9 How Deliberate, Spontaneous, and Unwanted Memories Emerge in a Computational Model of Consciousness

Bernard J. Baars, Uma Ramamurthy, and Stan Franklin

> And as soon as I had recognized the taste of the piece of madeleine soaked in her decoction of lime-blossom which my aunt used to give me . . . immediately the old grey house upon the street, where her room was, rose up like a stage set to attach itself to the little pavilion opening on to the garden which had been built out behind it for my parents . . . ; and with the house the town, from morning to night and in all weathers, the Square where I used to be sent before lunch, the streets along which I used to run errands, the country roads we took when it was fine . . .
>
> – Marcel Proust, *Remembrance of Things Past*

INTRODUCTION

In these words the novelist Marcel Proust described a flood of unbidden memories evoked by the taste of what must be the most famous cookie in the world, Proust's madeleine soaked in lime-blossom tea. It is of course an experience of spontaneous recall. Judging by numerous thought-monitoring studies, spontaneous recall is the norm in everyday thought. But because it is more difficult to study experimentally than deliberate recall, we know much less about it.

In this chapter we describe how a current theory of conscious cognition, global workspace theory, or GWT, leads naturally to a model of both deliberate and spontaneous recall. Deliberate recall is intended; spontaneous memories are not. They can be divided into two categories:

1. *acceptable* spontaneous recall (ASR), like Proust's famous rush of memories evoked by the taste of the madeleine. Such memories are interesting or pleasant or at least tolerable;
2. *unwanted* spontaneous recall (USR), such as painful traumatic events, an annoying recurrent melody, or a memory of an unresolved argument with a loved one.

We therefore have three categories altogether, deliberate recall (DR), spontaneous recall that is acceptable (ASR), and unwanted spontaneous recall (USR). A large-scale computational model of GWT, called IDA, has been developed by Franklin and colleagues (Franklin et al., 2005). IDA[1] allows the detailed modeling of GW theory, together with other well-studied cognitive mechanisms, in challenging real-world tasks (Franklin et al., 1998; Franklin and Graesser 2001; Franklin 2001a; Ramamurthy, D'Mello, & Franklin, 2003, 2004; Franklin et al., 2005). This chapter will only focus on the question of consciousness and voluntary control as they apply to recall. Because IDA is able to simulate human functioning in at least one type of highly trained expertise, our approach here is to furnish a working proof of principle, showing that the basic computational mechanisms are adequate to generate human-like cognitive functioning in a real-world task. No added theoretical constructs are needed to show three kinds of recall we discuss here: deliberate, spontaneous, and unwanted. They emerge directly from the original model.

Unwanted memories are important in posttraumatic "flashbacks," as reported in the clinical literature. While there is controversy about the accuracy of claimed memories, for example, there is little debate that repetitive thoughts and fragments of memories can occur. Wegner has been able to evoke unwanted words in an "ironic recall" paradigm, that is, an experimental method in which subjects are asked not to think of some category of ideas, such as white bears or pink elephants (1994). Unwanted memories can be annoying, or in the case of obsessional thinking, they may become disabling. In everyday life, one can simply ask people to bring to mind an intensely embarrassing personal memory, which can be quite uncomfortable. A number of clinical categories (the Axis I disorders) involve unwanted thoughts, feelings, actions, or memories. These conditions are at the more dysfunctional pole of unwanted mental events, and the study of unwanted memories may help provide some insight into them (see also Steel & Holmes, chapter 4, this volume, for an extensive treatment of involuntary memories in clinical populations).

GLOBAL WORKSPACE THEORY

Global Workspace Theory (GWT) attempts to integrate a large body of evidence, much of which has been known for decades, into a single conceptual

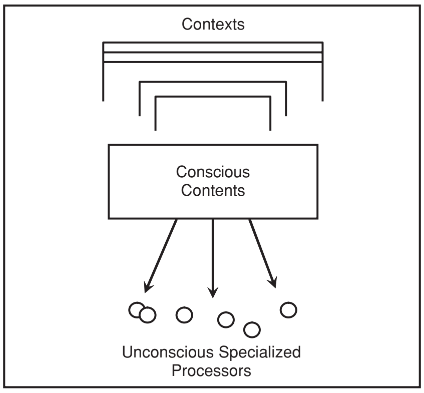

Figure 9.1 Contexts in Global Workspace Theory (Baars, 1988).

framework focused on the role of consciousness in human cognition (Baars 1988, 1997, 2002; Baars & Franklin 2003). Like other theories, GWT postulates that human cognition is implemented by a multitude of relatively small, special-purpose processors, almost always unconscious (Minsky 1985, Ornstein 1986, Edelman 1987). Processors are comparatively simple, and communication between them is relatively rare, occurring over a narrow signal bandwidth. A coalition of such processors is a collection that works together to perform a specific task. Coalitions normally perform routine actions, in pursuit of sensory, motor, or other problem-solving tasks. GWT suggests that the brain supports a global workspace capacity which allows for the integration and distribution of separate processors (for brain evidence, see Dehaene, Sergent, & Changeux, 2003; Schneider and Chein 2003; Baars 2002). A coalition of processors that gains access to the global workspace can broadcast a message to all the unconscious processors, in order to recruit new components to join in interpreting a novel situation, or in solving a current problem.

In GWT, consciousness allows the brain to deal with novel or problematic situations that cannot be dealt with efficiently, or at all, by habituated unconscious processes. In particular, it enables access to informational or memorial resources whose relevance cannot be predicted ahead of time, a problem known as "the relevance problem" in computational theories of cognition (Newell, 1990).

GWT suggests an answer to the paradox of cognitive limited capacity associated with conscious experience, immediate memory, and immediate goals. Although the brain seen from the outside has tens of billions of neurons, with trillions of connections, this massively parallel and distributed architecture has remarkable capacity limits under non-routine task conditions. Consciousness and other limited-capacity processes are very expensive biologically – if an animal is distracted while fleeing a predator, it may well die. Evolutionary pressures would therefore be expected to trend toward massive parallelism rather than a very narrow bottleneck of information processing. GWT suggests that the compensating advantage is the ability to mobilize many unconscious resources in a non-routine way to address novel challenges.

This theory offers an explanation for consciousness being serial in nature rather than parallel, as is common in the rest of the nervous system. Messages broadcast in parallel would tend to overwrite one another, making understanding difficult. It similarly explains the limited capacity of consciousness as opposed to the huge capacity typical of long-term memory and other parts of the nervous system. Large messages would be overwhelming to small, special-purpose processors.

Functions of consciousness

GWT postulates several functional roles for consciousness (Baars 1988). "Consciousness is a supremely functional adaptation" (Baars 1997). Among the several functions, the following seem relevant in the context of memory: prioritizing; aiding, recruiting, and controlling actions; error detection; learning and adaptation; and the access function. The access function is crucial in the context of memory: all functions of consciousness involve novel access between separate elements of the mental theater. Hence it seems that the most prominent function of consciousness is to increase access between separate sources of information.

Figure 9.2 (Baars 1997) illustrates this universal access function of consciousness to different sources of information, including the various memory systems. Everything in Figure 9.2 is connected to every other element through the spotlight of consciousness. Memory systems are unconscious. This theory suggests "consciousness is needed to recruit unconscious specialized networks that carry out detailed working memory functions" (Baars & Franklin 2003, p. 166).

One principal function of consciousness is therefore to recruit the relevant resources needed for dealing with novel or problematic situations. These resources may include both knowledge and procedures. They are recruited internally, but may be partly driven by stimulus input.

