2006

How the IDA model supports the Crick and Koch Framework

S. Franklin

How the IDA model supports the Crick and Koch Framework[1]

By Stan Franklin
Institute for Intelligent Systems
& Computer Science Division
The University of Memphis

Abstract

Crick and Koch recently introduced a Framework for Consciousness intended as a means of facilitating the explanation of the neural correlates of consciousness. The IDA model of global workspace theory is both a conceptual and computational model of consciousness, and of much of cognition as well. Here we systematically compare the ten aspects of the Crick and Koch Framework with hypotheses derived from the IDA model. On the seven aspects to which the IDA model relates there's a surprising amount of agreement. In addition the IDA model serves to flesh out the Framework with detailed hypotheses about the functionality of each of the seven relevant aspects, and often of its structure as well. The agreement of the Framework and the IDA model would tend to lend support to each.

The Framework

Recently Crick and Koch offered a "Framework for Consciousness" that they believe "offers a coherent scheme for explaining the neural correlates of (visual) consciousness" (2003). They explain that a framework must not be confused with a set of hypotheses. Rather a framework for consciousness offers a point of view from which to address the problems of consciousness. It's intended to guide research. A good framework should fit within current scientific knowledge reasonably well and should be roughly correct. It needn't be correct in all its details, but rather should guide research to fill in and correct it details. Such frameworks have proved useful in Biology and Physics. This one can be expected to be useful in consciousness studies. It consists of ten "aspects" each of which will be considered individually below.

Frameworks, by their very nature contain large gaps that can be filled with the results of later research. They can also be filled, temporarily, by hypotheses derived from cognitive models to give more specific guidance to research than does the framework by itself. In this case, hypotheses from the IDA model of consciousness and cognition can serve to clarify some aspects of this framework and to flesh out others.

The IDA Models

IDA is both a conceptual and a computational model of human cognition (Franklin 2000a, 2000c). The computational IDA is implemented as a software agent (Franklin & Graesser, A. C. 1997). This IDA "lives" on a computer system with connections to the Internet and various databases and does personnel work for the US Navy, performing all

[1] An earlier, briefer version of this paper appeared on line in *Science and Consciousness Review* as "A computer-based model of Crick and Koch's Framework for Consciousness" http://psych.pomona.edu/scr/editorials/20030301.html. 1 Mar 2003.

the specific personnel tasks of a human (Franklin 2001). In particular, IDA negotiates with sailors in natural language, deliberates, and makes voluntary action selections in the process of finding new jobs for sailors at the end of their current tour of duty. IDA senses strings of characters, and negotiates with sailors via email. The computational IDA is a running software agent that has been tested to the satisfaction of the Navy.

In addition to the computational model, we will also describe the richer IDA conceptual model, which models several cognitive faculties whose mechanisms have been designed, but not yet implemented. Among these faculties are automazation (Negatu et al. in review), problem solving, feelings and emotions (Franklin & McCaulley 2004), and several types of learning (Ramamurthy et al. 2004 to appear). In what follows, examples will often be drawn from the computational IDA model to help clarify the concepts. However, in the more general discussions, and in particular in the context of comparisons with the Crick and Koch Framework, an unspecified references to the IDA model should be interpreted as meaning the IDA conceptual model. Each cognitive faculty included in the IDA conceptual model is supported by an underlying computational mechanism, though sometimes unimplemented. Each design decision of such a mechanism give rise to a hypothesis concerning human cognition, that humans do it like the model. It's these hypotheses that constitute the scientifically useful output of the model. (More on this in the next section.)

Also note that software agent cognitive models differ in several ways from more traditional computational cognitive models. These more usual cognitive models typically produce data sets that can be compared with data sets produced by experiments. Thus their hypotheses can be tested quantitatively. Software agent models such as the IDA model, however, produce testable, but qualitative, hypotheses (Baars & Franklin 2003, Franklin et al. in review). Also, where traditional computational cognitive models are typically narrow in scope, software agent cognitive models like IDA must, by their nature, model a broad swath of cognition (Franklin & Graesser, A. C. 1997).

