. It makes minimal assumptions:

1. That the brain may be viewed as a collection of distributed *specialized networks* (processors);
2. That consciousness is associated with a *global workspace* -- a fleeting memory capacity whose focal contents are widely distributed (“broadcast”) to many unconscious specialized networks;
3. That some unconscious networks, called *contexts*, shape conscious contents (for example, unconscious parietal maps of the visual field modulate feature cells needed for conscious vision);
4. That such contexts may work together to jointly constrain conscious events;
5. That motives and emotions can be viewed as part of *goal contexts*;
6. That executive functions work as *hierarchies of goal contexts*.

A number of these functions have plausible brain correlates, and the theory has recently gathered considerable interest in cognitive neuroscience and philosophy (see Table 1).

Table 1. Globalist views of consciousness⁵

“Conscious contents provide the nervous system with <i>coherent, global information</i> .”	Baars 1983
“Each specialized neural circuit enables the processing and mental representation of specific aspects of conscious experience, with the outputs of the distributed circuits integrated via a large-scale neuronal ‘workspace.’” (p. 161)	Cooney & Gazzaniga 2003
“... time-locked multiregional retroactivation of widespread fragment records...can become contents of consciousness.”	Damasio 1989
“We postulate that this global availability of information through the workspace is what we subjectively experience as the conscious state.”	Dehaene & Naccache 2001

⁵ Adapted from Baars, 2002

“Theorists are converging from quite different quarters on a version of the global neuronal workspace model of consciousness...”	Dennett 2001
“When we become aware of something ... it is as if, suddenly, many different parts of our brain were privy to information that was previously confined to some specialized subsystem. ... the wide distribution of information is guaranteed mechanistically by <i>thalamocortical and corticocortical reentry</i> , which facilitates the interactions among distant regions of the brain.”	Edelman & Tononi 2000
“The activity patterns that are formed by the dynamics are spread out over large areas of cortex, not concentrated at points. Motor outflow is likewise globally distributed.... In other words, the pattern categorization (corresponds) to an induction of a global activity pattern.”	Freeman in press
“Evidence has been steadily accumulating that information about a stimulus complex is distributed to many neuronal populations dispersed throughout the brain.”	John et al 2001
“... awareness of a particular element of perceptual information must entail not just a strong enough neural representation of information, but also access to that information by most of the rest of the mind/brain.”	Kanwisher 2001 p.105.
“... <i>the thalamus represents a hub from which any site in the cortex can communicate with any other such site or sites.</i> ... temporal coincidence of specific and non-specific thalamic activity generates the functional states that characterize human (conscious) cognition.	Llinas et al 1998
“The high resting brain activity is proposed to include the global interactions constituting the subjective aspects of consciousness. “	Shulman et al 2003
“... the brain ... transiently settling into a globally consistent state (is) the basis for the unity of mind familiar from everyday experience.”	F. Varela et al 2001 p. 237.

The IDA Model

We now turn to the software agent model of Global Workspace Theory (Baars 1988, 1997, 2002) that gave rise to the hypotheses listed above. IDA (Intelligent Distribution Agent) is an intelligent software agent (Franklin and Graesser 1997) developed for the US Navy (Franklin et al. 1998). At the end of each sailor's tour of duty, he or she is assigned to a new billet. This assignment process is called distribution. The Navy employs some 300 people, called detailers, full time to effect these new assignments. IDA facilitates this process by completely automating the role of the human detailer (Franklin 2001a). The IDA software agent is up and running, and has matched the performance of Navy detailers.

The IDA model includes the architecture and mechanisms of both the computational model, the currently implemented portions, and the overarching conceptual IDA model. The IDA model implements Baars' global workspace theory of consciousness (1988, 1997, 2002), and is compatible with a number of other psychological theories (Baddeley 1993; Ericsson & Kintsch 1995; Gray 1995; Glenberg 1997; Barsalou 1999; Sloman 1999; Conway 2001).

Most computational models of cognitive processes are designed to predict experimental data. When successful, their structure and parameters serve to explain the

cognitive data. These models are typically narrow in scope, though a few attempt to provide a single, unified mechanism for cognition (see for example Laird et al. 1987 or Anderson 1990).

IDA, on the other hand, is an intelligent software agent (Franklin and Graesser 1997) designed and developed for practical personnel work for the U. S. Navy. While the IDA model was not initially developed to reproduce experimental data, it is based on psychological theories, and does generate hypotheses and qualitative predictions. But why should one give any credence to hypotheses derived from the IDA model? One argument is that IDA successfully implements much of global workspace theory, and there is a growing body of empirical evidence supporting that theory (Baars 2002, Dehaene et al 2003). IDA's flexible cognitive cycle has also been used to analyze the relationship of consciousness to working memory at a fine level of detail, offering explanations of such classical working memory tasks as visual imagery to gain information and the rehearsal of a telephone number (Baars and Franklin 2003). There's also evidence from neurobiology in the form of "...hemisphere-wide, self-organized patterns of perceptual neural activity..." recurring aperiodically at intervals of 100 to 200 ms (Freeman 2003a, Freeman 2003b, Lehmann et al 1998, Koenig, Kochi and Lehmann 1998). In the present paper, hypotheses from the IDA model will be used to analyze several major findings of existing memory literature in relation to consciousness (see the section on Basic Facts above.) Other supporting evidence related to the frame-rate is mentioned below early in the section on IDA's cognitive cycle and the memory hypothesis.

Memory Systems and Terminology

In this section, we will briefly discuss the various human memory systems that will play a role in the rest of the article. It will be helpful to the reader to specify here how we plan to use the various terms, as there isn't always agreement in the literature. Figure 1 displays some of the relations between the memory systems we'll discuss.

Sensory memory holds incoming sensory data in sensory registers and is relatively unprocessed. It provides a workspace for integrating the features from which representations of objects and their relations are constructed. There are different sensory memory registers for different senses, iconic (visual), echoic, haptic, and likely a separate sensory memory for integrating multimodal information. Sensory memory has the fastest decay rate, measured in hundreds of milliseconds.

Working memory is the manipulatable scratchpad of the mind (Miyake and Shah 1999). It holds sensory data, both endogenous (for example, visual images and inner speech) and exogenous (sensory), together with their interpretations. Its decay rate is measured in seconds. Again, there are separate working memory components associated with the different senses, the visuo-spatial sketchpad and the phonological loop for example (Baddeley 1993; Franklin & Baars 2003). Also, there are long-term processing components of working memory (Ericsson and Kintsch 1995). Baars & Franklin (2003) have suggested that conscious input, rehearsal and retrieval are necessary for the normal functions of working memory.

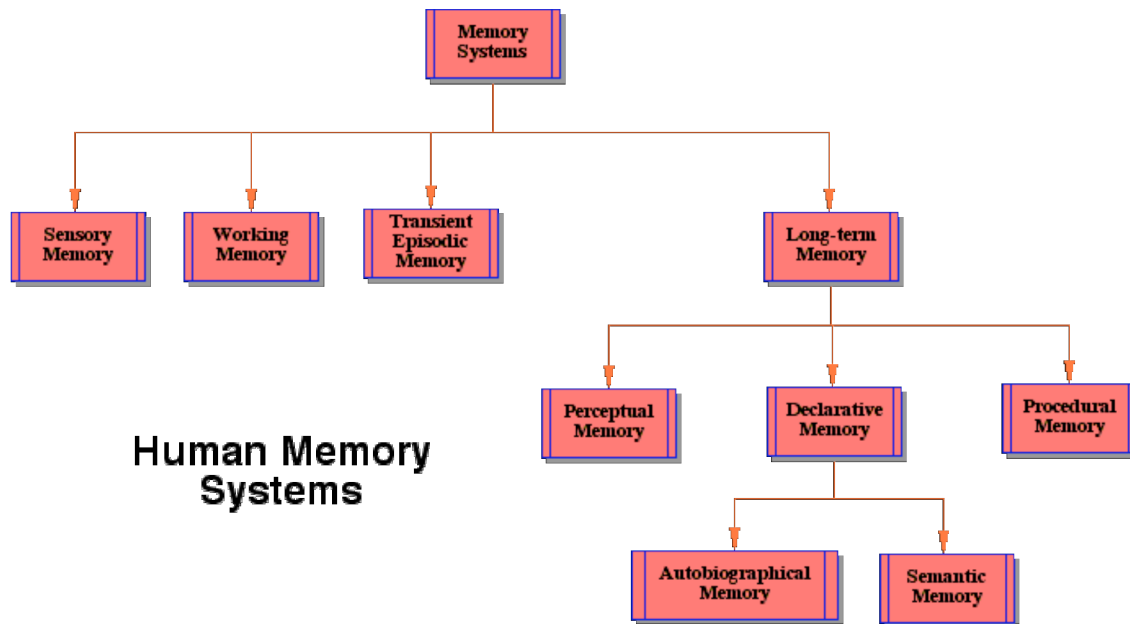


Figure 1. Human Memory Systems

Episodic or autobiographical memory is memory for events having features of particular time and place (Baddeley, Conway and Aggleton 2001). This memory system is associative and content-addressable.

An unusual aspect of the IDA model is transient episodic memory (TEM), an episodic memory with a decay rate measured in hours. Though often assumed (Panksepp 1998, p 129 assumes a “transient memory store”), the existence of such a memory has rarely been explicitly asserted (Donald 2001; Conway 2001; Baars and Franklin 2003). It will play a major role in the hypotheses about memory systems generated by the IDA model.

Humans are blessed with a variety of long-term memory types that decay exceedingly slowly, if at all. Memory researchers typically distinguish between procedural memory, the memory for motor skills including verbal skills, and declarative memory. In the IDA model, declarative memory (DM) is composed of autobiographical memory, described in a previous paragraph, and semantic memory, memories of fact or belief typically lacking a particular source with a time and place of acquisition. In contrast, semantic memories have lost their association with their original autobiographical source. DM is a single system within the IDA model. These declarative memory systems are accessed by means of specific cues from working memory. The IDA model presumes that DM decays inversely with the strength of the memory traces (see Figure 5 for the general shape of the curve).

Though “perceptual memory” is often used synonymously with “sensory memory,” following Taylor, we use the term differently (1999 p. 29). Like semantic memory, perceptual memory is a memory for individuals, categories, and their relations. The IDA model distinguishes between semantic memory and perceptual memory (PM) and hypothesizes distinct mechanisms for each. According to the model, PM plays the major role in recognition, in categorization, and more generally in the assignment of interpretations. Upon presentation of features of an incoming stimulus, PM returns interpretations. The content of semantic memory is hypothesized to be a superset of that

of PM. All this essentially restates the most controversial part of our Perceptual Memory Hypothesis, the claim of distinct mechanisms for PM and semantic memory. Several types of evidence, of varying degrees of persuasiveness, support this dissociation.

We speculate that perceptual memory is evolutionarily older than TEM or declarative memory. The functions of perceptual memory, the recognition of food items, nest-mates, prey, predators, potential mates, etc. seems almost universal among animals including insects (Beekman et al 2002), fish (Kelley and Magurran 2003) and crustaceans (Zulandt Schneider et al 2001). Even mollusks learn to recognize odors (Watanabe et al 1998). While common (perhaps universal) in mammals (Chrobak and Napier 1992) and birds (Clayton et al 2000), and conjectured for all vertebrates (Morris 2001), the functions of episodic memory, memories of events, seems beyond the reach of most invertebrates (Heinrich 1984). This suggests that PM in humans may be evolutionarily older than TEM in humans making it likely, though not at all certain, that they have different neural mechanisms. Since the contents of TEM consolidate into DM, which contains semantic memory, these facts suggest the possibility of separate mechanisms for PM and semantic memory.

One can also argue developmentally. Infants who have not yet developed object permanence (TEM) are quite able to recognize and categorize. This argues, though not conclusively, for distinct systems for PM and semantic memory. In this same direction Mandler distinguishes between perceptual and conceptual categorization in infants, the latter being based upon what objects do (2000) (see also Glenberg 1997). The IDA model would suggest PM involvement in perceptual categorization while semantic memory would play the more significant role in conceptual categorization. It also employs distinct mechanisms.