GWT therefore has only a few constructs: unconscious processors and a global workspace whose activity corresponds to conscious experience (under additional conditions; see Baars, 1988, chapter 4). The third basic construct is called a "context," defined as a set of unconscious constraints on conscious experiences that are reportable in the usual fashion (Figure 9.1). Constructs may involve different data types: there are goal contexts, perceptual contexts, conceptual contexts, and even cultural contexts that are shared by members

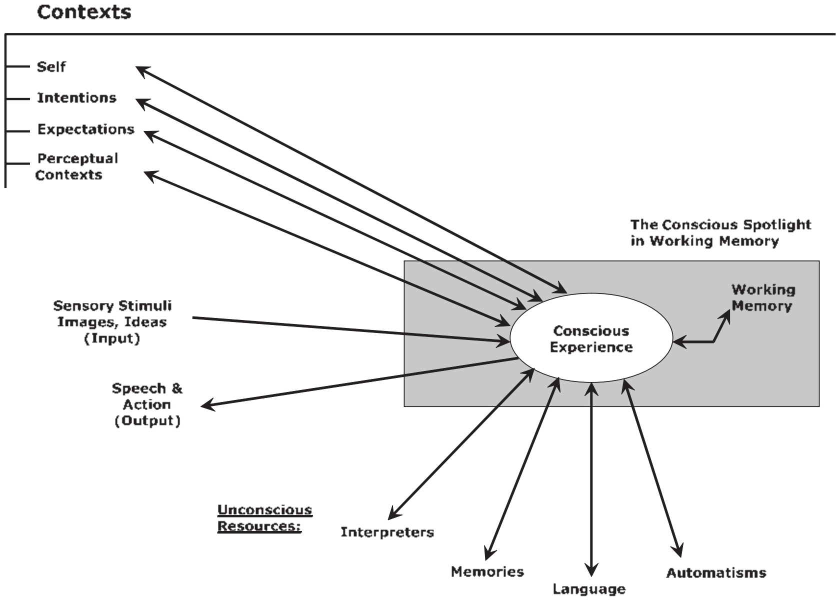

Figure 9.2 Consciousness *Creates* Access (Baars, 1997).

of a group. Each context may be considered to be a stable coalition of processors. Contexts are very similar to mental representations, semantic networks, schemas, scripts, plans – various kinds of knowledge representations. They only differ in that contexts are explicitly defined as unconscious representations which act to influence a conscious one.

Contexts develop over time in a process of coalition-formation and competition. Currently active contexts are called Dominant Contexts – such as the reader's intention to finish this chapter, his or her semantic assumptions about involuntary memories, and a host of other unconscious factors that shape the experience of this sentence and its meaning. At any point in time, therefore, the current Dominant Context influences what will come to consciousness.

William James (1890) reflects more than a century of scholarly discussion and informal experimentation on the topic of volitional control. One well-known phenomenon to students of psychology in the nineteenth century was "ideomotor control" – the tendency of people to behave according some mental image of a goal. Hypnotic arm-raising – simply by imagining one's arm floating upward – was a particularly well-known example, but there were many others, such as the Chevreul Pendulum, in which subjects were instructed to "will" a small pendulum to swing in a North–South direction, while at the same time visualizing the object moving in an East–West direction. In the Chevreul Pendulum demonstration one's intended goal is commonly overcome by the visualized direction of motion, a fact many people find astonishing. Experiments on "errors of agency" in the last decade show similar results (Wegner, 1994).

A number of established phenomena also suggest a strong influence of conscious goal imagery in the control of voluntary action. Goal imagery is commonly used to improve athletic or musical performance, and has been found in some cases to be as effective as overt practice (Beauchamp, Bray, & Albinson, 2002). On the other hand, frontal lobe damage is often associated with uncontrolled imitation of perceived others (Lhermitte, 1983), or an absence of impulse control in the case of violent or socially inappropriate actions. Phenomena like these were well known to medical practitioners and students of hypnosis (e.g., Discovery of the Unconscious).

For James, the relationship between conscious experience and the variety of unconscious brain processes posed a great philosophical problem. James was well aware of what we call "automaticity" today, which he called "habit," and indeed had written the definitive chapter on Habit that was read throughout introductory psychology courses in the United States and elsewhere, in his *Psychology – Briefer Course* (1892). Because the mind, for James, must be conscious by definition, it was difficult to conceive how a conscious thought could evoke a largely unconscious motor action. The answer, to James, was the ideomotor theory of voluntary control.

James suggests that any conscious goal image tends to trigger a habitual action unless the goal evokes some opposing idea. In a classic passage of the *Principles of Psychology* (1890), he explores the example of making a voluntary decision contrary to one's own desire: deciding to get up from a warm bed in an unheated room in the dead of winter, not an unusual event in Boston in the nineteenth century. "Now how do we ever get up under such circumstances? If I may generalize from my own experience, we more often than not get up without any struggle at all" (Baars, 1988).

The essence of voluntary control, therefore, was to keep a goal image in consciousness long enough, for unconscious habits to trigger the appropriate action. As long as opposing thoughts were kept away from consciousness, the willed action would tend to take place with no need for a great mental struggle. Human beings do what they allow themselves to imagine they will do. "This case," wrote James, "seems to me to contain in miniature form the data for an entire psychology of volition."

Quite surprisingly, James's ideomotor hypothesis fits hand in glove with the theoretical framework of GWT. Conscious goals can activate several unconscious action plans and motor routines. If conscious contents are broadcast widely among specialized unconscious processors, the goals that need to recruit, organize, and execute the plans and motor routines would be conscious. GWT says that detailed intelligence resides in specialized members of the processor-population who can interpret global messages as they relate to local conditions. Once a goal context is chosen, conscious goals tend to execute automatically. Further, conscious feedback about the results of an action is required for correcting errors. When we become conscious of a speech error, we repair it as quickly as we can, without being conscious of the details of the repair. Consciousness of errors goes along with the ability to fix the errors unconsciously, thus creating effective access to unconscious resources.

GWT adopts James's ideomotor theory as is, and provides a functional architecture for it (Baars 1997, chapter 6), which is implemented in the IDA model of GW theory (Franklin 2000b). The ideomotor hypothesis explains the phenomenology of voluntary control; the fact that we are often conscious of goals, but not of the means by which we carry out those goals in muscular actions. The intelligence of the GWT architectures is highly distributed, just as it appears to be in the brain. The essential role of consciousness, therefore, is not to compute syntactic structures or the degrees of freedom of a moving arm. Rather, it is to trigger existing unconscious "habits" – processors or automatisms – to carry out a sufficiently long-lasting conscious goal (assuming there is a fit between the goal and the effector mechanisms, of course).

What if there are no specialized habits to carry out the goal? Here GWT suggests an essential role for consciousness in learning – of actions, of cognitive routines, and even of perceptual entities like phonemes and faces. And of course, the question of learning brings us back to memory –- the lifelong archive of learned experiences, perceptual units, concepts, linguistic rules and regularities, vocabulary items, action components, common associations, coping strategies, cognitive automatisms, and the like. Thus GWT makes a strong claim for the necessary role of conscious experience in evoking memorial processes in the brain.[2]

VOLUNTARY AND NON-VOLUNTARY MEMORIES

Humans encounter both voluntary and non-voluntary memories. Imagine you are at the drive-in window of the local pharmacy to pick up your prescription medication. You are looking through the window and see a familiar-looking man in a blue blazer walk into the pharmacy. You only get a fleeting glance of this man. Though you are unable to identify and place this person, you do have a strong feeling of knowing that person. As you drive out of that place and other events take over your attention, the memory of this person you saw through the drive-in window keeps coming back to you non-voluntarily. There are two types of non-voluntary memories: wanted and unwanted. Wanted non-voluntary memories have a positive or neutral emotional content and affect associated with them, while unwanted non-voluntary memories have a negative emotional content and affect.

We also encounter voluntary memories. These are episodes that we consciously recall. For example, consider your volitional act of finding out who is the person you saw through that drive-in window at the pharmacy. As a volitional act, recalling that episode of seeing that person in a blue blazer is a voluntary memory.

We hypothesize that future non-voluntary memories may be in the service of this volitional act. We will return to these voluntary and non-voluntary memories later in this chapter after we see the IDA model, its Cognitive Cycle, and the mechanisms by which volitional actions happen in that model.

Memory systems

In this section, we will briefly discuss the various human memory systems that will play a role in our model and analyses. It will be helpful to the reader for us to specify here how we plan to use the various terms, as there is not always agreement in the literature. Figure 9.3 displays some of the relations between the memory systems we describe below.

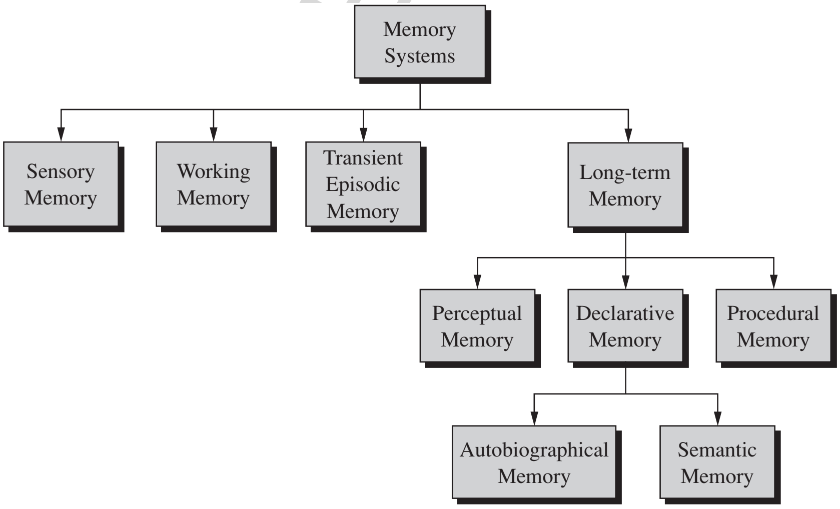

Figure 9.3 Human Memory Systems (Franklin et al., 2003).