Based on Baars' global workspace theory of consciousness (Maes 1994, Drescher 1998, Baars 2002) the IDA model is also consistent with several other psychological theories (Glenberg 1997, Barsalou 1999, Sloman 1999, Conway 2002). The IDA model fleshes out global workspace theory so as to yield a fine-grained functional account of the steps involved in perception, several kinds of memory, consciousness, context setting, and action selection. We refer to this particular sequence of steps as a *cognitive cycle,* as described in some detail in the attached text box (Baars & Franklin 2003), in Figure 1, and in the next section. The reader is encouraged to read through, or at least skim, the cognitive cycle text box now, and to refer back to it as needed while reading the rest of this article. Though the description of the cognitive cycle in the text box contains examples from the computational IDA, the cognitive cycle, as described, also contains elements from the IDA conceptual model. More details on the various steps in the cycle can be found in the references. Codelets are described just below.

The IDA model includes a 'consciousness' module whose processes appear in steps 4 and 5 of the cognitive cycle. We consider the computational IDA to be *functionally conscious* in that the consciousness mechanism postulated in global workspace theory is fully implemented. This functionality is described in steps 4 and 5 of the cognitive cycle (see text box). We make no claim of phenomenal (subjective) consciousness for the IDA computational model (software agent), since we wouldn't know how to test such a claim (Franklin 2003).

As described in the text box, a number of cognitive processes play significant roles within a single cognitive cycle. Other cognitive processes such as deliberation (Franklin 2000b), problem solving, voluntary goal selection (Franklin 2000b), metacognition, etc. must take place over a number of such cycles. But, the mechanism for each of these multi-cyclic processes is built on a foundation of cognitive cycles. The IDA architecture comprises a unified theory of cognition in the sense of Newell (1990).