Yet another line of empirical evidence comes from experiments with amnesiacs such as H.M. Pattern priming effects involving recognition of letters, words, shapes and objects have been demonstrated that are comparable to unimpaired subjects (Gabrieli et al 1990). These studies suggest that H.M. can encode into PM (recognition) but not into DM (no memory of having seen the original patterns), including semantic memory. These results are consistent with, and even suggest, distinct mechanisms.

Our final, and quite similar, line of argument for distinct PM and semantic memory mechanisms come from studies of rats in a radial arm maze (Olton, Becker and Handelman 1979). With four arms baited and four not (with none restocked), normal rats learn to recognize which arms to search (PM) and remember which arms they have already fed in (TEM) so as not to search there a second time. Rats with hippocampal systems excised lose their TEM but retain PM, again suggesting distinct mechanisms.

In the recognition memory literature dual-process models have been put forward proposing that two distinct memory processes, referred to as familiarity and recollection, support recognition (Mandler 1980, Jacoby and Dallas 1981). Familiarity allows one to recognize the butcher in the subway acontextually as someone who is known but not to recollect the context of the butcher shop. In the IDA model, PM alone provides the mechanism for such a familiarity judgment, while both PM and DM are typically required for recollection. Recent brain imaging results from cognitive neuroscience support a dual-process model (Rugg and Yonelinas 2003), and so, are compatible with our Perceptual Memory Hypotheses. (For an analysis, see the section of Recognition Memory below.)

Table 2 lays out the correspondence between the different types of memory discussed, their implementation, or planned implementation, in IDA, and some hypotheses concerning their neural basis. Sensory memory, procedural memory, perceptual memory, working memory, and transient episodic memory are each implemented in the IDA model by a distinct mechanism, while the two forms of declarative memory use yet another (single) mechanism. These different mechanisms were chosen to facilitate the implementation of the distinct roles they play in cognition according to the IDA model. We hypothesize that *they each have distinct neural implementations in the brain*, though not necessarily in different anatomical structures.

There is some evidence from neuroscience in support of pieces of these hypotheses. As mentioned above, experiments with rats learning to search properly for food in an eight-armed maze support the role of the hippocampal system in transient episodic memory, but not in perceptual memory (Olton, Becker and Handelman 1979). An eye blink in response to a tone can be classically conditioned without appeal to the hippocampus if the unconditional response immediately follows the conditioned response. If the response is delayed, the hippocampus (transient episodic memory) is needed (Lavond, Kim and Thompson 1993). fMRI studies produced dissociation between possible neural correlates of PM and TEM. “Engagement of hippocampal and parahippocampal computations during learning correlated with later source recollection, but not with subsequent item recognition. In contrast, encoding activation in perirhinal cortex correlated with whether the studied item would be subsequently recognized, but failed to predict whether item recognition would be accompanied by source recollection.” (Davachi, Mitchell & Wagner 2003)

Memory type	Implementation in IDA	Neural Correlates
Sensory Memory	Perceptual Workspace	Temporal-prefrontal Neocortex (echoic) (Alain Woods & Knight 1998)
Procedural Memory	Codelets	
Perceptual memory	Slipnet	Perirhinal Cortex (Davachi, Mitchell & Wagner 2003)
Working Memory	Workspace	Temporo-parietal and frontal lobes (Baddeley 2001)
Transient Episodic Memory	Sparse Distributed Memory with “I don’t care” ⁶ and decay	Hippocampal System (Shastri 2001, 2002; Panksepp 1998. p 129)
Declarative Memory	Sparse Distributed Memory with “I don’t care”	Neo-cortex

Table 2. Memory Implementation

⁶ Kanerva’s Sparse Distributed Memory is modeled in binary space – both the address and content are binary in character. To accommodate for flexible cuing with fewer features than the actual memory trace, IDA’s TEM and DM are designed with a modified SDM that is set in ternary address space, but has a semi binary content-space: namely, 0, 1 and “don’t cares”.

Transient episodic and declarative memories have distributed representations in IDA. There is evidence that this is also the case in the nervous system. In IDA, these two memories are implemented computationally using a modified version of Kanerva's Sparse Distributed Memory (SDM) architecture (Kanerva 1988).

SDM models a random access memory that has an address space and a set of allowable addresses. The address space considered is enormous, of the order of 2^{1000} . Of this space, you choose a uniform random sample, say 2^{20} , of locations. These are called *hard locations*. Thus the hard locations are sparse in this address space. Many hard locations participate in storing and retrieving of any datum, resulting in the distributed nature of this architecture.

Each hard location is a bit vector of length 1000, storing data in 1000 counters with a range of -40 to 40. Each datum to be written to SDM is a bit vector of length 1000. Writing 1 to a counter results in incrementing the counter, while writing a 0 decrements the counter. To write in this memory architecture, you select an access sphere from the location X. Any hard location in this sphere is accessible from X. So, to write a datum to X, you simply write to all the hard locations (roughly about a 1000 of them) within X's access sphere. This results in distributed storage. This also provides naturally for memory rehearsal – a memory trace being rehearsed can be written many times and each time to about 1000 locations.

To read/retrieve from location Y, you compute the bit vector read at Y – by assigning its k^{th} bit the value 1 or 0, based on Y's k^{th} counter being positive or negative. Hence each bit of the bit vector read at Y is a majority rule decision of all the data that have been written at Y. Effectively, the read data at Y is an aggregate of all data that have been written to Y, but may not be any of them exactly. Similar to writing, retrieving from SDM involves the same concept of access sphere – you read all the hard locations within the access sphere of location Y, pool the bit vectors read from all these hard locations and let each of the k^{th} bit of those locations participate in a majority vote for the k^{th} bit of Y. Effectively, you reconstruct the memory trace in every retrieval operation.

This memory can be cued with noisy versions of the original memory trace. To accomplish this, you employ *iterated reading* – first read at X to obtain the bit vector, X_1 . Next read at X_1 to obtain the bit vector X_2 . Next read at X_2 to obtain the bit vector, X_3 . If this sequence of reads *converges* to Z, then Z is the result of iterated reading at X. *Convergence* happens very rapidly in this architecture, while *divergence* is indicated by an iterated read that bounces randomly in this address space.

The SDM architecture has several similarities with human memory (Kanerva 1988) and provides for “reconstructed memory” in its retrieval process.

IDA's Cognitive cycle and the Memory Hypotheses

IDA functions by means of flexible, serial but overlapping cycles of activity that we refer to as *cognitive cycles*. The overarching hypothesis, **the Cognitive Cycle Hypothesis**, that emerges from the IDA model claims that *human cognition functions by means of similar flexible cognitive cycles* each taking approximately 100-200 milliseconds, but because of overlapping, potentially occurring at a rate of more than five to ten cycles per second. Citing evidence from Thompson *et al.* (1996) and from Skarda and Freeman (1987), Cotterill speaks of “... the time usually envisioned for an elementary cognitive event ... about 200 ms (2003). From our cognitive cycle hypothesis

it might seem reasonable to call one such cycle an elementary cognitive event. Freeman (1999) suggests that conscious events succeed one another at a “frame rate” of 6 Hz to 10 Hz as would be expected from our cognitive cycle hypothesis (See also Freeman 2003b). The rate of such cycles coincides roughly with that of other, perhaps related, biological cycles such as saccades (Steinman, Kowler and Collewijn 1990), systematic motor tremors⁷, and vocal vibrato (Seashore 1967). Could these hypothesized cycles be related to hippocampal theta waves (at 6-9 hz) with gamma activity superimposed on them (VanRullen and Koch. 2003)?

The cognitive cycle is so named because it specifies the functional roles of memory, consciousness and decision-making in cognition according to global workspace theory. We'll next explore the cognitive cycle in detail (see Figure 2) to show the interactions between conscious events and memory systems.

Though in humans the cognitive cycle would most often begin with an expectation-driven action pointing sensors in a certain direction, here we will begin its description with incoming sensations. In Step 1 of the cycle, sensory stimuli, external or internal, are received and interpreted by IDA's (preconscious) perception module where recognition and categorization take place, and interpretations are assigned. Sensory memory facilitates this process by providing a workspace within which potential interpretation structures can be constructed and evaluated (Hofstadter and Mitchell 1994). In humans, sensory memory decay time is measured in milliseconds. The decision as to which interpretations (recognitions, categorizations, ideas, meanings) to assign is made by perceptual memory (PM) that's implemented in IDA by a slipnet, an activation passing semantic net. The slipnet settles into the various interpretations, represented by its nodes (Hofstadter and Mitchell 1994; Zhang et al 1998). *The existence of a separate, long-term, perceptual memory with its own distinct mechanism and updating procedure is a major construct of the IDA model.*

Though interpretations are assigned during each cycle, the full meaning of a particular stimulus, say a sentence, may accumulate only over several cycles. The activation of nodes in PM decays, but it does so at such a rate that interpretations acquired during one cycle are still somewhat available during the next few cycles (Tanaka and Sagi 1998). Thus we should think of this preconscious perceptual process as an iterative one.

In Step 2, the (still unconscious) percept, containing some of the data, plus its initial interpretations, is stored in one or more of the preconscious buffers of working memory (WM) (Baddeley 1993). In humans, these buffers may include the (preconscious) visuo-spatial sketchpad, the (preconscious) phonological loop, and possibly others. The decay rate of WM in humans is measured in seconds. Its capacity is typically taken to be quite small, though we hypothesize that *this limitation results from the limited capacity of consciousness rather than of the working memory buffers, the **WM Limitation Hypothesis***.⁸

⁷ “In the last 15 years or so, it has become clear that the 8-12Hz rhythmicity of physiological tremor is observed not only during voluntary movement, but also ... during maintained posture and in supported limbs at rest (Marsden et al. 1984).” (Llinas 2001)

⁸ Global workspace theory postulates the broadcast of conscious contents to a host of quite small, special purpose processors (Baars 1988) similar to [neuronal groups (Edelman 1987), agents (Minsky 1985), small minds (Ornstein 1986), demons (Jackson 1987)]. These limited capability processors would be ill equipped to handle a long or complex message.

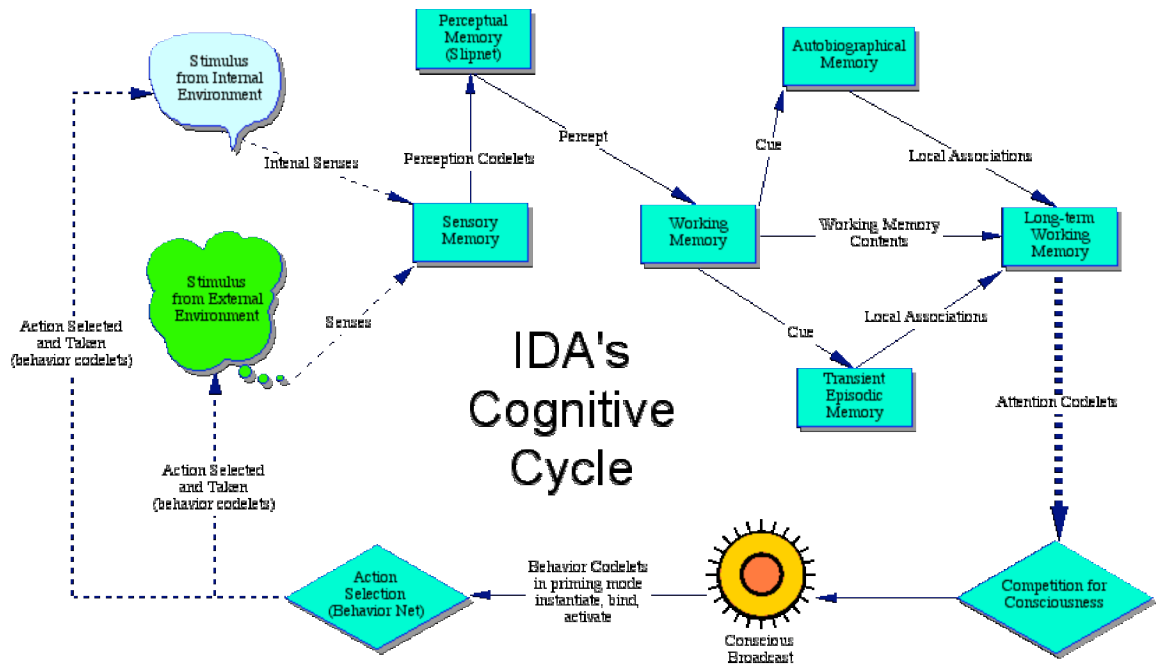


Figure 2. IDA's Cognitive Cycle

In Step 3 of the cycle the IDA model employs two different associative memories, each implemented, or being implemented, with a distinct, enhanced version of Kanerva's Sparse Distributed Memory (Kanerva 1988, Anwar et al. 1999, Anwar and Franklin 2003, Franklin 2001b), and subject to different decay rates. Declarative memory (DM), including both semantic memory and autobiographical memory, decays exceedingly slowly, if at all. Decades old memories of distinctive events are routinely retrieved (Bahrick, 1984). The capacity of this memory is very large. We speak of DM rather than of long-term memory in order to exclude both procedural memory and PM, because all three are implemented so differently in IDA. Their differences in function lead us to hypothesize that *they are implemented differently neurally as well*.