Sensory memory holds incoming sensory data in sensory buffers 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 tens of milliseconds.

Working memory is the "manipulable scratchpad of the mind" (Miyake & 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 tens of seconds. Again, there are separate working-memory components associated with the different senses, the visuospatial sketchpad and the phonological loop, for example (Baddeley, 1993; Baars & Franklin, 2003). Also, there are long-term processing components of working memory (Ericsson & Kintsch, 1995). Baars and Franklin (2003) have suggested that conscious input, rehearsal, and retrieval are necessary for the normal functions of working memory.

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

An unusual aspect in our memory model is the transient episodic memory (TEM), an episodic memory with a decay rate measured in hours. Though often assumed (e.g., Panksepp, 1998, p. 129 assumes a "transient memory store"), the existence of such a memory has rarely been explicitly asserted (Donald, 2001; Conway, 2002; Baars & Franklin, 2003; Franklin et al., 2005).

Humans are blessed with a variety of long-term memory types that may decay exceedingly slowly, if at all. Memory researchers typically distinguish between procedural memory, the memory for motor skills (including verbal skills), and declarative memory. Declarative memory (DM) is composed of autobiographical memory and semantic memory, memories of fact or belief typically lacking a particular source with a time and place of acquisition. Semantic memories have lost their association with their original autobiographical source. These declarative memory systems are accessed by means of specific cues from working memory.

Though "perceptual memory" is often used synonymously with "sensory memory," we follow Taylor (1999) and use the term differently (p. 29). Somewhat like semantic memory, perceptual memory is a memory for individuals, categories, and their relations. Our model distinguishes between semantic memory and perceptual memory (PM), and hypothesizes distinct mechanisms for each (Franklin et al., 2005). According to this model, PM plays the major role in recognition, categorization, and more generally the assignment of interpretations. Upon presentation of features of an incoming stimulus, PM returns interpretations, the beginnings of meaning. The content of semantic memory is hypothesized to be a superset of that of PM.

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 & Magurran, 2003), and crustaceans (Zulandt Schneider, Huber, & Moore, 2001). Even mollusks learn to recognize odors (Watanabe, Kawahara, & Kirino, 1998). While common (perhaps universal) in mammals (Chrobak & Napier 1992) and birds (Clayton, Griffiths, & Dickinson, 2000), and conjectured for all vertebrates (Morris 2001), the functions of episodic memory, memories of events, seem 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 from the results of developmental studies. 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 (2000) distinguishes between perceptual and conceptual categorization in infants, the latter being based upon what objects are used for (see also Glenberg, 1997). Our model would suggest PM involvement in perceptual categorization, while semantic memory would play the more significant role in conceptual categorization.

Yet another line of empirical evidence comes from experiments with amnesiacs, such as HM. Pattern priming effects involving recognition of letters, words, shapes, and objects have been demonstrated that are comparable to the capabilities of unimpaired subjects (Gabrieli et al., 1990). These studies suggest that HM 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.

Additional support for the dissociation of PM and semantic memory comes from another study of human amnesics (Fahle & Daum, 2002). Half a dozen amnesic patients learned a visual hyperacuity task as well as did control groups though their declarative memory was significantly impaired.

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, & 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 their 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 & 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 also support a dual-process model (Rugg & Yonelinas, 2003).

Transient episodic and declarative memories have distributed representations in our model. There is evidence that this is also the case in the nervous system. In our model, these two memories are implemented computationally using a modified version of Kanerva's Sparse Distributed Memory (SDM) architecture (Kanerva, 1988; Ramamurthy et al., 2004). The SDM architecture has several similarities to human memory (Kanerva, 1988), and provides for "reconstructed memory" in its retrieval process.

"CONSCIOUS" SOFTWARE AGENTS

An *autonomous agent* is a system situated within and part of an environment. The agent senses that environment and acts on it, over time, in pursuit of its own agenda, effecting what it senses in the future (Franklin & Graesser, 1997). The agent is structurally coupled to its environment (Maturana, 1975; Maturana & Varela, 1980; Varela, Thompson, & Rosch, 1991). An autonomous software agent is one which "lives" in computer systems, and connects to the networks and email systems. When an autonomous software agent is equipped with computational versions of cognitive features, such as multiple senses, perception, various forms of memory including transient episodic memory and declarative memory, learning, emotions, and multiple drives, it is called a *cognitive software agent* (Franklin, 1997). Such cognitive software agents promise to be more flexible, more adaptive, more human-like than the classical, existing software because of their ability to learn, and to deal with novel input and unexpected situations.

One way to design and implement cognitive software agents is to do it within the constraints of GWT, discussed earlier in this chapter. Agents built within the constraints of GWT are called *"conscious" software agents*. No claim of sentience or phenomenal consciousness is being made. IDA is such a "conscious" software agent; her architecture is described in the next section.

THE IDA ARCHITECTURE

IDA's task presents both communication problems and action-selection problems involving constraint satisfaction. She must communicate with sailors via email and in natural language, understanding the content and producing life-like responses. Sometimes IDA will initiate conversations. She must access a number of databases, again understanding the content. She must see that the Navy's needs are satisfied, for example, the required number of sonar technicians on a destroyer with the required types of training. In doing so, IDA must adhere to a number of Navy policies. She must hold down moving costs. And IDA must cater to the needs and desires of the sailors as well as is possible. This includes negotiating with the sailor via email, in natural language. IDA employs deliberative reasoning in the service of action selection. Before going further, it is important that we distinguish between IDA as a computational model and as a conceptual model. The computational IDA is a running piece of Java code, an actual software agent. The conceptual IDA model includes everything in the computational model with relatively minor changes. It also includes, however, additional functionality that has been designed but not yet implemented.

As we hypothesize that humans also do, the IDA model runs in a rapidly continuing sequence of partially overlapping cycles, called cognitive cycles (Baars & Franklin, 2003). These cycles will be discussed in detail below, after the IDA architecture and its mechanisms are described.

"Conscious" software architecture and mechanisms

The IDA architecture is partly symbolic and partly connectionist, at least in spirit. Although there are no artificial neural networks as such, spreading activation abounds. The mechanisms used in implementing the several modules have been inspired by a number of different new AI techniques (Hofstadter & Mitchell, 1994; Holland, 1986; Jackson, 1987; Kanerva, 1988; Maes, 1989; Minsky, 1985). The architecture is partly composed of entities at a relatively high level of abstraction, such as behaviors, message-type nodes, emotions, etc., and partly of low-level codelets.

Each codelet is a small piece of code performing a simple, specialized task. They correspond to Baars's processors in GWT (1988). Most codelets are, like demons in an operating system, always watching for a situation to arise, making it appropriate to act. Codelets come in many varieties: perceptual codelets, information codelets, attention codelets, behavior codelets, expectation codelets, etc. Though most codelets subserve some high-level entity, many codelets work independently. Codelets do almost all the work. IDA can almost be viewed as a multiagent system, though not in the usual sense of the term.

As noted above, most of IDA's various entities, both high-level entities and codelets, carry and spread some activation. Such activation typically hopes to measure some sort of strength or relevance. Unless told otherwise, it is safe to assume that every activation decays over time. Finally, note that though the IDA architecture is conveniently described in terms of modules, it is, in fact, tightly linked. Like the brain, the IDA architecture is both modular and highly interconnected.

Perception

IDA perceives both exogenously and endogenously (Zhang et al., 1998). (In humans, the concept of endogenous perception includes visual imagery, inner speech, and memory-based percept-like experiences.) The stimuli of IDA's single sense are strings of characters. We use Barsalou's perceptual symbol systems as a guide (1999). The perceptual knowledge base of this agent, called perceptual memory, takes the form of a semantic net with activation called the slipnet. The name is taken from the Copycat architecture that employs a similar construct (Hofstadter & Mitchell, 1994). Nodes of the slipnet constitute the agent's perceptual symbols, representing individuals, categories, and higher-level ideas and concepts. A link of the slipnet represents a relation between its source node and its sink node.