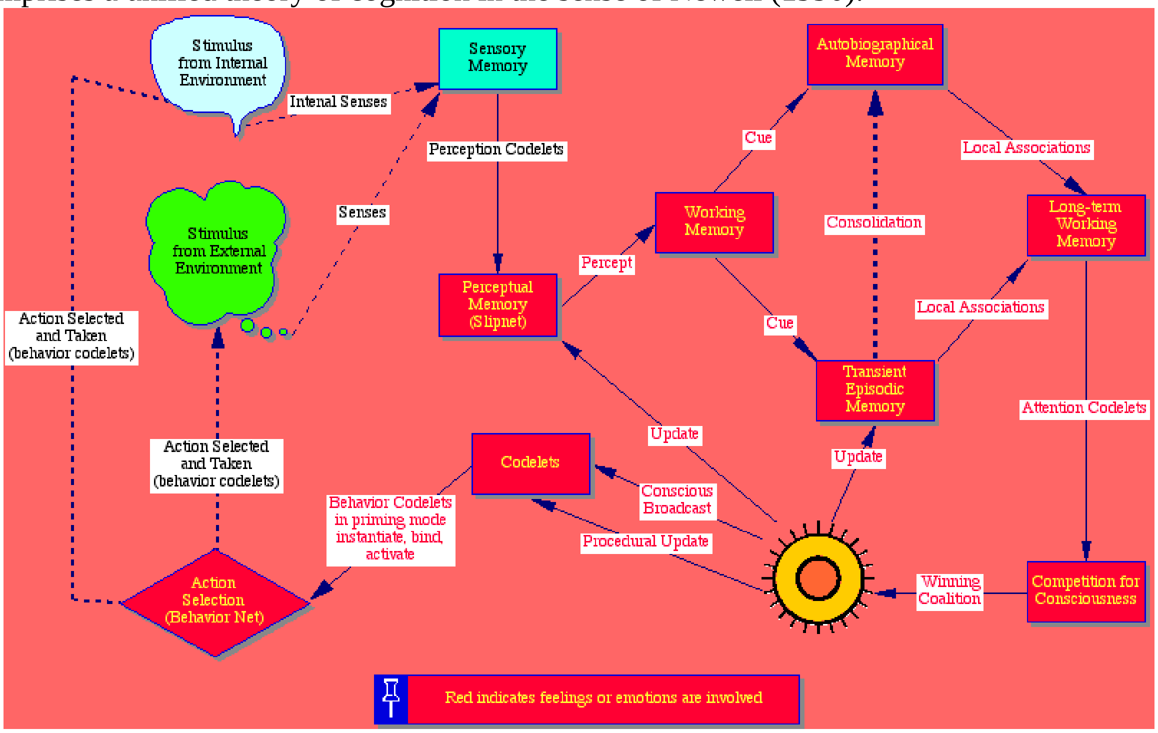

Figure 1. IDA's Cognitive Cycle

The Cognitive Cycle

The IDA model hypothesizes that cognitive processing in humans, as in IDA, consists of rapid, continually repeated, asynchronous traversals through these cognitive cycles. The cycles will cascade so as to act somewhat in parallel, but will carefully maintain the seriality of consciousness. A full cognitive cycle may take on the order of 100-200ms to run (Freeman 2003a). But due to overlapping and to automaticity, which shortens the cycle, up to a dozen cycles may be active during any given second. Freeman's work suggests that a change from one cycle to the next typically involves a hemisphere wide change of state in the brain over a time period of 5-10ms. Such sudden state changes are compatible with the conscious broadcast proposed by global workspace theory. This cognitive cycling hypothese implies that conscious contents come in temporally discrete steps, but rapidly enough to give the illusion of continuity (VanRullen & Koch 2003). Might consciousness also be spatially discrete? The need for a thalmocortical core suggests not (Edelman & Tononi 2000).

Several of the steps in the cognitive cycle refer to codelets. Computationally a codelet is a small piece of code, a little program that performs one specialized, simple task.

Codelets often play the role of demons[2] waiting for a particular type of situation to occur and then acting as per their specialization. Codelets in the IDA model implement the processors postulated by global workspace theory. They may also correspond to Edeleman's neuronal groups (1987), Minsky's agents (1985), Jackson's demons (1987) and to Ornstein's small minds (1986), The notion of cognition being organized as this kind of a multi-agent system is widespread.

As mentioned earlier, the role in science of any software agent model of cognition is to produce testable, qualitative hypotheses that can serve to guide the research of cognitive scientists and neuroscientists. The IDA model has produced a long list of such hypotheses (Baars & Franklin 2003, Franklin et al. in review) (see the text box also). The cognitive cycle hypothesis of five to ten cycles per second in humans allows an exceptionally fine-grained analysis of psychological phenomena (Baars & Franklin 2003, Franklin et al. in review). This fine-grained time span raises issues of testability. Are such fine-grained hypotheses actually testable? Some may be, but others are surely not given the current state of experimental technology. However, all are testable in principle, and may well guide experimentalists in developing new technology that would make them testable in practice.

The Framework and the IDA Model

In what follows the implications of IDA's cognitive cycling hypotheses for each of Crick and Kock's framework aspects will be discussed in turn. Quotes not otherwise identified will be from their paper (Crick & Koch 2003) and will be italicized. All in all, the IDA cognitive cycling hypotheses fits remarkably well into the Crick and Koch framework for consciousness, and serves to flesh it out with detailed functionality.

Keep in mind that the Crick and Koch framework is meant to guide research into the neural correlates of consciousness, and that it was developed with primate visual consciousness in mind. They also restrict their attention to processes occupying *"…time periods on the order of a few hundred milliseconds, or at the most, several seconds…"* This translates into one to a few tens of cognitive cycles.

The (unconscious?) homunculus: *"A good way to begin to consider the overall behavior of the cerebral cortex is to imagine that that the front of the brain is 'looking at' the sensory systems, most of which are at the back of the brain."* The IDA model would hypothesize that such looking and interacting happens over several consecutive cognitive cycles. Each individual cycle is largely, but by no means completely feedforward (see the text box). Crick and Koch propose that humans are only conscious of sensory representations of their thought, not of the thoughts themselves. The IDA model concurs for the most part. For IDA, as we see in Step 4 (competition for consciousness) and Step 5 (conscious broadcast) of the cognitive cycle, the contents of consciousness consist of the information carried by the member codelets of the winning coalition. These are for the most part sensory representations arising from the preconscious working memory buffers (Step 2) or from the local associations (Step 3). In addition, fringe attention codelets may bring fringe consciousness feelings or intuition to consciousness (Mangan 2001). Tip-of-the-tongue or déjà vu feelings are examples.

[2] This term comes from computer operating systems where such demons watch what's happening, and wait for an appropriate situation in which to act.