It is well established that in content-addressable associative memories interference results when similar specific events over time become merged into a general event, blurring their details (Chandler 1991, Lenhart and Freeman 1995, Lindsay and Read 1995). Thus, long-term associative memory cannot be counted on to help with the recall of where I parked my car in the parking garage this morning or what I had for lunch yesterday. These events are much too similar to a myriad of earlier such events. Yet such recall is essential for cognitive functioning. I need to know where to find my car. To solve this problem the IDA model makes use of a transient episodic memory (TEM) implemented as a content-addressable associative memory with a decay rate slower than that of WM but orders of magnitude faster than that of DM. For different empirical reasons, Conway postulates a similar TEM for humans with a decay rate measured in hours or perhaps a day and a sizable capacity (2001). Donald also assumes a quite similar TEM to which he refers as intermediate-term working memory (2001), while Panksepp speaks of a "transient memory store" (1998, page 129). *The existence of such a transient episodic memory in humans is a key hypothesis, the TEM Hypothesis, emerging from the*

IDA model. TEM is implemented in the IDA model with a version of Sparse Distributed Memory.

Using the incoming percept and the residual contents of the preconscious buffers as cues, during Step 3 local associations are retrieved from TEM and from DM. *These associations, together with the contents of the preconscious buffers of working memory, in humans would constitute Ericsson and Kintsch's long-term working memory (LTWM) (1995) (or Baddeley's episodic buffer (2000)), yielding another hypothesis from the IDA model. The IDA model predicts that LTWM uses working memory's preconscious buffers with their built-in rapid decay rates. Thus, in any cycle LTWM will contain both new local associations and those from previous cycles that haven't yet decayed. In the IDA model LTWM reuses the preconscious buffers of WM for its own storage and decays at the same rate as WM. We refer to these latter assertions for humans as the LTWM Hypothesis.*

In Step 4 of the cognitive cycle (above) we explicitly meet codelets for the first time.⁹ *Codelets* in IDA are small pieces of code running as independent threads, each of which is specialized for some relatively simple task. They often play the role of demons, waiting for a particular situation to occur in response to which they should act. Codelets in IDA implement the specialized processors of global workspace theory. Codelets also correspond more or less to Edelman's neuronal groups (1987), Jackson's demons (1987), Minsky's agents (1985), and Ornstein's small minds (1986). Codelets come in a number of varieties, as we will see. Perceptual codelets would have already participated in Step 1 above.

The task of attention codelets in IDA is to bring content to consciousness. In humans this content may include constructed sensory images, feelings, emotions, memories, intuitions, ideas, desires, goals, etc. Among these attention codelets are expectation codelets spawned by active behaviors (see below) that attempt to track the results of current actions or the lack thereof. In Step 4 of the cycle these attention codelets view LTWM. Motivated by novelty, unexpectedness, relevance to a goal, emotional content, etc., some of these attention codelets gather information codelets, form coalitions with them, and actively compete to bring their contents to consciousness (Bogner 1999; Bogner, Ramamurthy and Franklin 2000a; Franklin 2001b). Attention codelets are activated in proportion to the extent that their preferences are satisfied. This competition for consciousness may also include coalitions formed by attention codelets from a recently preceding cycle. The activation of such previously unsuccessful coalition members decays, making it more difficult, but not impossible, for their coalition to compete with newer arrivals. Thus we have a recency effect. The contents of unsuccessful coalitions remain in the preconscious buffers of LTWM until they decay away, and can take part in future competitions for consciousness.

IDA's mechanism for the competition for consciousness agrees quite nicely with Gray's neuroscience-based comparator model in which top-down and bottom-up processes interact to produce the contents of consciousness (1995, 2002). Attention codelets, and in particular the expectation codelets, constitute the top-down processes, while the information codelets from LTWM account for the bottom-up processes.

⁹ In Step 1 of the cognitive cycle, perceptual codelets scan the incoming stimuli (from the environment or internally generated) for items, or features, of interest to them. Successful perceptual codelets activate appropriate node in the slipnet, facilitating the assigning of interpretations.

Attention codelets are top-down in the sense that they are often instantiated by IDA's action selection mechanism, that is, by goal context hierarchies. Information codelets are bottom-up in that they carry information derived from incoming exogenous or endogenous stimuli or from some of the various forms of memory, WM, TEM, DM.

Within a single cognitive cycle the bringing of contents to consciousness is automatic. Existing attention codelets form coalitions and compete for access to the global workspace, i.e., to consciousness. All this is automatic. However, over multiple cycles the IDA model allows for conscious, voluntary attention. For example, while writing the phone rings. Shall I answer it or let the machine screen the call? A voluntary decision is taken to answer the call (Franklin 2000b), producing an intention to direct attention to a particular incoming stimulus. Each such voluntary decision spawns an attention codelet, in this case one that, likely, will bring the intention to consciousness, setting in motion the instantiation of a goal context hierarchy intent on answering the phone. (See Steps 6 and 7 below for details on how this happens.)

In Step 5 of the cognitive cycle a single coalition of codelets, typically composed of one or more attention codelets and their conveyer of relevant information codelets, wins the competition, gains access to the global workspace, and has its contents broadcast to all the codelets in the system. This broadcast is hypothesized to be a necessary condition for phenomenal consciousness. This "global access" broadcast step embodies the major hypothesis of global workspace theory, which has been supported by a number of recent neuroscience studies (Dehaene 2001; Dehaene and Naccache 2001; Baars 2002; Cooney and Gazzaniga 2003).

The IDA model leads us to hypothesize that *the conscious broadcast also results in the current contents of consciousness being stored in transient episodic memory*. The model also claims that at recurring times not part of a cognitive cycle, the undecayed contents of transient episodic memory are consolidated into long-term declarative memory. This suggests in humans a location of TEM in the hippocampal system (Shastri 2001, 2002).

In addition to broadcasting the contents of consciousness to the processors (codelets) in the system and storing them in TEM, the IDA model hypothesizes a third event occurring during Step 6 of the cognitive cycle. *Perceptual memory (the slipnet and its perceptual codelets in IDA) is updated using information from the contents of consciousness*. Details of this process will be discussed in the section entitled The Learning of Interpretations to be found below.

Also, the IDA model motivates us to hypothesize that *earlier actions are shaped by reinforcement derived from the current contents of consciousness*. This updating of procedural memory will be described in the section entitled Procedural Learning to be found below.

In summary, the IDA model suggests the following picture of the relationships between the various types of memory we have discussed (see Figure 3). A preconscious percept consisting of a selection of the contents of sensory memory, together with recognitions, categorizations and other interpretations produced in PM, are stored to WM. Only the conscious portion of the contents of WM (actually LTWM) is stored in TEM. Information from the same conscious content is used to update PM, TEM, and procedural memory. The undecayed contents of TEM are consolidated into DM at some other time. Retrieval from the content-addressable, associative memories TEM and DM uses recently

stored unconscious contents of WM as cues. A consequence of this account is that *contents can only be stored in DM via consciousness acting through TEM*. Another is that *new interpretations are learned in PM only via updating from conscious contents*. Yet another is that *skills are shaped only via conscious intervention*.

Thus a primary function of consciousness is to store relevant information in PM and in DM, and to shape skills in procedural memory. For this purpose the system employs a sequence of filters. The first filter selects from the internal and/or external sensory stimuli and the interpretations derived from them to produce the preconscious percept. This process involves sensory memory and PM¹⁰. The second filter uses this percept and the residual content of WM to select local associations from the vast store in DM and from the very much lesser store in TEM. The third filter selects the next contents of consciousness from among the various competing coalitions considering relevance, importance, urgency, insistence, etc.

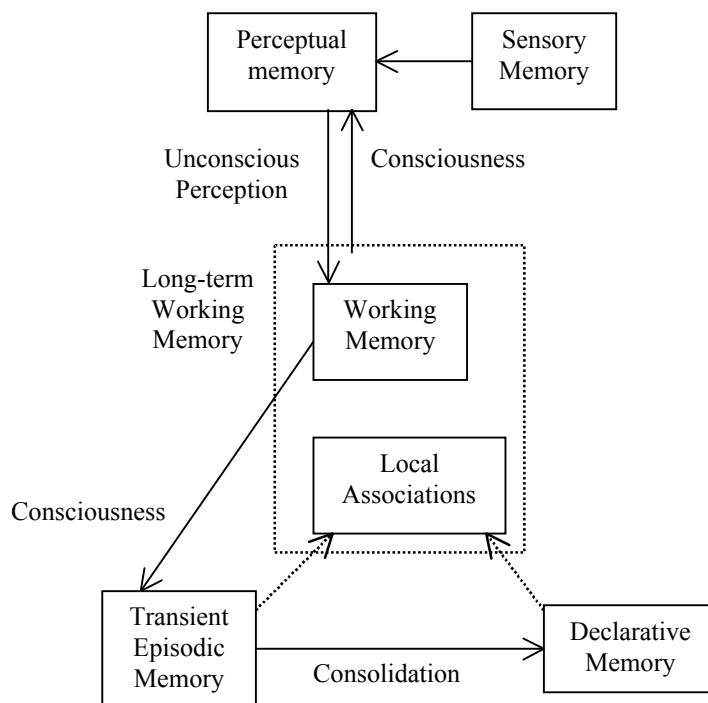


Figure 3. Relationships between some of the various memory systems

The remaining three steps of the cognitive cycle impact this account of what the IDA model has to say about memory by producing the actions that are shaped using conscious contents from subsequent cycles.

In Step 6 of the cycle, the recruiting function of consciousness, the central hypothesis of global workspace theory is implemented. Relevant behavior codelets respond to the conscious broadcast. These are typically codelets whose variables can be bound from information in the conscious broadcast. If the successful attention codelet(s) was an expectation codelet announcing an unexpected result from a previous action, the

¹⁰ In the computational IDA, PM is implemented by a semantic net with activations. The content of nodes with above threshold activations become part of the preconscious percept, thus the filtering.

responding codelets may be those some of whose actions can help to rectify the unfulfilled expectation.

In Step 7, some responding behavior codelets instantiate appropriate behavior streams (goal context hierarchies), if suitable ones are not already in place (Negatu & Franklin 2002). They also bind variables, and send activation to behaviors. Here we assume that there is such a behavior codelet and behavior stream. If not, then non-routine problem solving is called for, using additional mechanisms involving the execution over multiple cycles of a partial planning algorithm (Gerevin and Schuber 1996) implemented as a behavior stream. (This part of the IDA conceptual model is, as yet, unpublished.)

In Step 8, the behavior net (Negatu & Franklin 2002) chooses a single behavior and executes it. This choice may come from the just instantiated behavior stream or from a previously active stream. The choice is affected by internal motivation (activation from drives and goals as implemented by feelings and emotions), by the current situation, external and/or internal (environmental activation), by the relationship between the behaviors (the connections in the behavior net), and by the activations of the various behaviors. (The pervasive influence of feelings and emotions on the various memory systems and on action selection is another, as yet unpublished, part of the IDA conceptual model.)