An incoming stimulus, say an email message, is descended upon by a horde of perceptual codelets. Each of these codelets is looking for some particular string or strings of characters, say one of the various forms of the name of the city of Norfolk. Upon finding an appropriate character string, the codelet will activate an appropriate node or node in the slipnet. The slipnet will eventually settle down. Nodes with activations over threshold and their links are taken to be the constructed meaning of the stimulus. Pieces of the slipnet containing nodes and links, together with perceptual codelets with the task of copying the piece to working memory, constitute Barsalou's perceptual symbol simulators.

Perceptual learning

As described above, IDA's perceptual memory, including the slipnet and the perceptual codelets, is a fully implemented part of the running IDA computational model. The IDA conceptual model adds learning to this perceptual memory with updating during the broadcast (Step 5) of each cognitive cycle (Franklin et al., 2005). New nodes and links are added to the slipnet as needed, while existing node and links have their base-level activations and weights updated, respectively.

Memory

IDA employs sparse distributed memory (SDM) for the major associative, episodic memories (Anwar, Dasgupta, & Franklin, 1999; Anwar & Franklin, 2003; Ramamurthy et al., 2004). SDM is a content-addressable memory that, in many ways, is an ideal computational mechanism for use as a long-term associative memory (Kanerva, 1988). Content-addressable means that items in memory can be retrieved by using part of their contents as a cue, rather than having to know the item's address in memory.

The inner workings of SDM rely on large binary spaces, that is, spaces of vectors containing only zeros and ones, called bits. These binary vectors, called words, serve as both the addresses and the contents of the memory. The dimension of the space determines the richness of each word. These spaces are typically far too large to implement in any conceivable computer. Approximating the space uniformly with some manageable number of actually implemented, hard locations surmounts this difficulty. The number of such hard locations determines the carrying capacity of the memory. Features are represented as one or more bits. Groups of features are concatenated to form a word. When writing a word to memory, a copy of the word is placed in all close enough hard locations. When reading a word, a close enough cue would reach all close enough hard locations and get some sort of aggregate or average out of them. As mentioned above, reading is not always successful. Depending on the cue and the previously written information, among other factors, convergence or divergence during a reading operation may occur. If convergence occurs, the pooled word will be the closest match (with abstraction) of the input reading cue. On the other hand, when divergence occurs, there is no relation, in general, between the input cue and what is retrieved from memory.

SDM is much like human long-term declarative memory. A human often knows what he or she does or does not know. If asked for a telephone number you have once known, you may search for it. When asked for one you have never known, an immediate "I don't know" response ensues. SDM makes such decisions based on the speed of initial convergence. The reading of memory in SDM is an iterative process. The cue is used as an address. The content at that address is read as a second cue, and so on, until convergence, that is, until subsequent contents look alike. If it does not quickly converge, an

"I don't know" is the response. The "on-the-tip-of-my-tongue phenomenon" corresponds to the cue having content just at the threshold of convergence. Yet another similarity is the power of rehearsal, during which an item would be written many times and, at each of these, to a thousand locations – that is, the distributed part of sparse distributed memory. A well-rehearsed item can be retrieved with smaller cues. Another similarity is interference, which would tend to increase over time as a result of other similar writes to memory. The IDA conceptual model uses variants of SDM to implement both transient episodic memory and declarative memory (Franklin et al., 2005; Ramamurthy et al., 2004).

Transient episodic memory

Transient episodic memory is an unusual aspect of the IDA conceptual model. It is an episodic memory with a decay rate measured in hours. Though a "transient memory store" is often assumed (Panksepp, 1998, p. 129), the existence of such a memory has rarely been explicitly asserted (Donald, 2001; Conway, 2002; Baars & Franklin, 2003, Franklin et al., 2005). In the IDA conceptual model, transient episodic memory is updated during Step 5 of each cognitive cycle with the contents of "consciousness." We have expanded and tested our implementation of an experimental transient episodic memory using a ternary revision of sparse distributed memory allowing for an "I don't care" symbol (Ramamurthy et al., 2004).

"Consciousness"

In IDA, the processors postulated by GWT are implemented by codelets, small pieces of code. These are specialized for some simple task and often play the role of a demon waiting for an appropriate condition under which to act. The apparatus for producing "consciousness" consists of a coalition manager, a spotlight controller, a broadcast manager, and a collection of attention codelets that recognize novel or problematic situations (Bogner, 1999; Bogner, Ramamurthy, & Franklin, 2000). Each attention codelet keeps a watchful eye out for some particular situation to occur that might call for "conscious" intervention. Upon encountering such a situation, the appropriate attention codelet will be associated with the small number of information codelets that carry the information describing the situation. This association should lead to the collection of this small number of codelets, together with the attention codelet that collected them, becoming a coalition. Codelets also have activations. Upon noting a suitable situation, an attention codelet will increase its activation as a function of the match between the situation and its preferences. This allows the coalition, if one is formed, to compete for "consciousness."

The coalition manager is responsible for forming and tracking coalitions of codelets. Such coalitions are initiated on the basis of the mutual associations between the member codelets. During any given cognitive cycle, one of these coalitions finds its way to "consciousness," chosen by the spotlight controller, who picks the coalition with the highest average activation among its member codelets. GWT calls for the contents of "consciousness" to be broadcast to each of the codelets. The broadcast manager accomplishes this.

Action selection

IDA depends on an enhancement of Maes's behavior net (1989) for high-level action selection in the service of built-in drives (Song & Franklin, 2000; Negatu & Franklin, 2002). Each has several distinct drives operating in parallel and implemented in the IDA conceptual model by feelings and emotions. These drives vary in urgency as time passes and the environment changes. The goal contexts of GWT are implemented as behaviors in the IDA model. Behaviors are typically mid-level actions, many depending on several behavior codelets for their execution. A behavior net is composed of behaviors and their various links. A behavior looks very much like a production rule, having preconditions as well as additions and deletions. A behavior is distinguished from a production rule by the presence of an activation, which is a number indicating some kind of strength level. Each behavior occupies a node in a digraph (directed graph). The three types of links of the digraph are completely determined by the behaviors. If a behavior *X* will add a proposition *b*, which is on behavior *Y*'s precondition list, then put a successor link from *X* to *Y*. There may be several such propositions, resulting in several links between the same nodes. Next, whenever you put in a successor going one way, put in a predecessor link going the other. Finally, suppose you have a proposition m on behavior *Y*'s delete list that is also a precondition for behavior *X*. In such a case, draw a conflictor link from *X* to *Y*, which is to be inhibitory rather than excitatory.

As in connectionist models, this digraph spreads activation. The activation comes from four sources: from activation stored in the behaviors, from the environment, from drives (through feelings and emotions in the IDA conceptual model), and from internal states. The environment awards activation to a behavior for each of its true preconditions. The more relevant it is to the current situation, the more activation it is going to receive from the environment. This source of activation tends to make the system opportunistic. Each drive awards activation to every behavior that, by being active, will satisfy that drive. This source of activation tends to make the system goal directed. Certain internal states of the agent can also send activation to the behavior net. This activation, for example, might come from a coalition of behavior codelets responding to a "conscious" broadcast. Finally, activation spreads from behavior to behavior along links. Along successor links, one behavior strengthens those behaviors with preconditions that it can help fulfill by sending them activation. Along predecessor links, one behavior strengthens any other behavior with an add list that fulfills one of its own preconditions. A behavior sends inhibition along a conflictor link to any other behavior that can delete one of its true preconditions, thereby weakening it. Every conflictor link is inhibitory. A behavior is executable if all of its preconditions are satisfied. To be acted upon, a behavior must be executable, must have activation over threshold, and must have the highest such activation. Behavior nets produce flexible, tunable action selection for these agents.

Behaviors in these agents almost always operate as part of behavior streams, which correspond to goal context hierarchies in GWT. Visualize a behavior stream as a subgraph of the behavior net, with its nodes connected by predecessor links. A behavior stream is sometimes a sequence, but not always. It can fork in either a forward or backward direction. A behavior stream can be usefully thought of as a partial plan of action.

Deliberation

IDA's complex domain requires deliberation in the sense of creating possible scenarios and partial plans of actions and then choosing between them. For example, suppose IDA is considering a ranked list of several possible jobs for a given sailor produced by her constraint satisfaction module, all seemingly suitable. IDA must construct a scenario for at least one of these possible billets. In each scenario, the sailor leaves his or her current position during a certain time interval, spends a specified length of time on leave, possibly reports to a training facility on a certain date, and arrives at the new billet within a given time window. Such scenarios are judged on how well they fit the temporal constraints and on moving and training costs.