In the IDA model, most processing is carried out unconsciously as proposed in the Framework. For the most part IDA's planning, decision-making and creativity are all centered in her behavior net (Negatu & Franklin 2002), which isn't consciously accessible. The results of these processes may become conscious as the consequents of actions. IDA can make voluntary decisions (Franklin 2000b, Kondadadi & Franklin 2001). Though this process involves the behavior net and, thus, is partly unconscious, it is also partly conscious as it is in humans. The ideas that "pop into mind" are conscious, the mechanism for deciding between them is not. Humans sometime plan voluntarily, a capability seen in the computational IDA as she deliberates temporally about the suitability of a certain job. Again this is partly unconscious and partly conscious.

Zombie modes and consciousness: *"Many actions in response to sensory inputs are rapid, transient, stereotyped and unconscious. They could be thought of as cortical reflexes."* This aspect of the Framework draws attention to one of the innumerable gaps in the IDA model. Neither the computational model, nor even the conceptual model has, as yet, a facility for accommodating a zombie mode. Thus the IDA model adds nothing to this aspect.[3]

Coalitions of neurons: In this aspect of the Framework Crick and Koch speak of *"…competition among rivalrous cell assemblies…"* Step 4 of the cognitive cycle (competition for consciousness) describes the formation of such coalitions and the resulting competition for consciousness. Where Crick and Koch talk of competition among coalitions of neurons, the IDA model imposes more structure, coalitions of processors (Baars 1988) (codelets, cell assemblies). *"…at any moment, the winning coalition is somewhat sustained, and embodies what we are conscious of."* This occurs during Step 5 of the cycle (conscious broadcast.) According to the IDA model "embodies" should be interpreted as "provides the information" of which we become conscious.

"*Coalitions can vary both in size and in character. For example, a coalition produced by visual imagination (with one's eyes closed) may be less widespread than a coalition produced by a vivid and sustained visual input from the environment."* In the computational IDA model, the running software agent, much information is generated internally, say during deliberation, and is written to the preconscious working memory buffer. From there this information is available to be carried by information codelets into coalitions that compete for consciousness (Step 4). One of these coalitions often wins the competition. That is, internally generated information often becomes conscious (Step 5). Thus once again, the IDA model both supports and fleshes out the Framework.

In describing this aspect Crick and Koch go on to speak of *"…coalitions in the front of the cortex [that] may reflect feelings such as happiness and, perhaps, the feeling of 'authorship', which is related to free will."* In the IDA model, such feelings of happiness would originate in the preconscious perceptual module (the slipnet) whether engendered endogenously or exogenously.

[3] Since this writing, and partially as a result of it, plans are afoot to add a zombie mode called *perceptual automazation* to the IDA model. It would arise in much the same way as does behavioral automazation (Negatu et al. in review).

They go on to assert, *"There may be more than one coalition that achieves winning status and hence produces conscious experience."* Since IDA's coalitions may be somewhat larger and more structured that those envisioned by Crick and Koch, her coalitions may combine more than one of theirs. On the other hand, they may be proposing that more than one coalition can produce conscious experience during different cycles. The IDA model insists on exactly one coalition producing consciousness during a single cycle. Support for this generally accepted coherence property of consciousness comes introspectively from our inability to see both perspectives of a Necker cube simultaneously, and experimentally from numerous binocular rivalry studies. Of course, the IDA model may turn out to be simply wrong on this issue, and the quoted assertion just above may be shown to be correct as stated without further interpretation. In such a case we would have to repair the model.

Explicit representations: *"An explicit representation of a particular aspect of the visual scene implies that a small set of neurons exists that responds as a detector for that feature, without further complex neural processing."* In the IDA model the role of detector is played by perceptual codelets (Step 1a-early perception).