Action is taken in Step 9, the final step of the cycle. The execution of a behavior results in the behavior's underlying behavior codelets performing their specialized tasks, which may have external or internal consequences. This part of the process is IDA taking an action. The acting codelets also include at least one expectation codelet (see Step 4) whose task it is to monitor the action taken, and to try and bring its results to consciousness, particularly any failure.

The Learning of Interpretations

Though semantic memory, a part of declarative memory (DM), and perceptual memory (PM) have quite similar content, their functions and processes are different. In the IDA model, PM is implemented with an activation-passing semantic net called a slipnet (Hofstadter and Mitchell 1994) while DM is implemented using an expanded version of sparse distributed memory (Kanerva 1988; Anwar, Dasgupta and Franklin 1999; Anwar and Franklin 2003). These distinct mechanisms were chosen to reflect their different functions. DM is accessed in Step 3 of the cognitive cycle to retrieve local associations with the current contents of the WM buffers. PM, on the other hand, is the critical element in the process of assigning interpretations to sensory input in Step 1. These quite distinct functions lead us to hypothesize *distinct neural mechanisms for DM and PM in humans*.

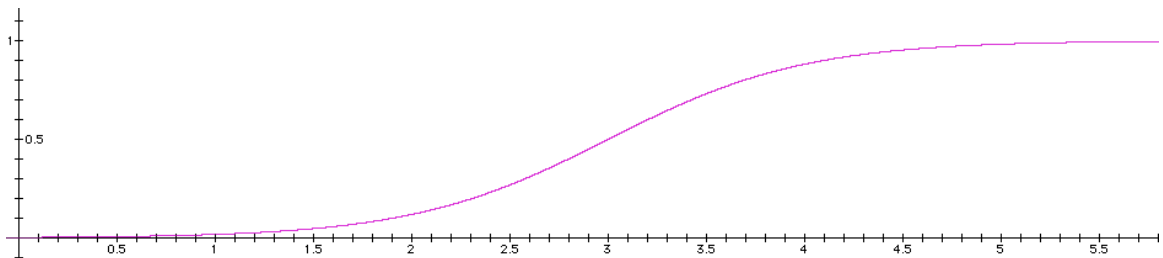


Figure 4. Base-level activation curve of slipnet nodes¹¹

PM is centrally involved in the processes of recognition, categorization, and concept formation. In the IDA model, recognition, categorization and the assignment of other interpretations happens in the slipnet. In Step 1 of the cognitive cycle, perceptual codelets search incoming stimuli, internally or externally generated, for features of interest to their special purpose. Those finding such features activate appropriate nodes in the slipnet. Activation in the slipnet spreads from node to node along relationship links. Eventually, the slipnet's state settles into a relatively stable pattern of activity. Nodes that are highly active at this point provide the interpretations assigned to the incoming stimuli. In Step 2 of the cognitive cycle, these interpretations (recognitions, categorizations, and others), together with some of the incoming data, are stored into WM buffers. The slipnet implements PM in the IDA model.

We've just described the process by which information is retrieved from PM in the IDA model. How is it stored (Gilbert, Sigman, and Crist. 2001)? To answer this question we need more detailed knowledge of the structure of the slipnet. Each node has a base-level activation analogous to the at-rest firing rate of a neuron. Additional activation provided by perceptual codelets, or spread from other nodes, is added to this base-level activation. The base-level activation of a slipnet node increases along a saturating, sigmoidal curve as the number of occurrences in consciousness of the node's interpretations increases (see Figure 4). The node's base-level activation decays at a rate inversely proportional to its magnitude. A low base-level activation decays quite rapidly, while a saturated base-level activation decays almost not at all (see Figure 5).

Links in the slipnet model relationships between nodes, concepts. Each link carries a weight that acts as a multiplier of activation spreading over the link. Similar to base-level activation of nodes, the weights on link increase along a saturating, sigmoidal curve according to the number of occurrences of its relationship in the contents of consciousness (see Figure 4). Also, the weight of the link decays inversely to its magnitude with a saturated weight decaying almost not at all (see Figure 5).

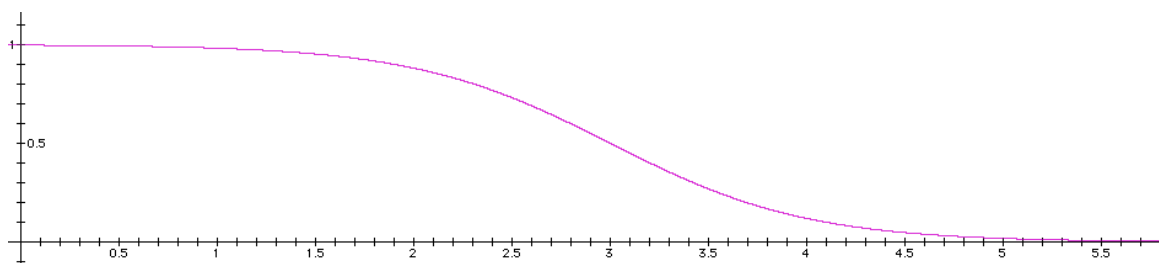


Figure 5. The decay curve of a slipnet node

Recall that Step 5 of each of IDA's cognitive cycles implements the broadcast postulated by global workspace theory. PM is updated simultaneously with the broadcast. The base-level activation of a slipnet node is increased whenever it represents an interpretation that appears in the broadcast contents of consciousness. Similarly, the weight on any slipnet link corresponding to a relationship appearing in the broadcast

¹¹ This and other such figures are included only for the general shape of the curve. The units are arbitrary.

contents of consciousness is increased. In this way, IDA's PM learns slowly with experience. With consciousness transpiring in cognitive cycles in humans occurring roughly five times a second, this incremental learning may not be so slow.

The previous paragraph describes the updating of existing nodes and links. But are such nodes and links originally learned and, if so, how? The general outline of how this is accomplished is as follows. An attention codelet succeeds in bringing to consciousness a new individual, a recent instance of a new category, or some other form of a new concept. The absence of an existing node for this concept triggers a mechanism that produces an appropriate new node within the slipnet. Each relationship of the new concept with another individual, category or other concept within the current contents of consciousness triggers the generation of a new link between their nodes in the slipnet.

A new node begins with a tiny base-level activation level; a new link begins with a tiny weight. Thus a new node or link is likely to decay away unless it receives reinforcement from the contents of consciousness during subsequent cognitive cycles. Should one, for example, be introduced to a new person at a party, shake hands, exchange a greeting and move on, the new node for the person just met would have come to consciousness only a dozen or so times, not producing much in the way of base-level activation. On the other hand, a ten-minute conversation with the new acquaintance might result in some three thousand appearances in subsequent conscious contents, resulting in a much higher base-level activation. In the first case, a sighting of the new person a few weeks later might result in no recognition at all. In the second case, recognition should occur and bring with it memories of the conversation.

One might reasonably ask at this point about the role of perceptual codelets in the learning of interpretations. Just as Edelman postulates a primary repertoire of neuronal groups (1987), we propose a primary repertoire of perceptual codelets evolved into natural systems and built into artificial systems. As do behavior codelets (see the next section), perceptual codelets have base-level activation that affect the total activation they can pass to appropriate nodes after noting their particular feature. Each conscious broadcast leads to the updating of these base-level activations just as happens for behavior codelets. In this way, degeneracy plays a role. Each broadcast also results in the appropriate associations between perceptual codelets and newly added nodes in perceptual memory.

One had best not think of IDA's slipnet implementation of PM as modeling neurons and their connections. A node would not correspond to a neuron. Nor do we think it would correspond to a collection of neurons. Perhaps it would be best to think of the slipnet as modeling, in the dynamical system sense, an attractor landscape with nodes corresponding to attractors and links to their boundaries (Skarda and Freeman 1987; Freeman and Skarda 1990; Freeman and Kozma. 2000). The well-known decades of difficulty computer scientists have had with pattern recognition leads us to believe that only a dynamical systems approach is likely to be successful in the real world.

Procedural Learning

Our account of procedural learning is based on consciousness providing reinforcement to actions. Global workspace theory has actions performed by special purpose processors. These are implemented in the IDA model by behavior codelets. We take them as neurally corresponding to neuronal groups (Edelman 1987). Edelman postulates an organism

being born with an initial repertoire of neuronal groups. An initial set of primitive actions, implemented by primitive codelets or some such, is a computational necessity for any autonomous agent, natural or artificial, software or robotic (Franklin 1997).

In the IDA model codelets have preconditions that must be satisfied in order that the codelet can act. They also have post-conditions that are expected after the actions take place. Each codelet also has a numerical activation that roughly measures its relevance and importance to the current situation. Part of this activation is base-level activation similar to that described for slipnet nodes above. Behavior codelets acquire environmental activation during Step 7 of each cognitive cycle (see above) in proportion to their preconditions being satisfied. In that same step, a behavior codelet may receive activation if its post-conditions help to satisfy some goal or drive as implemented by an emotion or feeling.

Collections of primitive behavior codelets form *behaviors*. This corresponds to collections of processors forming goal contexts in global workspace theory. Behaviors also have pre- and post-conditions derived from those of their underlying codelets. The activation of a behavior is always the sum of the activations of its codelets. Thus behaviors acquire activation for their codelets. They may also receive activation from other behaviors that they must share with their codelets.

Collections of behaviors, called *behavior streams*, with activation passing links between them, correspond to goal context hierarchies in global workspace theory. They can be thought of as partial plans of actions. Behavior streams are instantiated into an on going, activation passing, *behavior net*, also during Step 7. In Step 8 of each cycle, the behavior net chooses exactly one behavior to execute. The behavior net is IDA's action selection mechanism. More details of the operation of this action selection mechanism have been published elsewhere (Negatu and Franklin 2002).

Our account also employs Edelman's notion of degeneracy, postulating that several different neuronal groups can accomplish roughly the same task. In the IDA conceptual model the repertoire of behavior codelets is also taken to be degenerate. Thus two distinct behavior streams may be partial plans with the same goal, but may employ different behaviors to get there. Procedural learning in the IDA conceptual model is the consciously driven process of shaping a behavior stream via reinforcement.

According to the IDA model, procedural learning occurs according to the following scenario. In one cognitive cycle, a behavior is selected for execution by IDA's behavior net. The codelets that comprise that behavior all become active, including at least one expectation codelet concerned with bringing to consciousness the results of the actions taken by the behavior's behavior codelets. Typically, the results of some of these actions will be brought to consciousness by one of these expectation codelets during a subsequent cycle. During Step 5, the broadcast step, of that cycle, the base-level activations of the responsible behavior codelets are adjusted upward or downward by an internally generated reinforcement signal, according as to whether the action was successful or not as determined by the expectation codelet. The adjustment is made using a curve such as that in Figure 6 using the following procedure. Find the point on the curve whose y-value is the current base-level activation of the behavior codelet. According as the action succeeded or failed, extend a horizontal line segment of unit length to the right or left of the point. (The unit is an adjustable parameter.) A vertical

line through the endpoint of this segment intersects the curve at a point whose y-value is the new base-level activation.

Following this reinforcement procedure, successful behavior codelets with low base-level activation increase their activation rapidly. They learn quickly. Unsuccessful behavior codelets with low base-level activation decrease their activation even more rapidly. This learning on the part of a behavior codelet is done in the context of its current behavior (goal context) and behavior stream (goal context hierarchy). The change in the behavior codelet's base-level activation can be expected to change the likelihood of its behavior stream being instantiated in a future similar situation, and to change the likelihood of its behavior being executed. Learning in context will have occurred. For example, in the case of an unsuccessful behavior codelet, a similar behavior stream with a behavior like the original but with the unsuccessful codelet replaced with another with similar results may be chosen the next time. Degeneracy is required here.

The learning procedure just described is part of the IDA conceptual model that has not yet been implemented computationally.