As in humans, deliberation is mediated by the "consciousness" mechanism. Imagine IDA in the context of a behavior stream with a goal to construct a scenario to help evaluate a particular job for a particular sailor. She must first decide on a departure date within an allowable window, the first event of the scenario. Then, events for travel time (often in more than one segment), leave time (occasionally in several segments), training time (with specified dates), and arrival date must be decided upon, again within an appropriate window. If the first try does not work, IDA typically starts over with a suitably adjusted departure date. If still unsuccessful after several tries, IDA will give up on that particular job and go on to another. When successful, the job in question is so marked in working memory and becomes a candidate for voluntary selection (see below) to be offered to the sailor. Each step in this process will require several cognitive cycles, as described below. Thus, IDA is capable of temporal deliberation.

Voluntary action

Deliberation is also used in IDA to implement voluntary action in the form of William James's ideomotor theory as prescribed by GWT (Baars, 1988, chapter 7). Suppose scenarios have been constructed for several of the more suitable jobs. An attention codelet spots one that it likes, possibly due to this codelet's predilection for, say, low moving costs. The act of an attention codelet's bringing one of these candidates to consciousness serves to propose it. This is James's idea popping into mind. If no other attention codelet brings an objection to consciousness or proposes a different job, a timekeeper codelet assigned the particular task of deciding will conclude, after a suitable time having passed, that the proposed job will be offered, and starts the process by which it will be so marked in working memory. Objections and proposals can continue to come to consciousness, but the patience of the timekeeper codelet dampens as time passes. Also, the activation of a given attention codelet tends to diminish after winning a competition for consciousness in any given cognitive cycle. This lessening makes it less likely that this particular attention codelet will be successful in proposing, objecting, or supporting in the near future. This diminishing of patience and activation serve to prevent continuing oscillations in the voluntary action selection process.

Learning

The learning extension of the IDA model is the LIDA (Learning IDA) architecture. With the help of feelings and emotions as primary motivators and learning facilitators, the LIDA architecture adds three fundamental, continuously active, learning mechanisms that underlie much of human learning: (1) perceptual learning, the learning of new objects, categories, relations, etc.; (2) episodic learning of events, the what, where, and when; (3) procedural learning, the learning of new actions and action sequences with which to accomplish tasks new to our existing IDA system.

Perceptual learning in the LIDA model occurs with consciousness. This learning is of two forms, the strengthening or weakening of the base-level activation of existing nodes, as well as the creation of new nodes and links. Any existing concept or relation that appears in the conscious broadcast (Step 5 of the cognitive cycle) has the base-level activation of its corresponding node strengthened as a function of the arousal of the agent at the time of the broadcast.

Episodic learning in the LIDA architecture results from events taken from the contents of "consciousness" being encoded in our modified sparse distributed memory (SDM) representing TEM. In addition to the encoding of the sensory perceptual details of each episode manifested through the contents of consciousness, this episodic learning includes the encoding of feelings and emotions, and of actions taken by the agent. Periodically, and offline, the not yet decayed contents of TEM are consolidated into declarative memory (DM) (autobiographical memory plus semantic memory), which is also implemented as a modified SDM system. Conway stipulates that as an aftermath of the consolidation process, previously volatile events acquire high stability and durability (2001). This scenario mirrors our view of the still controversial question of how human episodic memory works.

Procedural learning in LIDA is a combination of both *instructionalist* as well as *selectionist* motivated agendas, with consciousness providing reinforcement to actions. Reinforcement is provided via a sigmoid function such that initial reinforcement becomes very rapid but tends to saturate. The inverse of the sigmoid function that produces the reinforcement curve serves as the decay curve. With such procedural learning, the agent is capable of learning new ways to accomplish new tasks by creating new actions and action sequences. With feelings and emotions serving as primary motivators and learning facilitators, every action, exogenous and endogenous, taken by an agent controlled with the LIDA architecture is self-motivated.

THE IDA COGNITIVE CYCLE

IDA functions by means of flexible, serial but cascading cycles of activity that we refer to as cognitive cycles. We will next explore the cognitive cycle in detail (as shown in Figure 9.4), in order to facilitate the reader's understanding of the later material on non-voluntary memory.

1. *Perception.* Sensory stimuli, external or internal, are received and interpreted by perception creating meaning. Note that this stage is unconscious.
   a. Early perception: Input arrives through senses. Specialized perception codelets descend on the input. Those that find features relevant to their specialty activate appropriate nodes in IDA's slipnet (a semantic net with activation).

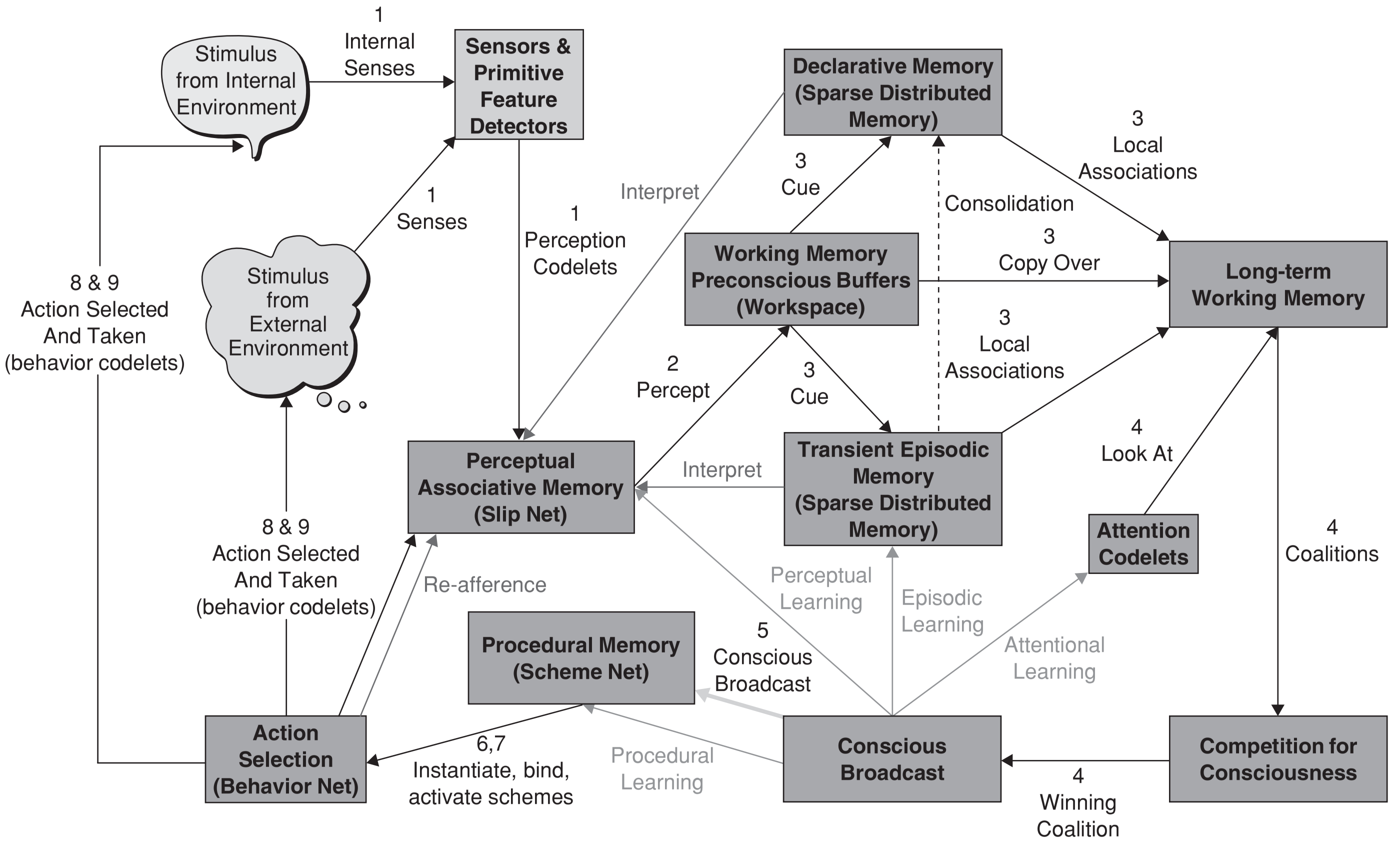
Stimulus from Internal Environment
1 Internal Senses
Sensors & Primitive Feature Detectors
Declarative Memory (Sparse Distributed Memory)
3 Local Associations
1 Senses
Interpret
3 Cue
Consolidation
1 Perception Codelets
Working Memory Preconscious Buffers (Workspace)
3 Copy Over
Long-term Working Memory
8 & 9 Action Selected And Taken (behavior codelets)
Stimulus from External Environment
2 Percept
3 Cue
3 Local Associations
4 Look At
Perceptual Associative Memory (Slip Net)
Interpret
Transient Episodic Memory (Sparse Distributed Memory)
8 & 9 Action Selected And Taken (behavior codelets)
Attention Codelets
4 Coalitions
Re-afference
Perceptual Learning
Episodic Learning
Procedural Memory (Scheme Net)
5 Conscious Broadcast
Attentional Learning
Action Selection (Behavior Net)
6,7 Instantiate, bind, activate schemes
Procedural Learning
Conscious Broadcast
4 Winning Coalition
Competition for Consciousness

Figure 9.4 IDA's Cognitive Cycle.

   b. Chunk perception: Activation passes from node to node in the slipnet. The slipnet stabilizes, bringing about the convergence of streams from different senses and chunking bits of meaning into larger chunks. These larger chunks, represented by meaning nodes in the slipnet, constitute the percept.