" if there are no such explicit neurons, or if they are all lost by brain damage, then the subject is unable to consciously perceive that aspect directly." Step 4 (competition for consciousness) describes the role of attention codelets and information codelets in the organization of coalitions of codelets representing some particular possible conscious experience. The loss of an attention codelet responsible for organizing some particular type of conscious event, or aspect of an event, would result in no further such events or aspects being experienced. (In the running example in the text box from the IDA software agent, the loss of a particular attention codelet could result in a sailor's preference for a particular location no longer coming to consciousness.) The loss of information codelets would have more particular consequences. At an earlier step in the cognitive cycle, (chunk perception) we describe the slipnet whose nodes extract meaning from incoming sensation. A missing node would result in certain types of stimuli not being recognized. Missing perception codelets from Step 1a (early perception) would again have more particular consequences. Thus the IDA model suggests that there are two distinct ways for losses of sensory consciousness such as achromatopsia to occur, one by defects in the preconscious perceptual system and the other from defects in the attention system. These would correspond to Crick and Koch's notion of the "back" and the "front" of the brain. Once again, there's good agreement between the Framework and the IDA model, with the later providing additional detail.

The higher level first: "*For a new visual input, the neural activity first travels rapidly and unconsciously up the visual hierarchy to a high level, possibly in the front of the brain... Signals then start to move backward down the hierarchy so that the first stages to reach consciousness are at the higher levels (showing the gist of the scene*." In the IDA model this quote describes the passing of activation among the nodes of the slipnet (perceptual memory) described in Step 1b (chunk perception). It also suggests that details of a particular visual stimulus may come to consciousness incrementally over several cognitive cycles. Though this doesn't occur in the IDA computational model (her sensing is too simple to require it), it is perfectly compatible with the IDA conceptual model.

Driving and modulating inputs: *"…connections to a cortical neuron fall roughly into two broad classes: driving and modulating inputs. For cortical pyramidal cells, driving* *inputs may largely contact the basal dendrites, whereas modulatory inputs include back-projections … or diffuse projections.*" There is nothing in the IDA model that sheds any additional light on this phenomenon.

Snapshots: *"We propose that conscious awareness … is a series of static snapshots… By this we mean that perception occurs in discrete epochs."* The IDA model makes precisely the same proposal with the contents of consciousness in Step 5 (conscious broadcast) of each cognitive cycle being one "snapshot." Others have also suggested that consciousness is discrete (VanRullen & Koch 2003).

"To reach consciousness, some (unspecified) neural activity for that feature has to cross a threshold…" The IDA model requires instead that a coalition of codelets win a competition with other such as described in Step 4 (competition for consciousness).

Crick and Koch go on to question *"What could be special about this activity that reaches above the consciousness threshold?"* In the IDA model only a single coalition can win the competition for consciousness. The winner is the most relevant, important, urgent coalition, that is, that coalition with the highest average activation among its constituent codelets. Perhaps, one day, this will be a testable hypothesis.

Attention and binding*: "Attention can usefully be divided into two forms: either rapid, saliency-driven and bottom-up or slower, volitionally controlled and top-down."* The IDA model implements volitional decision making via James' ideomotor theory. It does so using one particular behavior stream (goal context hierarchy) in the behavior net operating over several cognitive cycles (Franklin 2000b, Kondadadi & Franklin 2001, Franklin et al. in review). This includes voluntary decisions about what to attend to next. Saliency controlled attention (bringing to consciousness) occurs in single cognitive cycles. In the course of a conversation one may attend to both one's partner's voice and the movement of his or her lips during a single cognitive cycle. The computational IDA can attend to several distinct pieces of information from the same email message during a single cycle.

"Attention probably acts by biasing the competition among rival coalitions." Attention codelets do just that in Step 4 (competition for consciousness).

"…the 'binding' of the features of a single object/event is simply the co-membership in a particular coalition." This is equally true as stated in the IDA model in Step 4 (competition for consciousness).

Styles of firing*: "Synchronized firing (including various oscillations) may increase the effectiveness of a neuron, while not necessarily altering its average firing rate."* The IDA model has nothing to say about this aspect of the framework.

Penumbra and Meaning: *"The NCC [neural correlates of consciousness] at any one time will only directly involve a fraction of all pyramidal cells, but this firing will influence many neurons that are not part of the NCC. These we call the 'penumbra'… This penumbra includes past associations of NCC neurons, the expected consequences of the NCC, movements (or at least possible plans for movement) associated with NCC neurons, and so on."* The IDA model both supports and fleshes out this aspect of the Framework. Past associations are implemented in long-term working memory as described in Step 3 (local associations). Expectation is implemented by expectation codelets, a variety of attention codelet, as described in Step 6 (recruitment) and Step 9 (acting). Plans for movement are implemented as behavior streams (goal context hierarchies) as described in Step 7 (context setting) and Step 8 (action selection). Not only are all these features of this aspect of the Crick and Koch Framework present in the IDA model, but also relationships between them are explored.