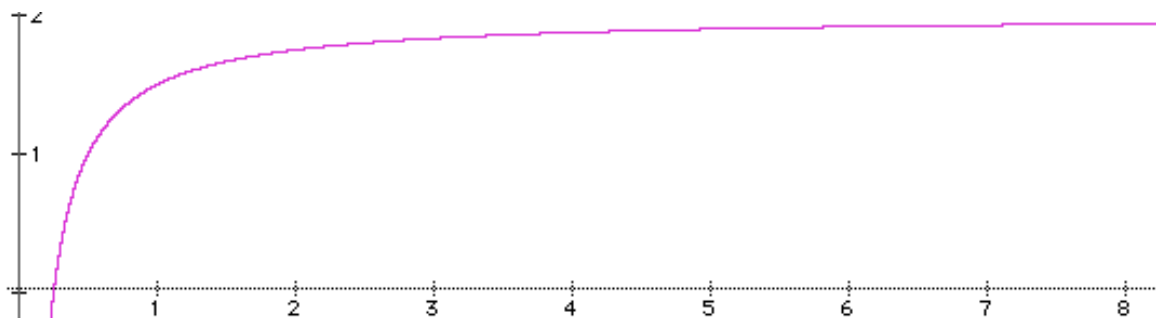


Figure 6. Procedural Learning Curve

The Availability Heuristic

It is well known that people tend to overestimate the frequency of divorce if they can quickly recall instances of divorced acquaintances. This principle also applies to frequency estimates, and is referred to as the *availability heuristic* (Kahneman, Slovic and Tversky 1982; Fiske and Taylor 1991). An online demonstration of the heuristic (Colston and Walter. 2001) asks the subject to review a list of names of well-known people, one such presented at each mouse click, to see if the subject knows them. No mention is made of gender. After viewing the last name in the list, the subject is presented a forced choice as to whether he or she had seen more men's names or women's. Since the men named tended to be more famous, and hence more easily recalled, the availability heuristic would correctly predict that most subjects would claim that there were more men's names on the list. There is, in fact, fourteen of each. In this section we will analyze this task using IDA's cognitive cycle to see what the IDA model would predict.

The initial instructions comprise a text of some thirty-seven words. To read and understand the instructions will likely occupy a few seconds and some few tens of cognitive cycles. During the last of these cycles, the gist of the meaning of the instructions will have accumulated in the appropriate preconscious working memory buffer. The conscious broadcast of these meanings will likely instantiate a goal context

hierarchy for sequentially clicking through and seeing the names, noting whether they are recognized. The action chosen during this cycle will likely be a mouse click bringing up the first name.

One to three or four cycles will likely suffice to preconsciously perceive the entire name, which will have accumulated in the appropriate preconscious working memory buffer. Here we must consider two cases: 1) The name has been recognized during the preconscious perceptual process (see the description of perceptual memory above), and 2) it has not. Recognition of the name John Doe has occurred if the subject can answer the question “who is John Doe?” In the first case, after the conscious broadcast the behavior net will likely chose a mouse click as an action. This would bring up another name. In the second case a conscious goal to consult declarative memory in search of recognition would likely arise over the several cycles (see major hypothesis 7 on voluntary memory retrieval above.) Such a conscious goal would produce an attention codelet on the lookout for information concerning the as yet unrecognized name. In Step 3 of the next cycle, the name in a preconscious working memory buffer will be used to cue declarative memory. This voluntary memory retrieval process may iterate over several additional cycles with the parts of local associations that make it to consciousness contributing to the cue for the next local associations. The subject will eventually recognize the name or will give up the effort on a subsequent cycle. In either case, the action chosen on this last cycle is likely to be a mouse click for the next name.

Thus the subject will work his or her way through the list of names recognizing many or most of them but missing some. At the mouse click following the last name a new set of instructions appears. These instructions, in sixteen words, ask the subject to decide whether more men’s or women’s names were on the list, and to click the mouse when the decision is made. A very few tens of cycles are spent understanding the instructions.

The conscious broadcast whose contents contain the full understanding of the instructions will likely recruit behavior codelets that instantiate a goal context hierarchy to comply. During Step 8 of some subsequent cycles, behaviors (goal contexts) will likely be selected that will attempt to query transient episodic memory (TEM), starting the recall process for names on the list. The behavior’s codelets will write this goal to preconscious buffers of working memory (WM) where it will serve to cue local associations from TEM and declarative memory (DM). The next contents of consciousness will be chosen from the resulting long-term working memory (LTWM).

This process can be expected to continue over many of the next number of cycles with each cue from WM containing material from previous local associations from DM. Since the names were encoded in TEM as distinct events, those retrieved from TEM can be expected to appear as such in LTWM when associated with the latest cue. From there each recovered name is likely to appear as the central content of its own coalition of attention and information codelets competing for consciousness. Thus, the IDA model would predict that only one name would be recalled at a time, since a single coalition must win the competition. More famous names would have their initial coalitions replaced during subsequent cycles with expanded coalitions including information from their many local associations in DM. Thus they would accrue an advantage in the next competition for consciousness; more famous names are more likely to be recalled. A name on the list could fail to be recalled for either of two reasons. One, it might not have been retrieved from TEM by any of the cues used during the process. Or, it may have

been retrieved into LTWM, but decayed away before it could be part of a coalition that won the competition for consciousness.

At some point in this process a decision as to more men's names or more women's is taken. How does that happen? Some subjects may actually instantiate a goal context hierarchy (behavior stream) that keeps separate tallies in WM of the number of men's names and of women's names recalled. These tallies are likely to be part of the conscious contents in at least some of the ongoing cycles in the process. When no more names are being recalled, a goal context (behavior) to decide is chosen. The decision is then made and, on a subsequent cycle, the mouse is clicked.

We think this is a relatively unlikely scenario in that most subjects will not keep such explicit running tallies. Rather, they'll decide on the basis of a fringe consciousness feeling that one gender or the other has been recalled more often (Mangan 2001). In the IDA model, such feelings are to be implemented as fringe attention codelets. In this case, a selected behavior would give rise to two such fringe attention codelets, one for each gender. As a name from one gender is recalled, the activation of the corresponding fringe attention codelet is increased. Each of these fringe attention codelets likely enters each competition for consciousness as the process progresses. The stronger of them will be able to win only after names are no longer being recalled; a coalition with a name to be newly recalled would simply have a higher average activation. Such fringe consciousness feelings are easily defeated.

Thus the IDA model predicts the experienced outcome of this demonstration of the availability heuristic, and supports the commonly given, functional explanation. What the model adds is a hypothesized detailed mechanism for its functional process.

Voluntary action in process dissociation

Our example of an analysis of a process dissociation experiment using the IDA cycle will require some understanding of how the IDA model handles voluntary action selection (Franklin 2000b; Kondadadi and Franklin. 2001). In this section, we provide a brief summary of that process.

We humans most often select actions subconsciously, that is, without conscious thought. But we also make voluntary choices of action. Baars argues that such voluntary choice is the same as conscious choice (1997, p. 131). We must carefully distinguish between being conscious of the results of an action and consciously deciding to take that action, that is, of consciously deliberating on the decision. It's the latter case that constitutes voluntary action. William James proposed the *ideomotor theory* of voluntary action (James 1890). James suggests that any idea (internal proposal) for an action that comes to mind (to consciousness) is acted upon unless it provokes some opposing idea or some counter proposal. GW theory adopts James' ideomotor theory as is (Baars 1988, Chapter 7) and provides a functional architecture for it. The IDA model furnishes an underlying mechanism that implements that theory of volition and its architecture in a software agent.

Suppose that in the experiment of the previous section, the subject has reached the stage of deciding upon a more frequent gender. The players in this decision making process include proposing and objecting attention codelets and a timekeeper codelet. A proposing attention codelet's task is to propose a certain decision, in this case a gender. Choosing a decision to propose on the basis of the codelet's particular pattern of

preferences (in this case for one gender), it brings information about itself and the proposed decision to “consciousness” so that the timekeeper codelet and others can know of it. Its preference pattern may include several different issues with differing weights assigned to each, though that doesn’t happen in our example. If no other objecting attention codelet objects (by bringing itself to “consciousness” with an objecting message), and no other proposing attention codelet proposes a different job within a given span of time, the timekeeper codelet will decide on the proposed decision. If an objection or a new proposal is made in a timely fashion, it stops timing or resets the timing for the new proposal.

Two proposing attention codelets may alternatively propose the same two jobs several times. Several mechanisms tend to prevent continuing oscillation. Each time a codelet proposes the same job it does so with less activation and, so, has less chance of coming to “consciousness.” Also, the timekeeper loses patience as the process continues, thereby diminishing the time span required for a decision.

Analysis of a Process Dissociation Experiment

The IDA model also serves as a tool with which to analyze and qualitatively predict and/or explain the results of psychological experiments. Here, we offer as an example an analysis of a well-known process dissociation experiment designed to measure conscious vs. unconscious memory (Jacoby 1991). After having been given a list of twenty or so two syllable words to study one at a time for a few seconds each, subjects were tested under an inclusion and under an exclusion condition. When given the first syllable of a word, a subject was asked under the inclusion condition to complete the word with one on the list, if possible, and to guess otherwise. Under the exclusion condition the subject was to complete the word while avoiding those words previously seen on the list. Errors during the exclusion condition illustrated unconscious (implicit) memory, i.e. the unconscious influence of previously seen, but not consciously remembered, words. Such errors were made at above chance levels. Can the IDA model account for these results?

Our analysis begins with the subject being told that he or she will be shown word stems and given a forced choice to produce a complete, two-syllable word based on that stem under either the inclusion or exclusion condition. For example, suppose the word *mo-tive* had appeared in the studied list of words. The corresponding word stem would be *mot---*. We presume that the subject intends to comply with the choice. Suppose the instruction prescribes the exclusion condition, asking that the subject complete the word-stem avoiding words previously studied in the list. At the end of the few tens of cycles required to understand the instructions, a behavior stream (goal context hierarchy) for the required voluntary decision-making will be instantiated (Franklin 2000b; Kondadadi and Franklin 2001) with its various behaviors (goal contexts) activated and their variables bound (see previous section). Attention codelets for bringing word completions to consciousness (proposers) will also be instantiated, as well as at least one attention codelet whose role it is to object to completion proposals for words on the studied list. The voluntary decision making process will have started. The action taken during this cycle would include attending to an upcoming word-stem. Over the next several cognitive cycles, a word-stem completion decision will be taken.

When a stem, say *mot---*, is presented, perceptual codelets note its features, e.g., letters and their order, and activate appropriate nodes in the slipnet (PM). The incoming

stem stimulus may be assigned meaning in the form of preconscious recognition if an individual word-stem node achieves sufficient activation. If not, it will likely still be preconsciously recognized as a word stem (belonging to the class of word stems) by virtue of a more abstract WORD-STEM node being activated. This assignment would occur during Step 1 of a single cognitive cycle. In Step 2, this preconscious percept consisting of an image of the stimulus work-stem together with whatever meaning it has acquired would be written to the preconscious buffers of WM. There, it would cue local associations in TEM and DM during Step 3. The local associations from TEM would likely include any recently encountered word that constituted a completion of the stimulus word-stem, possibly *motive*, together with, possibly, the episodic event in which it was encountered, say being on the studied list if that were the case. Local associations from DM would likely include various completions of the stimulus word-stem recovered from semantic memory, say *motel*, *motion*, *motor*, *motive*, etc.

With *motive* on the studied list and *mot---* as the presented word-stem we must now consider several possibilities:

1. The local associations from TEM include the word *motive* and its relation to the studied list.
2. The local associations from TEM include *motive* but NOT its relation to the studied list.
3. The local associations from TEM do not include *motive*.

Each of these cases may be expected to differently affect results in the word-stem completion experiment. We will continue our analysis in each of the three cases, and also consider how these three different local associations can come about. In all three cases, the instructions are likely to remain available in LTWM.