2. *Percept to preconscious buffer*. The percept, including some of the data plus the meaning, is stored in preconscious buffers of IDA's working memory.
3. *Local associations*. Using the incoming percept and the residual contents of the preconscious buffers as cues, local associations are automatically retrieved from transient episodic memory and from declarative memory. The contents of the preconscious buffers, along with the local associations retrieved from transient episodic memory and declarative memory, together constitute long-term working memory (Ericsson & Kintsch, 1995; Baddeley, 2000).
4. *Competition for "consciousness."* Attention codelets, whose job it is to bring relevant, urgent, or insistent events to "consciousness," view long-term working memory. Some of them gather information, form coalitions, and actively compete for access to "consciousness." The competition may also include attention codelets from a recent previous cycle.

   The activation of unsuccessful attention codelets decays, making it more difficult for them to compete with newer arrivals. However, the contents of unsuccessful coalitions remain in the preconscious buffer and can serve to prime ambiguous future incoming percepts. The same is true of contents of long-term working memory that are not picked up by any attention codelet.
5. *"Conscious" broadcast*. A coalition of codelets, typically an attention codelet and its covey of related information codelets carrying content, gains access to the global workspace and has its contents broadcast. The current contents of "consciousness" are also stored in transient episodic memory. At recurring times not part of a cognitive cycle, the contents of transient episodic memory are consolidated into long-term associative memory.
6. *Recruitment of resources*. 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 was an expectation codelet calling attention to an unexpected result from a previous action, the responding codelets may be those that can help to rectify the unexpected situation. Thus "consciousness" solves the relevancy problem in recruiting resources.
7. *Setting goal context hierarchy*. Some responding behavior codelets instantiate an appropriate behavior stream, if a suitable one is not already in place. 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 using additional mechanisms is called for.

8. *Action chosen.* The behavior net chooses a single behavior (goal context) 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), by the current situation, external and internal conditions, by the relationship between the behaviors, and by the activation values of various behaviors.
9. *Action taken.* The execution of a behavior (goal context) results in the behavior codelets performing their specialized tasks, which may have external or internal consequences. This is IDA taking an action. The acting codelets also include an expectation codelet (see Step 6) whose task it is to monitor the action, and to try and bring to "consciousness" any failure in the expected results.

IDEOMOTOR THEORY AND ITS IMPLEMENTATION AS VOLITION IN IDA

Long ago, 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. He speaks at length of the case of deciding to get out of a warm bed into an unheated room in the dead of a New England winter. The act of standing up occurs when a goal image, namely the thought of standing, comes to consciousness and remains long enough to trigger unconscious effectors. If a conscious inner debate occurs for and against rising out of bed, the action would be inhibited. "This case seems to me to contain in miniature form the data for an entire psychology of volition." GWT adopts James's ideomotor theory as is, and provides a functional architecture for it (Baars, 1997, chapter 6; Franklin, 2000).

We humans most often select actions subconsciously, that is, without conscious thought about which action to take. Sometimes when we speak, we are surprised at what we say. But we humans also make voluntary choices of action, often as a result of deliberation. Baars argues that 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, being conscious of the decision. We are typically conscious of our speech (the results of actions) but not typically conscious of the decision to speak. However, sometimes, as in a formal meeting, we may consciously decide to speak and then do so. The decision itself becomes conscious. It is the latter case that constitutes voluntary action. Here we provide an underlying mechanism that implements that theory of volition and its architecture in IDA.

Though voluntary action is often deliberative, it can also be reactive in the sense of Sloman (1999), who allows for the possibility or the action-selection mechanism being quite complex. Suppose that, while sitting on the couch in your living room, you decide you would like a cup of coffee and thereafter head for the kitchen to get it. The decision may well have been taken voluntarily, that is, consciously, without your having deliberated about it by considering alternatives and choosing among them. Voluntary actions may also be taken metacognitively (by Sloman's meta-management processes). For example, you may consciously decide to be more patient in the future with your child. That would be a voluntary metacognitive decision.

What about action-selection decisions in IDA? Are they voluntary or not? Both kinds occur. When IDA reads a sailor's projected rotation date from the personnel database, she formulates and transmits a query to the database and accepts its response. The decision to make the query, as well as its formulation and transmission, is done unconsciously. The results of the query, the date itself, do come to "consciousness." This situation is analogous to that of almost all human actions. On the other hand, IDA performs at least one voluntary action, that of choosing a job or two, or occasionally three, to offer a sailor. How is this done?

In the situation in which this voluntary action occurs, at least one scenario has been successfully constructed in the workspace, as described in the previous section. The players in this decision-making process include several proposing-attention codelets and a timekeeper codelet. A proposing-attention codelet's task is to propose that a certain job be offered to the sailor. This is accomplished by it bringing information about itself and about the proposed job to "consciousness" so that the timekeeper codelet can know of it. This proposing-attention codelet (and its brethren) choose a job to propose on the basis of its particular pattern of preferences. The preferences include several different issues with differing weights assigned to each. The issues typically include priority (stated on the job requisition list), gap (how well the departing and arriving time intervals are respected), cost of the move, fitness value, and others.

For example, our proposing attention codelet may place great weight on low moving cost, some weight on fitness value, and little weight on the others. This codelet may propose the second job on the scenario list because of its low cost and high fitness, in spite of low priority and a sizable gap. What happens then? There are several possibilities. If no other proposing attention codelet objects (by bringing itself to "consciousness" with an objecting message) and no other such codelet proposes a different job within a span of time kept by the timekeeper codelet, the timekeeper codelet will mark the proposed job as being one to be offered. If an objection or a new proposal is made in a timely fashion, it will not do so.

Two proposing attention codelets may alternatively propose the same two jobs several times. What keeps IDA from oscillating between them forever? There are three possibilities. The second time a codelet proposes the same job it carries less activation and so has less chance of being selected for the spotlight of "consciousness." Also, the timekeeper loses patience as the process continues, thereby diminishing the time span required for a decision. Finally, the metacognitive module watches the whole process and intervenes if things get too bad.

A job proposal may also alternate with an objection, rather than with another proposal, with the same kinds of consequences. These occurrences may also be interspersed with the creation of new scenarios. If a job is proposed but objected to, and no other is proposed, the scenario building may be expected to continue yielding the possibility of finding a job that can be agreed upon. We hypothesize that this procedure mimics the way humans make such decisions. It provides a mechanism for voluntary action.

VOLUNTARY VERSUS NON-VOLUNTARY EPISODIC MEMORIES IN HUMANS AND IN IDA

Non-voluntary memories

Let us consider that episode of you seeing a man in a blue blazer through the drive-in window at the local pharmacy. Suppose you recognize the person as your optometrist. That recognition is an example of non-voluntary perceptual memory. If you also remember an incident from your last check-up, that's non-voluntary episodic memory. Such memories are non-voluntary, since they were not the result of any conscious volitional action. Though the memories came to consciousness, they arose through unconscious processes.

Suppose you do not recognize the person right away, yet there is a strong feeling of knowing who he is. GWT postulates that such strong feelings of knowing are due to *fringe consciousness* (á la William James) (Baars, 1997; Mangan, 2001). From the spotlight of global workspace, we get immediate and detailed experiences. The spotlight has a hazy penumbra to represent the fringe consciousness. From this fringe, we have reliable access to information, without being able to experience the sensory events explicitly in detail. Such fringe consciousness is also non-voluntary since it arises from unconscious processes.