Crick and Koch propose that "… *the penumbra neurons may be the site of unconscious priming."* The role in IDA of the preconscious buffer and of long-term working memory in priming is described in Step 4 (competition for consciousness).

Memory

The careful reader may note an apparent inconsistency between the several memory systems postulated as explicated parts of the IDA architecture, as opposed to memory, learning and knowledge in the framework being attributed to stable coalitions of cell assemblies. This apparent inconsistency is heightened by the fact that memory systems are known to be widely distributed over the brain (Squire 1987).

Memory in humans is quite complex. Declarative memory (autobiographical and semantic) (Baddeley et al. 2001) is known to be widely distributed over the neocortex. In the IDA model declarative memory is implemented using Sparse Distributed Memory (Kanerva 1988, Anwar & Franklin 2003), allowing for such distributed representations as stable cell assemblies.

The IDA architecture also models other human memory systems. Perceptual memory modeled in IDA using a slipnet (Hofstadter & Mitchell 1994). There are good arguments for a distinct neural mechanism for perceptual memory (Davachi et al. 2003, Freeman 2003b, Franklin et al. in review). Similarly, both working memory (Baddeley 2001) and transient episodic memory (Conway 2002, Shastri 2002, Franklin et al. in review) also have neural distinct mechanisms. Finally, procedural memory, the memory for behavioral skills, must certainly have its own distinct mechanism.

These various neural memory mechanisms all likely involve stable cell assemblies as proposed in the Crick and Koch framework. The apparent inconsistency results from a too narrow view of human memory, and from overlooking the possibility of distributed mechanisms being used to implement specific modules in an architecture. We conclude that there is no real inconsistency here between the specific memory modules contained in the IDA architecture, and the position of the framework regarding memory.

Hypotheses

The computational IDA, as a "conscious" software agent, is not a cognitive model in the sense of SOAR (Laird et al. 1987) or ACT-R (Anderson & Lebiere 1998) that simulate experiments in an attempt to reproduce their data. Rather the computational IDA is a running, embodied (Franklin 1997) software agent who performs a practical, real world task. The usefulness of both the computational and the conceptual IDA to cognitive scientists and neuroscientists lies in the many hypotheses generated by the IDA model that can be used to guide empirical research (Franklin 2000c, Franklin & Graesser, A. 2001, Baars & Franklin 2003, Franklin et al. in review). The IDA model's fleshing out of the Crick and Koch Framework serves to highlight several of these hypotheses, which we list in this section.

1. ***The Cognitive Cycle Hypotheses:*** Much of human cognition functions cyclically by means of continual interactions between consciousness, the various memory systems and the action selection mechanism. We call these cyclical interactions, as modeled in IDA, cognitive cycles. Such a cycle occurs

over a time scale of 100-200 ms. Cognitive cycles cascade, producing parallel actions. However, they preserve the seriality of consciousness.

2. ***The Discreteness Hypothesis:*** Consciousness occurs in a sequence of discrete episodes (VanRullen & Koch 2003) separated by quite short periods (~5 ms) of unconsciousness (Freeman 2003a). Both conscious and unconscious processing takes place in parallel during these episodes.
3. ***The Voluntary and Automatic Attention Hypothesis:*** Attention is the process of bringing contents to consciousness. Automatic attention occurs automatically and unconsciously during a single cognitive cycle. Attention may also occur consciously and voluntarily over multiple cycles in a consciously goal-directed way (Franklin 2000b).
4. ***The Coalition and Binding Hypothesis:*** During each cognitive cycle a single coalition sends its content to consciousness. Binding occurs both during a single cycle and incrementally over a number of consecutive cycles.
5. ***The Meaning Hypothesis:*** Meaning is assigned to both exogenous and endogenous stimuli early in the perceptual process of a single cycle. Perceptual memory plays the central role (Franklin et al. in review). Additions to this meaning come incrementally from local associations during the current and immediately subsequent cycles.