Case 1. In this case, TEM has done its expected job and the word *motive* has been explicitly but as yet unconsciously recalled into LTWM, along with the event of it having been on the list of studied words. During Step 4, attention codelets observe LTWM looking for opportunities to form coalitions and compete for consciousness. Among these attention codelets are likely various proposer attention codelets, who would like to bring particular word-stem completions to consciousness, thereby proposing them. A coalition formed around one of these would likely win the competition, and the “idea” of a particular completion would “pop into mind” during Step 5. In this case, a coalition built around the stem completion *motive* would likely win the competition due to the recency effect, and would “pop into mind” during the conscious broadcast of Step 5 as a proposed word completion of *mot---*. In this situation, Steps 6 and 7 would not likely add additional behavior streams (goal context hierarchies) to the behavior net (Negatu and Franklin 2002). The action selected in Step 8 and taken in Step 9 would surely include waiting for a possible objection.

Step 1 of the next cycle would then encounter the continued presentation of the presented word-stem, *mot---* as an exogenous stimulus. In Step 2, this image, along with its assigned meaning would be written as a preconscious percept to the preconscious WM buffers, where they would help cue TEM and DM for local associations during Step 3. These associations would likely leave LTWM only slightly changed from the previous cycle and still containing the given instructions, the word *motive*, its association with the list the subject studied, and the word *motive* having been proposed as the desired completion. This situation is precisely what an objector attention codelet has been

waiting for. It gathers a coalition carrying the idea *not motive, it's on the list* and wins the competition for consciousness in Step 4. The conscious broadcast in Step 5 brings an inner voice objecting to the choice of *motive*. The results of Steps 6 and 7 bind variables and activate behaviors, but leave the dominant goal context hierarchy (active behavior stream) unchanged. The action selected in Step 8, when taken in Step 9, stops and resets the unconscious timer.

During the next cycle, several of the alternate completion possibilities from Step 3's local associations from DM become the nuclei of coalitions brought together by proposer attention codelets. One of these coalitions, say that carrying *motel*, wins Step 4's competition and is brought to consciousness during Step 5's broadcast, and is thus proposed as the desired completion. The activations of these competing coalitions may depend indirectly on the frequency of the word they carry due to its affect on their storage in DM.

On a subsequent cycle, sufficient time will have passed without objection for the proposal of the word completion *motel* to be accepted during Step 8 when an action is initiated to speak the word *motel*. Note that the subject only becomes conscious of the choice having been made when the results of the action taken in Step 9 come to consciousness in Step 5 of the next cycle. We are conscious of the ideas that "pop into mind" during voluntary action selection, and also of the internal debates about them, but not of the actual taking of the voluntary decision.

This completes the analysis of Case 1.

Case 2. In this case, the word content *motive* has been retrieved as a local association from TEM during Step 3 of a cognitive cycle, but its source content, it having been among the words studied, has not. How can such a case occur? The transient episodic memory of the general event of studying a list of words is reinforced by coming to consciousness several times during each of the twenty or so five-second or so periods of study of individual words. Having come to consciousness a thousand or more times during a short time period, it would have been well established in TEM and would, no doubt, be eventually consolidated into DM. In contrast, each individual specific event of a word studied would tend to lose its detail due to the similarity of the related stored events. Among the details lost or blurred in some of these events might well be the connection with individual words. This blurring would occur in the Sparse Distributed Memory (Kanerva 1988) used in the IDA model to implement TEM by virtue of its distributed mode of storage. It's more of a matter of retrieval being degraded rather than of decay of the storage, which should affect both the word content and the source content equally.

Our analysis in this case would parallel that of Case 1 up to the point that the objector attention codelet won the competition for consciousness in Step 4. In this case, the situation would be different for the objector since the word *motive* having been among the words studied wasn't recalled. Thus no objection would come to consciousness and, after an appropriate wait, *motive* would be selected as the word completion. Hence an exclusion error would have occurred, an example of the influence of unconscious (implicit) memory in operation.

Case 3. If the word *motive* was not recalled, no such exclusion error would result unless *motive* was selected from among the local association from DM. In such a case, the competition would likely depend largely on the frequency of the words in question.

But how can motive not be recalled from TEM when it had been recently studied? Due to the combination of the primacy effect and the recency effect, the recall curve upon studying words in such a list would be u-shaped. That is, words early or late in the list would be more likely to be recalled than those in the middle. Why is this? According to the IDA model, words early in the list have the advantage of close and persistent attention being paid to them with this attention diminishing as the subject becomes habituated to the experiment. Early words have the disadvantage of a longer time in which to decay. Words late in the list suffer from lack of attention, but have the advantage of less time for decay. Words in the middle of the list tend to suffer both disadvantages. The result is a U-shaped recall curve allowing for a studied word like *motive* not to be recalled.

Retrieved memory information is less conscious in recognition than in recall tasks

As mentioned above, analysis using the IDA model can offer an explanatory account of the cognitive processes involved in the recognition of a person in an unfamiliar context, say a butcher encountered in the subway. Such an encounter can be expected to produce one of three different responses:

1. A recollection of the butcher including the context of the butcher shop.
2. A feeling of definite familiarity (“I know this person. Now, who is it?”)
3. A feeling of possible familiarity (“Don’t I know this person?”).

Though the feeling of familiarity may be more of a continuum than cases 2 and 3 might suggest, subjects in experiments are generally asked to choose between them.

Such memory situations have been studied experimentally giving rise to the dual-process proposal of two distinct memory processes being involved (Mandler 1980, Jacoby and Dallas 1981, Rugg and Yonelinas 2003). Subjects are often asked whether they recognize a target word as having been recently seen. A reliability judgment often accompanies the answer to measure the “continuum” of familiarity. A positive response is followed by a query as to recollection of the context in which the word was seen.

Let’s first consider *case 1* of recollection. Suppose a sensory stimulus of the butcher’s face. During Step 1 of the IDA cognitive cycle, it would generate activation (via perceptual codelets responding to features) in the Perception Memory (PM) node for that individual’s face. (Were this to occur in the butcher shop instead of in the subway, activation from other sensed nodes would also be passed to the face node in question.) The total activation of the individual’s face node will depend on its strength (base activation), the activation it receives from the perceptual codelets noting features, and the activation passed by other nodes over relation links.

If the total activation is sufficiently strong, the face image, and its activation, are passed in Step 2 to the appropriate working memory (WM) preconscious buffer along with its total activation. If the total activation doesn’t reach this threshold, none of the three responses listed above can occur. If sufficient activation passes from the individual face node to the butcher shop node as in this case, the latter relation may also be sent to working memory. In this case, recognition will likely occur¹².

In Step 3, local associations are sought from transient episodic memory (TEM) and from long-term declarative memory (DM) cued by the contents of these preconscious working memory buffers. In this case, the butcher shop node made it to working memory,

¹² It could fail in the unlikely case of a simultaneous distraction, say a loud explosive sound that would compete for consciousness in Step 4 of the cycle and win.

the local association may well contain information about the most recent visit to the shop and other more general, semantic information about it. Attention codelets looking at long-term working memory (LTWM), would in Step 4, gather coalitions of information codelets and compete for consciousness. The local associations could be reported during subsequent cycles to an experimenter checking for recollection, provided they make it to consciousness.

If the face of the butcher together with an image of the butcher shop comes to consciousness in Step 5, we would say that recognition had occurred via the recollection process. In the IDA model, this process would have employed PM, WM, TEM or DM or both, LTWM, and the consciousness mechanism, a considerable part of the cognitive apparatus.

For *case 2*, let's suppose that the butcher shop node was not sufficiently activated for an image to make it to WM in Step 2. Let's also suppose that the node for the butchers face in PM was strongly activated so that the image, together with its *high* activation, was sent to WM as part of the preconscious percept. In this case, the local associations of Step 3 may contain nothing of value to the recognition of the butcher. However, an attention codelet may well succeed in Step 4 in bringing the image to consciousness in Step 5, along with the definite feeling of knowing that person based on familiarity. That is, response 2 above may occur.

We would tend to classify response 3 above, a vague sense of familiarity, as part of fringe consciousness (Mangan 2001). In this case, *case3*, the face image together with its relatively *low* activation would have come to working memory in Step 2, and thus to long-term working memory during Step 4. (As in case 2, the local associations of Step 3 would play little role.) Response 1 would occur if a "sense of familiarity" fringe attention codelet were attracted by the face image and its low activation level, and were to win the competition for consciousness. In this case the subject might produce the internal sentence "Don't I know this person?" over the next several cycles.

Cases 2 and 3 seem to be instances of the familiarity process of recognition. This process uses the same cognitive mechanisms as does the recollection process, but uses all of them less heavily.

Case 2 may often lead to an attempt at voluntary memory retrieval over the next several cognitive cycles (see Hypothesis 8 above and the section on Voluntary Action). With the conscious goal of recognizing the given face, various structures may be employed in hopes of successfully cueing TEM and/or DM for local associations that would identify the butcher. A conscious goal in IDA that cannot be immediately accomplished generates an intention codelet, a particular kind of attention codelet that watches LTWM so as to bring to consciousness material that can help achieve its goal. It is the action of this persistent codelet that causes the recognition of the butcher to occur some time after case 2.

This analysis supports the dual-process model of recognition memory by explicating the empirically observed characteristics of the two processes, familiarity and recollection. Although utilizing the same underlying mechanisms, the two processes do so differently. Recollection is relatively slow (since it requires several cycles), and yields qualitative information about prior events (from local associations). Familiarity, on the other hand, offers a simpler "strength-like" memory signal (deriving from the activation level of the butcher's face node in PM).

Based on neuropsychological, ERP and fMRI studies of recognition memory, Rugg and Yonelinas (2003) conclude that “recognition is supported by functionally dissociable processes...” Our analysis supports and explicates that conclusion. They also suggest that the two processes “depend upon qualitatively distinct neural mechanisms.” Our analysis suggests considerable common mechanisms for the two processes (PM, WM), along with some additional mechanisms (TEM, DM) for recollection. They also claim that their “results strongly support the view that the hippocampus plays a crucial role in recollection, and argue for a less important role for this structure in familiarity.” If one views the hippocampus as the seat of TEM (Shastri 2001, 2002; Panksepp 1998, p 129) and the gateway to DM (Eichenbaum 2002, p. 137), then our analysis also supports this conclusion.

Dream Amnesia

Introspection suggests that dreaming is a conscious activity, and that idea is supported by the waking-type EEG patterns that typically accompany REM sleep. Yet perhaps all of us have experienced awakening from a dream and only remembering its very last portion, if any of it at all. This phenomenon is referred to as dream amnesia. Neuroscientist Paul Adams theorizes that both slow wave sleep and REM sleep are required for the offline updating of thalamocortical circuits during learning (consolidation of DM) (Cox & Adams 2000). Updating one of the circuits requires activity provided by the random auto-associations of dreaming, while the other needs the relative quiet of deep sleep. Thus the imaginary contents of dreams would play no useful role in DM; they are not to be learned. The IDA model suggests that during sleep, just as motor activity is suppressed, so also is the function detailed in Step 5 whereby conscious contents are stored in TEM. Hobson and Stickgold have suggested a neural mechanism for this phenomenon in terms of cholinergic activity during REM sleep (1994). Thus the only memory available for the recall of a dream is the small capacity WM, resulting in dream amnesia. The limited capacity of WM would yield a memory only of the final small portion of the dream. The rapid decay of WM would account for no memory of the dream at all after a slow awakening.

The “Unconscious” Driving Experience: Recall of absorbed automatic skills

The belief that one can drive an automobile unconsciously is widely held. A quick web search turned up a number of quotes attesting to this belief. Here are a few such. “Have you ever been driving down a highway and all of a sudden snap out of a daydream only to briefly panic as you realize you've been driving unconsciously for quite awhile?” “The laughter faded with the memory as he realized he had been driving unconsciously.” “When you drive a car for long periods you tend to slip into this bicameral state and are actually driving unconsciously.” Philosopher Peter Carruthers offers “absent-minded driving” as one example of a “non-conscious experience” (1998).