Now suppose that in response to the feeling that you should know who he is, you begin to consciously try to remember him. Should you succeed, this would be an example of a voluntary memory, in the sense that it arose from a consciously taken decision to try to remember.

Let us analyze these example episodes through IDA's cognitive cycle. Suppose you see the person in the blue blazer for a fleeting moment through the drive-in window of the local pharmacy and recognize him. In step 1 of the cognitive cycle, the stimuli from this person in the blue blazer come through the senses. The specialized perception codelets process these stimuli. Those codelets that find relevant features in the stimuli activate the appropriate nodes in IDA's slipnet. In step 2, a percept containing these nodes comes to IDA's preconscious working memory buffers. Step 3 cues episodic memories, resulting in local association of some explicit incident during your last check-up visit. If all these items survive the competition for consciousness (step 4), you will become conscious of who the person is and of the previous incident concerning him, both in step 5, the conscious broadcast.

Now suppose you don't recognize the person, but experience the "I know that person" feeling. In this case the node in the slipnet for that man in the blue blazer will not be activated sufficiently (step 2) for it to become part of the percept (step 3). Instead, a fringe consciousness-attention codelet notices something in long-term working memory (step 4) that allows it to bring to consciousness (step 5) the "I should know him" idea. The IDA model predicts that this something would arise from the local associations produced in step 3. "I know that person" is a specialization of a fringe-consciousness feeling of familiarity to a stimulus which is clearly from a person (Mangan, 2001). In the IDA model, episodic memory is implemented with sparse distributed memory (Kanerva, 1988), which is able to distinguish between likely knowing something but not being able to find it, and not knowing it at all. This is an example of a testable cognitive hypothesis coming from the IDA model. A corollary to this hypothesis is that an animal without episodic memory will not experience such a feeling of familiarity.

Voluntary memories

Suppose the non-voluntary memory of that man in the blue blazer keeps coming back to you, and you make a volitional decision to remember who this man is. When you go about that volitional act, we hypothesize that you go through multiple cognitive cycles, both to make the volitional decision and to get as many detailed features as you can from the percept. The former was described above in the section on ideomotor theory and volition. In particular, an *intention codelet*, a particular kind of attention codelet, will be generated whose task is to bring to consciousness any information likely to help with the recognition of the man in the blue blazer. For the latter, those precepts in the multiple cycles bring out local associations from the transient episodic memory and the declarative memory. The persistence of these percepts and local associations make them likely to succeed in the competition for consciousness in step 4 of each of the cognitive cycles. Hence they will likely be in the conscious broadcasts of step 5. Those local associations and conscious broadcasts over multiple cognitive cycles finally bring about the recognition of who this person in the blue blazer is or, perhaps, they will fail. Until the intention codelet decays away, it will continue to watch long-term working memory for information perhaps useful to the goal (step 4). This is how, according to the IDA model, recognition of the man in the blue blazer suddenly "pops into mind" after some time of thinking of something else. Thus the voluntary memory is complete.

Non-voluntary memories support voluntary memories

We hypothesize that non-voluntary memories occur in service of volitional acts of retrieving voluntary memories (see Mace, chapter 3, this volume, for a similar view). Further, we argue that the fine-grained analysis using IDA's cognitive cycles allows us to see how these non-voluntary memories arise through the action of particular unconscious processes. As we have seen, these include perceptual associative memory (step 2), working memory (step 3), long-term working memory (step 4), and the competition for consciousness (step 5). Even the latter steps in a cognitive cycle may contribute when we select the internal action (steps 6, 7, and 8) of following a particular line of thought (internal dialog). This internal dialog is produced by the deliberation process described above. In contrast to this fine-grained analysis provided by the IDA model, traditional models analyze in a coarse-grained fashion, and such high-level abstractions cause these non-voluntary memories to be subsumed by the voluntary memories, missing an important aspect of the cognitive structure.

Consider again the example of your fleeting sight of the man in the blue blazer at the drive-in window of the local pharmacy. For the rest of the day, that fleeing sight of the man in the blue blazer keeps coming back to you as non-voluntary memory. You finally decide to find out who that person is. This volitional act follows William James's ideomotor theory, as we discussed earlier. Our hypothesis says that all the future non-voluntary memories, which are produced in single cognitive cycles, may be in the service of this volitional act of wanting to figure out who this person in the blue blazer is.

WANTED VERSUS UNWANTED NON-VOLUNTARY MEMORIES IN HUMANS AND IN IDA

Non-voluntary memories are of two types, namely wanted and unwanted. Let us look at our example of the man in the blue blazer again. In IDA's perceptual memory, the slipnet nodes have links to emotion-nodes. If the "recognition" of this person has positive emotional content and/or effect, then this non-voluntary memory will recur with that positive emotional effect. So, while you go through the rest of the day, this non-voluntary memory coming back to you has a positive effect. Hence it becomes a *wanted* non-voluntary memory.

On the other hand, if this person's features in the slipnet has associations with negative emotion-nodes (for example, fear or anger), then when this non-voluntary memory comes back to you through the day, it has a negative affect. Even if you make a meta-cognitive, volitional decision not to think about this person in the blue blazer, the non-voluntary memory may keep coming back. Thus it becomes an *unwanted* non-voluntary memory.

How can such a situation occur? The volitional decision to suppress the memory will produce an intention codelet (see above) that will attempt to bring to consciousness strategies for such suppression; for example, occupy yourself with a task that requires attention, or think of something pleasant when such unwanted memories occur. Though such strategies may be successful to some extent, unwanted, non-voluntary memories can continue to occur as a result of random stimuli producing local associations from episodic memory. Here's a possible scenario. The appearance of a woman in a blue skirt results in slipnet nodes for Person, Blue, Article-of-Clothing being part of the percept produced in a particular cognitive cycle (step 2). This percept-cueing episodic (declarative or transient episodic) memory (step 3) produces as a local association (step 3) a record of the emotionally laden event of encountering the man in the blue blazer. Some attention codelet (step 4), intent on bringing, say, fearful situations to consciousness, succeeds in winning the competition (step 4), resulting in the unwanted memory coming to consciousness (step 5). All these mostly unconscious processes occur independently of the volitional (conscious) decision to suppress such memories. That decision only produced an intention codelet that may not be able to continually compete successfully with highly emotionally charged local associations for access to consciousness.

Do note that the situation described in the previous paragraph can well recur over time. In the short run, the unwanted memory itself, via internal sensing (step 1), can result, by the same processes described above, in the elaboration via local association of details of the current incident and/or of details of an earlier incident which led to the original, say, fear. Over a longer time span, different random stimuli can trigger, with varying frequency, the same emotionally charged unwanted non-voluntary memory and its short-term elaboration. With each recurrence, the voluntary decision to suppress the unwanted memory is likely to be renewed, renewing or extending the life of the resulting intention codelet and that of the only partially effective strategies it espouses. These strategies are most effective in reducing the elaboration phase of the unwanted memory by competing with it for access to consciousness. As we have seen, they are not always effective in preventing the unwanted memories.

CONCLUSION

We have presented the IDA model which includes a large-scale computational model of GWT. The cognitive cycle of the IDA model provides an important tool for fine-grained analysis of cognitive processes. Using this tool, we have presented two types of memories: voluntary and non-voluntary. We hypothesize that the fine-grained analyses provide us with the ability to see the non-voluntary memory recalls which may be in the service of voluntary memories through volitional actions. We have classified non-voluntary memories into two types (wanted and unwanted), based on emotional content and effect.

NOTES

1 IDA (Intelligent Distribution Agent) is a "conscious" software agent that was developed for the US Navy (Franklin, Kelemen, & McCauley, 1998). At the end of each sailor's tour of duty, the sailor is assigned to a new billet. This assignment process is called *distribution*. The Navy employs some 280 people, called detailers, to effect these new assignments. IDA's task is to facilitate this process by completely automating the role of detailer.

2 Note that implicit learning fits this description, in that subjects are always asked to be conscious of a stimulus set, which triggers unconscious memorial and interpretive processes in the brain. What is "implicit" in implicit learning is not the process of learning, but rather the unconscious acquisition of knowledge triggered by conscious access to a set of stimuli. Along the same lines, GWT defines "selective attention" simply as the set of mechanisms that enable access to consciousness.

REFERENCES

Anwar, A., Dasgupta, D., & Franklin, S. (1999, July). *Using genetic algorithms for sparse distributed memory initialization.* Paper presented at the international conference on Genetic and Evolutionary Computation (GECCO).