Not everyone adheres to this belief. Dennett suggests describing the unconscious driving experience as “a case of rolling consciousness with swift memory loss” (1991, p. 137). Commenting on the Carruthers paper mentioned above, Lyvers writes: “In the case of ‘absent-minded driving’, many of the relatively mundane events one routinely encounters in everyday driving experiences (such as a particular traffic light being red or green) are simply not encoded by the memory processes into a readily accessible or

enduring form, most likely because such events do not have any particular long-term significance. The absence of enduring memory for such events can lead to the illusory retrospective impression that one has somehow been driving ‘unconsciously’ (1999).”

The IDA model suggests that Dennett and Lyvers are correct in thinking that the driving was done consciously, but that the details were not recalled. According to the model, sensory input during driving was (unconsciously) perceived, stored in preconscious WM, brought to consciousness, and stored in TEM (Steps 1 through 5 of the cognitive cycle). If the driving experiences were so stored, why can’t they be recalled? TEM, a content-addressable associative memory, blurs details of similar events. In driving to work, for example, one residential stretch is quite similar to another. Intersections are similar to one another, as are commercial stretches. Also, similar events from the previous morning may not have yet decayed from TEM. If the driving was time shared with something else, say listening to NPR on the radio or talking to a passenger, fewer of the details of the driving experience would have become conscious (Step 4) and been stored in TEM. Thus it’s not surprising that on arriving at work the driving experiences can’t be recalled.

Conclusion

In the introductory section above we listed seven empirical facts about the role of consciousness in memory that required accounting for. The IDA model can account for almost all of these. In this concluding section, we guide the reader to these accountings in the current article and in a couple of others.

1. **Subliminal vs. supraliminal learning.** Although subliminal acquisition of information appears to occur, the effect sizes are quite small compared to conscious learning. Conscious access greatly facilitates most types of learning. The IDA model’s qualitative account of several of these types of supraliminal learning appears in the sections above entitled ‘IDA’s Cognitive cycle and the Memory Hypotheses,’ ‘The Learning of Interpretations’ and ‘Procedural Learning.’
2. **Process dissociation.** In a section above entitled ‘Analysis of a Process Dissociation Experiment,’ we provide a careful, fine-grained analysis of a well-known process dissociation experiment designed to measure conscious vs. unconscious memory (Jacoby 1991). The IDA model accounts qualitatively for errors observed during the exclusion condition that illustrates unconscious (implicit) memory, i.e. the unconscious influence of previously seen, but not consciously remembered, words.
3. **Implicit learning.** Aside from noting that consciousness plays a central role in implicit learning by virtue of the stimulus set always being conscious, the IDA model is as yet unable to adequately account for implicit learning. We’re working on it.
4. **Recall vs. recognition.** That people are conscious of retrieved memories in recall, but not necessarily in recognition tasks (e.g., Gardiner et al, 1998) is accounted for by the IDA model in a section above entitled ‘Retrieved memory information is less conscious in recognition than in recall tasks.’ The role of transient episodic memory and/or declarative memory in recall is contrasted with that of the fringe consciousness observed during recognition.

The fine-grained analysis of a typical recall vs. recognition experiment provides a qualitative account of the evidence.

5. **Working memory.** A companion article by Baars & Franklin (2003) describes the way the IDA model accounts for the active operations of working memory, -- namely input, rehearsal, recall and report, being conscious (Baddeley 1993). In contrast, contents of working memory prior to retrieval are not conscious.
6. **Availability heuristic.** In problem solving, people tend to select the first solution that becomes conscious. This heuristic can lead to predictable patterns of errors (Kahneman, Slovic and Tversky 1982; Fiske and Taylor 1991). In a section above entitled 'The Availability Heuristic,' a fine-grained, careful analysis allows the IDA model to account for these patterns qualitatively.
7. **Automaticity.** Predictable skills tend to become more automatic, that is less conscious, with practice. When such automatized procedures repeatedly fail to accomplish their objectives, conscious access tends to return. The IDA model's account of these processes is the subject of a separate article (Negatu, McCauley and Franklin in preparation).

The above analyses demonstrate the usefulness of the fine-grained cognitive cycle of the IDA model as an aid to thinking about cognitive processes, as well as a source of, hopefully testable, hypotheses. (See Franklin & Graesser 2001; Baars and Franklin 2003 for more such hypotheses.) But, as any computational/conceptual model of human cognition must be, the IDA model is replete with gaps, areas in which it can't offer explanations. With no hope of filling them all, we intend to fill an occasional gap as time and energy permit.

Acknowledgments

The form of presentation of the cognitive cycle used above was adapted from a paper about the interaction of consciousness and WM (Baars and Franklin 2003). We are grateful to Art Graesser, Danielle McNamara, Lee McCauley, Bill Marks, Aregahagn Negatu, Sidney D'Mello, and Igor Beliaev for most useful conversations on the topics of this article. B.J.B gratefully acknowledges support from the Neurosciences Institute and the Neurosciences Research Foundation (10640 John Jay Hopkins Drive, San Diego, CA 94549, USA; www.nsi.edu).

References

- Alain, C., D. L. Woods, and R. T. Knight. 1998. A distributed cortical network for auditory sensory memory in humans. *Brain Res.* 812:23-37.
- Anderson, J. R. 1990. *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum
- Anwar, A., D. Dasgupta, and S. Franklin. 1999. Using Genetic Algorithms for Sparse Distributed Memory Initialization. International Conference Genetic and Evolutionary Computation(GECCO). July, 1999.
- Anwar, A., and S. Franklin. 2003. Sparse Distributed Memory for "Conscious" Software Agents. *Cognitive Systems Research* 4:339-354.

- Baars, B. J. 1983. Conscious contents provide the nervous system with coherent, global information." in R.J. Davidson, G.E. Schwartz & D. Shapiro, *Consciousness & Self-regulation*. P. 41 NY: Plenum Press.
- Baars, B. J. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Baars, B. J. 1993. How does a serial, integrated and very limited stream of consciousness emerge from a nervous system that is mostly unconscious, distributed, and of enormous capacity? In *CIBA Symposium on Experimental and Theoretical Studies of Consciousness*, ed. G. R. Bock, and J. Marsh. London: John Wiley and Sons.
- Baars, B. J. 1997. *In the Theater of Consciousness*. Oxford: Oxford University Press.
- Baars, B. J. 2002. The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Science* 6:47–52.
- Baars, B. J., and S. Franklin. 2003. How conscious experience and working memory interact. *Trends in Cognitive Science* 7:166-172.
- Baddeley, A. D. 1993. Working memory and conscious awareness. In *Theories of memory*, ed. A. Collins, S. Gathercole, M. Conway, and P. Morris. Howe: Erlbaum.
- Baddeley, A. D. 2000. The episodic buffer: a new component of working memory? *Trends in Cognitive Science* 4:417-423.
- Baddeley, A. D. 2001. The concept of episodic memory. In *Episodic Memory*, ed. A. D. Baddeley, M. A. Conway, and A. Aggleton. Oxford: Oxford University Press.
- Baddeley, A., M. Conway, and J. Aggleton. 2001. *Episodic Memory*. Oxford: Oxford University Press.
- Bahrck, H. P. 1984. Semantic memory content in permastore: fifty years of memory for Spanish learned in school. *J. Exp. Psychol. Gen.* 113:1-29.
- Barsalou, L. W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22:577–609.
- Beekman, M., J. N. M. Calis, B. P. Oldroyd, and F. L. W. Ratnieks. 2002. When do honeybee guards reject their former nestmates after swarming? *Insectes Sociaux* 49:56-61.
- Bogner, M. 1999. Realizing "consciousness" in software agents. Ph.D. Dissertation. University of Memphis.
- Bogner, M., U. Ramamurthy, and S. Franklin. 2000. Consciousness" and Conceptual Learning in a Socially Situated Agent. In *Human Cognition and Social Agent Technology*, ed. K. Dautenhahn. Amsterdam: John Benjamins. 113-135.
- Broadbent, D. E. 1958. *Perception and Communication*. New York: Pergamon Press.
- Carruthers, P. 1998. Natural theories of consciousness. *European Journal of Philosophy* 6:203–222.
- Chandler, C. C. 1991. How memory for an event is influenced by related events: Interference in modified recognition tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17:115-125.
- Chrobak, J. J., and T. C. Napier. 1992. Intraseptal bicuculline produces dose- and delay dependent working/episodic memory deficits in the rat. *Neuroscience* 47:833-841.
- Clayton, N. S., D. P. Griffiths, and A. Dickinson. 2000. Declarative and episodic-like memory in animals: Personal musings of a scrub jay or When did I hide that worm over there? In *The Evolution of Cognition*, ed. C. M. Heyes, and L. Huber. Cambridge, MA: MIT Press.

- Colston, H., and D. Walter. 2001. walter@uwp.edu. The Common Thinking Errors Page. <http://www.uwp.edu/academic/psychology/demos/UTICdemo/UTICdemo.html>. 17 Jun 2001. 8 Jun 2003.
- Conway, M. A. 2001. Sensory-perceptual episodic memory and its context: autobiographical memory. In *Episodic Memory*, ed. A. Baddeley, M. Conway, and J. Aggleton. Oxford: Oxford University Press.
- Cooney, J. W., and M. S. Gazzaniga. 2003. Neurological disorders and the structure of human consciousness. *Trends in Cognitive Sciences* 7:161-165.
- Cotterill, R. M. J. 2003. CyberChild: A Simulation Test-Bed for Consciousness Studies. *Journal of Consciousness Studies* 10:31-45.
- Cox, K. J. A., and P. R. Adams. 2000. Implications of synaptic digitization and error for neocortical function. *Neurocomputing* 32:673-678.
- Damasio, A. R. 1989. Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33:25-62.
- Davachi, L., J. P. Mitchell, and A. D. Wagner. 2003. Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proc. Natl. Acad. Sci. USA*, 100:2157-2162.
- Dehaene, S. 2001. *The Cognitive Neuroscience of Consciousness*. Cambridge, MA; Amsterdam: MIT Press; Elsevier.
- Dehaene, S., and L. Naccache. 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79:1-37.
- Dehaene, S., C. Sergent, and J.-P. Changeux. 2003. A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. USA*, 100:8520-8525.
- Dennett, D. C. 1991. *Consciousness explained*. Boston: Little, Brown & Co.
- Dennett, D. C. 2001. Are we explaining consciousness yet? *Cognition* 79:221-237.
- Donald, M. 2001. *A Mind So Rare*. New York: Norton.
- Edelman, G. M. 1987. *Neural Darwinism*. New York: Basic Books.
- Ebbinghaus, H. 1885/1964. *Memory: A contribution to experimental psychology*. New York: Dover.
- Edelman, G. M. 1987. *Neural Darwinism*. New York: Basic Books.
- Edelman, G. M., and G. Tononi. 2000. *A Universe of Consciousness*. New York: Basic Books.
- Eichenbaum, H. 2002. *The Cognitive Neuroscience of Memory*. Oxford: Oxford University Press.
- Elliott, R., and R. J. Dolan. 1998. Neural response during preference and memory judgments for subliminally presented stimuli: a functional neuroimaging study. *J. Neurosci.* 18:4697-4704.
- Ericsson, K. A., and W. Kintsch. 1995. Long-term working memory. *Psychological Review* 102:21-245.
- Fiske, S. T., and S. E. Taylor. 1991. *Social Cognition*, 2nd ed. New York: McGraw-Hill.
- Franklin, S. 1997. Autonomous Agents as Embodied AI. *Cybernetics and Systems* 28:499-520
- Franklin, S. 2000a. Modeling Consciousness and Cognition in Software Agents. In *Proceedings of the Third International Conference on Cognitive Modeling, Groningen, NL, March 2000*, ed. N. Taatgen. Veenendal, NL: Universal Press.

- Franklin, S. 2000b. Deliberation and Voluntary Action in ‘Conscious’ Software Agents. *Neural Network World* 10:505-521.
- Franklin, S. 2001a. Automating Human Information Agents. In *Practical Applications of Intelligent Agents*, ed. Z. Chen, and L. C. Jain. Berlin: Springer-Verlag.
- Franklin, S. 2001b. Conscious Software: A Computational View of Mind. In *Soft Computing Agents: New Trends for Designing Autonomous Systems*, ed. V. Loia, and S. Sessa. Berlin: Springer (Physica-Verlag).
- Franklin, S. 2002. ed. IDA on Will: It's no Illusion. *Science and Consciousness Review*. 2002. http://psych.pomona.edu/scr/ED_Sep02_IDAFreeWill.htm.
- Franklin, S., and A. C. Graesser. 1997. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In *Intelligent Agents III*. Berlin: Springer Verlag. 21-35.
- Franklin, S., and A. Graesser. 2001. Modeling Cognition with Software Agents. In *CogSci2001: Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, ed. J. D. Moore, and K. Stenning. Mahwah, NJ: Lawrence Erlbaum Associates; August 1-4, 2001.
- Franklin, S., A. Kelemen, and L. McCauley. 1998. IDA: A Cognitive Agent Architecture. In *IEEE Conf on Systems, Man and Cybernetics*. : IEEE Press.
- Freeman, W. J. 1999. Consciousness, Intentionality and Causality. In *Reclaiming Cognition*, ed. R. Núñez, and W. J. Freeman. Exeter: Imprint Academic.
- Freeman, W. J. 2003a. High resolution EEG brings us another step closer to the NCC?. Plenary. ASSC7: Seventh Conference of the Association for the Scientific Study of Consciousness; Association for the Scientific Study of Consciousness; Memphis, TN, USA; May 30th to June 2nd, 2003.
- Freeman, W. J. 2003b. Evidence from human scalp electroencephalograms of global chaotic itinerancy. *Chaos* 13:1067-1077.
- Freeman, W. J. in press. The Wave Packet: An Action Potential for the 21st Century. *Biological Cybernetics*.
- Freeman, W. J., and R. Kozma. 2000. Local-global interactions and the role of mesoscopic (intermediate-range) elements in brain dynamics. *Behavioral and Brain Sciences* 23.
- Freeman, W. J., and C. Skarda. 1990. Representations: Who Needs Them? In *Brain Organization and Memory Cells, Systems, and Circuits*, ed. J. L. McGaugh. New York: Oxford University Press. 375-380.
- Fuster, J. M., M. Bodner, and J. K. Kroger. 2000. Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 405:347-351.
- Gabrieli, J. D., W. Milberg, M. M. Keane, and S. Corkin. 1990. Intact priming of patterns despite impaired memory. *Neuropsychologia* 28:417-427.
- Gardiner, J. M., C. Ramponi, and A. Richardson-Klavehn. 1998. Experiences of remembering, knowing, and guessing. *Consciousness and Cognition* 7:1-26.
- Gerevin, A., and L. Schuber. 1996. Accelerating partial-order planners: Some techniques for effective search control and pruning. *Journal of Artificial Intelligence Research* 5:95-137.
- Gilbert, C. D., M. Sigman, and R. E. Crist. 2001. The neural basis of perceptual learning. *Neuron* 31:681-697.
- Glenberg, A. M. 1997. What memory is for. *Behavioral and Brain Sciences* 20:1–19.
- Gray, J. 2002. The Sound Of One Hand Clapping. *Psyche* 8:8–11.

- Gray, J. A. 1995. The contents of consciousness: a neuropsychological conjecture. *Behavioral and Brain Sciences* 18:659–676.
- Greenwald, A. G., S. C. Draine, and R. L. Abrams. 1996. Three cognitive markers of unconscious semantic activation. *Science and Consciousness Review* 273:1699-1702.
- Halgren, E., C. Boujon, J. Clarke, C. Wang, and P. Chauvel. 2002. Rapid distributed fronto-parieto-occipital processing stages during working memory in humans. *Cerebral Cortex* 12:710-728.
- Heinrich, B. 1984. Learning in Invertebrates. In *The Biology of Learning*, ed. P. Marler, and H. S. Terrace. Berlin: Springer-Verlag.
- Hobson, J. A., and R. Stickgold. 1994. A neurocognitive approach to dreaming. *Consciousness and Cognition* 3:16–29.
- Hofstadter, D. R., and M. Mitchell. 1994. The Copycat Project: A model of mental fluidity and analogy-making. In *Advances in connectionist and neural computation theory, Vol. 2: logical connections*, ed. K. J. Holyoak, and J. A. Barnden. Norwood N.J.: Ablex.
- Jackson, J. V. 1987. Idea for a Mind. *Siggart Newsletter*, 181:23–26.
- Jacoby, L. L. 1991. A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language* 30:513-541.
- Jacoby, L. L., and M. Dallas. 1981. On the relationship between autobiographical memory and perceptual learning. *J. Exp. Psychol. Gen.* 110:306-350.
- James, W. 1890. *The Principles of Psychology*. Cambridge, MA: Harvard University Press.
- John, E. R. 2001. A Field Theory of Consciousness. *Consciousness and Cognition* 10:184-213.
- John, E. R., L. S. Prichep, W. Kox., P. Valdes-Sosa, J. Bosch-Bayard, E. Aubert, M. Tom, F. diMichele, and L. D. Gugino. 2001. Invariant reversible QEEG effects of anesthetics. *Consciousness and Cognition* 10:165-183.
- Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgement under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.
- Kanerva, P. 1988. *Sparse Distributed Memory*. Cambridge MA: The MIT Press.
- Kanwisher, N. 2001. Neural events and perceptual awareness. *Cognition* 79:89-113.
- Kelley, J. L., and A. E. Magurran. 2003. Learned predator recognition and antipredator responses in fishes. *Fish and Fisheries* 4:216.
- Koenig, T., K. Kochi, and D. Lehmann. 1998. Event-related electric microstates of the brain differ between words with visual and abstract meaning. *Electroencephalogr. Clin. Neurophysiol.* 106:535-546.
- Kondadadi, R., and S. Franklin. 2001. A Framework of Deliberative Decision Making in "Conscious" software Agents. In *Proceedings Of Sixth International Symposium on Artificial Life and Robotics (AROB-01)*.
- Laird, E. J., A. Newell, and Rosenbloom P. S. 1987. SOAR: An Architecture for General Intelligence. *Artificial Intelligence* 33:1–64.
- Lavond, D. G., J. J. Kim, and R. F. Thompson. 1993. Mammalian brain substrates of aversive classical conditioning. *Ann. Rev. Psych.* 44:317-342.

- Lehmann, D., W. K. Strik, B. Henggeler, T. Koenig, and M. Koukkou. 1998. Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. Visual imagery and abstract thoughts. *Int. J. Psychophysiol.* 29:1-11.
- Lenhart, M. D., and W. J. Freeman; 1995. A Neurophysiological Explanation of the Retroactive Inhibition Memory Effect. *Computation and Neural Systems*. Monterey CA; July 11 through July 14, 1995.
- Lindsay, D. S., and J. D. Read. 1995. MEMORY, REMEMBERING, AND MISREMEMBERING. *Ptsd Research Quarterly* 6:1-4.
- Llinas, R. R. 2001. *I of the Vortex: from Neurons to Self*. Cambridge MA: MIT Press.
- Llinas, R., U. Ribary, D. Contreras, and C. Pedroarena. 1998. The neuronal basis for consciousness. *Philos Trans R Soc Lond B Biol Sci* 353:1841-1849.
- Logan, G. D. 1992. Attention and preattention in theories of automaticity. *American Journal of Psychology* 105:317-339.
- Lyvers, M. 1999. Who Has Subjectivity? *Psyche* 5:31.
- Mandler, G. 1980. Recognizing: the judgement of previous occurrence. *Psychological Review* 87:252-271.
- Mandler, J. 2000. Perceptual and Conceptual Processes in Infancy. *Journal of Cognition and Development* 1:3-36.
- Mangan, B. 2001. Sensation's Ghost: The Non-Sensory "Fringe" of Consciousness. *Psyche* 7:<http://psyche.cs.monash.edu.au/v7/psyche-7-18-mangan.html>.
- Marsden, C. D., J. C. Rothwell, and B. L. Day. 1984. The use of peripheral feedback in the control of movement. *Trends Neurosci.* 7:253-257.
- Minsky, M. 1985. *The Society of Mind*. New York: Simon and Schuster.
- Miyake, A., and P. Shah. 1999. *Models of Working Memory*. Cambridge: Cambridge University Press.
- Morris, R. G. M. 2001. Episodic- like memory in animals. *Philos Trans R Soc Lond B Biol Sci* 356:1453-1465.
- Negatu, A., and S. Franklin. 2002. An action selection mechanism for 'conscious' software agents. *Cognitive Science Quarterly* 2.
- Negatu, A., T. L. McCauley, and S. Franklin. in preparation. Automatization for Software Agents. .
- Olton, D. S., J. T. Becker, and G. H. Handelman. 1979. Hippocampus, space and memory. *Behavioral and Brain Sciences* 2:313-365.
- Panksepp, J. 1998. *Affective Neuroscience*. Oxford: Oxford University Press.
- Posner, M. I. 1982. Cumulative Development of Attentional Theory. *American Psychologist* 37:168-179.
- Reber, A. S., F. F. Walkenfeld, and R. Hernstadt. 1991. Implicit and explicit learning: individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17:888-896.
- Rugg, M. D., and A. P. Yonelinas. 2003. Human recognition memory: a cognitive neuroscience perspective. *Trends in Cognitive Science* 7:313-319.
- Seashore, C. 1967. *he Psychology of Music*. New York: Dover.
- Shastri, L. 2001. A computational model of episodic memory formation in the hippocampal system. *Neurocomputing* 38-40:889-897.

- Shastri, L. 2002. Episodic memory and cortico-hippocampal interactions. *Trends in Cognitive Sciences* 6:162–168.
- Shulman, R. G., F. Hyder, and D. L. Rothman. 2003. Cerebral metabolism and consciousness. *CR BIOL.* 326:253-273.
- Skarda, C., and W. J. Freeman. 1987. How Brains Make Chaos in Order to Make Sense of the World. *Behavioral and Brain Sciences* 10:161-195.
- Sloman, A. 1999. What Sort of Architecture is Required for a Human-like Agent? In *Foundations of Rational Agency*, ed. M. Wooldridge, and A. Rao. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Standing, L. 1973. Learning 10,000 pictures. *Quarterly journal of Experimental Psychology* 25:207-222.
- Steinman, R. M., E. Kowler, and H. Collewijn. 1990. New directions for oculomotor research. *Vision Research* 30:1845-1864.
- Tanaka, Y., and D. Sagi. 1998. A perceptual memory for low-contrast visual signals. *Proc. Natl. Acad. Sci. USA*, 95:12729-12733.
- Taylor, J. G. 1999. *The Race for Consciousness*. Cambridge, MA: MIT Press.
- Thompson, K. G., D. P. Hanes, N. P. Bichot, and J. D. Schall. 1996. Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology* 76:4040-4055.
- Tulving, E. 1985. Memory and consciousness. *Canadian Psychology* 26:1-12.
- VanRullen, R., and C. Koch. 2003. Is perception discrete or continuous? *Trends in Cognitive Science* 7:207-213.
- Varela, F., J.-P. Lachaux, E. Rodriguez, and J. Martinerie. 2001. The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience* 2:229-239.
- Warrington, E. K., and L. Weisdrantz. 1968. New method for testing long-term retention with special reference to amnesic patients. *Nature* 217:972-974.
- Watanabe, S., S. Kawahara, and Y. Kirino. 1998. Morphological characterization of the bursting and nonbursting neurones in the olfactory centre of the terrestrial slug *Limax marginatus*. *J. Exp. Biol.* 201:1851-1861.
- Zhang, Z., S. Franklin, B. Olde, Y. Wan, and A. Graesser. 1998. Natural Language Sensing for Autonomous Agents. In *Proceedings of IEEE International Joint Symposium on Intelligence Systems 98*.
- Zulandt Schneider, R. A., R. Huber, and P. A. Moore. 2001. Individual and status recognition in the crayfish, *Orconectes rusticus*: The effects of urine release on fight dynamics. *Behaviour* 138:137-154.