Anwar, A., & Franklin, S. (2003). Sparse distributed memory for "conscious" software agents. *Cognitive Systems Research*, *4*, 339–354.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baars, B. J. (1997). *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., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Science*, *7*, 166–172.

Baddeley, A., Conway, M., & Aggleton, J. (2001). *Episodic memory*. Oxford: Oxford University Press.

Baddeley, A. D. (1993). Working memory and conscious awareness. In A. Collins, S. Gathercole, M. Conway, & P. Morris (Eds.), *Theories of memory* (pp. 11–26). Hove, England: Lawrence Erlbaum.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science*, *4*, 417–423.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–609.

Beauchamp, M. R., Bray, S. R., & Albinson, J. G. (2002). Pre-competition imagery, self-efficacy and performance in collegiate golfers. *Journal of Sports Sciences*, *20*, 697–705.

Beekman, M., Calis, J. N. M., Oldroyd, B. P., & Ratnieks, F. L. W. (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, TN.

Bogner, M., Ramamurthy, U., & Franklin, S. (2000). "Consciousness" and conceptual learning in a socially situated agent. In K. Dautenhahn (Ed.), *Human cognition and social agent technology* (pp. 113–135). Amsterdam: Benjamins.

Chrobak, J. J., & Napier, T. C. (1992). Intraseptal bicuculline produces dose-and delay dependent working/episodic memory deficits in the rat. *Neuroscience*, *47*, 833–841.

Clayton, N. S., Griffiths, D. P., & Dickinson, A. (2000). Declarative and episodic-like memory in animals: Personal musings of a scrub jay or When did I hide that worm over there? In C. M. Heyes & L. Huber (Eds.), *The evolution of cognition* (pp. 273–288). Cambridge, MA: MIT Press.

Conway, M. A. (2001). Sensory-perceptual episodic memory and its context: autobiographical memory. *Philosophical Transactions of the Royal Society of London B*, *356*, 1375–1384.

Conway, M. A. (2002). Sensory-perceptual episodic memory and its context: Autobiographical memory. In A. Baddeley, M. Conway, & J. Aggleton (Eds.), *Episodic memory* (pp. 53–70). Oxford: Oxford University Press.

Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences of the USA*, *1001*, 8520–8525.

Donald, M. (2001). *A mind so rare*. New York: Norton.

Edelman, G. M. (1987). *Neural Darwinism*. New York: Basic Books.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

Fahle, M., & Daum, I. (2002). Perceptual learning in amnesia. *Neuropsychologia*, *40*, 1167–1172.

Franklin, S. (1997). Autonomous agents as embodied AI. *Epistemological Aspects of Embodied AI* [Special issue] *Cybernetics and Systems*, *28*, 499–520.

Franklin, S. (2000). Deliberation and voluntary action in "conscious" software agents. *Neural Network World*, *10*, 5–521.

Franklin, S. (2001a). Automating human information agents. In Z. Chen & L. C. Jain (Eds.), *Practical applications of intelligent agents* (pp. 27–54). Berlin: Springer-Verlag.

Franklin, S. (2001b). Conscious software: A computational view of mind. In V. Loia & S. Sessa (Eds.), *Soft computing agents: New trends for designing autonomous systems* (pp. 1–46). Berlin: Springer (Physica-Verlag).

Franklin, S., Baars, B. J., Ramamurthy, U., & Ventura, M. (2005). The role of consciousness in memory. In *Brains, minds and media* (Vol. 1) (urn:nbn:de:0009-3-1505).

Franklin, S., & Graesser, A. C. (1997). Is it an agent, or just a program? A taxonomy for autonomous agents. In *Intelligent agents III* (pp. 21–36). Berlin: Springer.

Franklin, S., & Graesser, A. C. (2001). Modeling cognition with software agents. In J. D. Moore & K. Stenning (Eds.), *CogSci2001: Proceedings of the 23rd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Franklin, S., Kelemen, A., & McCauley, L. (1998). IDA: A cognitive agent architecture. *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, 2646–2651.

Gabrieli, J. D., Milberg, W., Keane, M. M., & Corkin, S. (1990). Intact priming of patterns despite impaired memory. *Neuropsychologia*, *28*, 417–427.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*, 1–19.

Heinrich, B. (1984). Learning in invertebrates. In P. Marler & H. S. Terrace (Eds.), *The biology of learning* (pp. 135–47). Berlin: Springer.

Hofstadter, D. R., & Mitchell, M. (1994). The Copycat Project: A model of mental fluidity and analogy-making. In K. J. Holyoak & J. A. Barnden (Eds.), *Advances in connectionist and neural computation theory, Vol. 2: Logical connections* (pp. 31–112). Norwood NJ: Ablex.

Holland, J. H. (1986). A mathematical framework for studying learning in classifier systems. *Physica*, *22 D*, 307–317.

Jackson, J. V. (1987). Idea for a mind. *Siggart Newsletter*, *181*, 23–26.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology General*, *110*, 306–350.

James, W. (1890). *The principles of psychology*. Cambridge, MA: Harvard University Press.

James, W. (1892). *Psychology – briefer course*. New York: Henry Holt. New edition, New York: Harper & Row, 1961.

Kanerva, P. (1988). *Sparse Distributed Memory*. Cambridge, MA: MIT Press.

Kelley, J. L., & Magurran, A. E. (2003). Learned predator recognition and antipredator responses in fishes. *Fish and Fisheries*, *4*, 216–226.

Lhermitte, F. (1983). "Utilization behavior" and its relation to lesions of the frontal lobes. *Brain*, *106*, 237–55.

Maes, P. (1989). How to do the right thing. *Connection Science*, *1*, 291–323.

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).

Maturana, H. R. (1975). The organization of the living: A theory of the living organization. *International Journal of Man–Machine Studies*, *7*, 313–32.

Maturana, H. R., & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living*. Dordrecht, The Netherlands: Reidel.

Minsky, M. (1985). *The society of mind*. New York: Simon & Schuster.

Miyake, A., & Shah, P. (1999). *Models of working memory*. Cambridge: Cambridge University Press.

Morris, R. G. M. (2001). Episodic-like memory in animals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *356*, 1453–1465.

Negatu, A., & Franklin, S. (2002). An action selection mechanism for "conscious" software agents. *Cognitive Science Quarterly*, *2*, 363–386.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Olton, D. S., Becker, J. T., & Handelman, G. H. (1979). Hippocampus, space and memory. *Behavioral and Brain Sciences*, *2*, 313–365.

Ornstein, R. (1986). *Multimind*. Boston: Houghton Mifflin.

Panksepp, J. (1998). *Affective neuroscience*. Oxford: Oxford University Press.

Ramamurthy, U., D'Mello, S., & Franklin, S. (2003). *Modeling Memory Systems with Global Workspace Theory*, Seventh Conference of the Association for the Scientific Study of Consciousness (ASSC7), Memphis, May 30–June 2.

Ramamurthy, U., D'Mello, S., & Franklin, S. (2004, October). *Modified Sparse Distributed Memory as Transient Episodic Memory for Cognitive Software Agents*, IEEE International Conference on Systems, Man and Cybernetics (SMC2004), The Hague, The Netherlands.

Rugg, M. D., & Yonelinas, A. P. (2003). Human recognition memory: A cognitive neuroscience perspective. *Trends in Cognitive Science*, *7*, 313–319.

Schneider, W., & Chein, J. M. (2003). Controlled and automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science*, *27*, 525–559.

Sloman, A. (1999). What sort of architecture is required for a human-like agent? In M. Wooldridge & A. Rao (Eds.), *Foundations of rational agency*. Dordrecht, The Netherlands: Kluwer Academic.

Song, H., and Franklin, S. (2000). A behavior instantiation agent architecture. *Connection Science*, 12: 21–44.

Taylor, J. G. (1999). *The race for consciousness*. Cambridge, MA: MIT Press.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge, MA: MIT Press.

Watanabe, S., Kawahara, S., & Kirino, Y. (1998). Morphological characterization of the bursting and nonbursting neurones in the olfactory centre of the terrestrial slug Limax marginatus. *Journal of Experimental Biology*, *201*, 1851–1861.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*, 32–54.

Zhang, Z., Franklin, S., Olde, B., Wan, Y., & Graesser, A. (1998). *Natural Language Sensing for Autonomous Agents*. In Proceedings of IEEE International Joint Symposia on Intelligence Systems 98.

Zulandt Schneider, R. A., Huber, R., & Moore, P. A. (2001). Individual and status recognition in the crayfish, Orconectes rusticus: The effects of urine release on fight dynamics. *Behaviour*, *138*, 137–154.