Many other more specific hypotheses can be mined from the text describing the relationship between the IDA model and the various aspects of the Framework. Still others can be found in each step of the description of the cognitive cycle given in the text box.

Conclusion

As our trip through the ten aspects of the Crick and Koch Framework has shown, there seems to be a remarkable concurrence between the aspects of the Framework and the various steps in IDA's cognitive cycle. Though the IDA model adds nothing to three of the aspects, in each of the other cases it provides relatively detailed hypotheses about the functionality of the aspect and often about its structure. This concurrence between the framework and the model lends credence to both.

There are, of course, differences as well. The IDA model represents cognitive structures at a much more abstract level than does the C&K framework. In Sloman's terms, the IDA model operates on a higher level virtual machine (Sloman & Chrisley 2003). This difference is to be expected, and accounts for the IDA model having nothing to add to a couple of the framework aspects. It doesn't account for one such omission, however. The lack of an IDA fleshing out of the zombie modes is just an gap in the IDA model, one we're in the process of filling.

One would think that the various aspects of the C&K framework would provide the basis for many testable, in principle, hypotheses concerning cognition. The same is true of the IDA model with the "in principle" emphasized. Unfortunately, our current techniques for studying cognitive phenomena at a fine grain, neural level, PET, fMRI, EEG, implanted electrodes, etc., are all lacking either in scope, in spatial resolution, or in temporal resolution. PET and fMRI have temporal resolution problems, EEG is well-known to have localizability difficulties, and implanted electrodes (in epileptic patients), while excellent in temporal and spatial resolution, can only sample a limited number of neurons, that is, is lacking in scope. As a result, many of our hypotheses, while testable in principle, are not testable at the present time for lack of technologies with suitable scope and resolution. (Einstein found himself in the same situation for several decades with his curvature of space hypothesis.) Nonetheless, the integrative nature of the IDA/GW model suggests that these hypotheses can will prove useful in helping to guide the research of cognitive scientists and neuroscientists interested in such, cognitive phenomena.

Acknowledgements

The author wishes to acknowledge the essential contributions to the IDA project of the many members of the "Conscious" Software Research Group and, particularly of the coauthors of relevant papers cited here. He is grateful to the Office of Naval Research for financial support of this project. Finally, sincere appreciation is due Christof Koch for suggesting improvements to an earlier draft, and to an anonymous reviewer for cogent comments that led a considerably improved depth.

References

Anderson, J. R., and C. Lebiere. 1998. *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Anwar, A., and S. Franklin. 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. 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. 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.

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

Conway, M. A. 2002. Sensory-perceptual episodic memory and its context: autobiographical memory. In *Episodic Memory*, ed. A. Baddeley, M. Conway, and J. Aggleton. Oxford: Oxford University Press.

Crick, F., and C. Koch. 2003. A framework for consciousness. *Nature Neuroscience* 6:119–126.

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.

Donald, M. 2001. *A Mind So Rare*. New York: Norton.

Drescher, G. L. 1998. Learning from Experience Without Prior Knowledge in a Complicated World. In *Proceedings of the AAAI Symposium on Parallel Models*. : AAAI Press.

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.
Franklin, S. 1997. Autonomous Agents as Embodied AI. *Cybernetics and Systems* 28:499–520.
Franklin, S. 2000a. A "Consciousness" Based Architecture for a Functioning Mind. In *Proceedings of the Symposium on Designing a Functioning Mind, Birmingham, England, April 2000*, ed. A. Sloman.
Franklin, S. 2000b. Deliberation and Voluntary Action in 'Conscious' Software Agents. *Neural Network World* 10:505–521.
Franklin, S. 2000c. Modeling Consciousness and Cognition in Software Agents. In *Proceedings of the Third International Conference on Cognitive Modeling, Groeningen, NL, March 2000,* ed. N. Taatgen. Veenendal, NL: Universal Press.
Franklin, S. 2001. Automating Human Information Agents. In *Practical Applications of Intelligent Agents*, ed. Z. Chen, and L. C. Jain. Berlin: Springer-Verlag.
Franklin, S. 2003. IDA: A Conscious Artifact? *Journal of Consciousness Studies* 10:47–66.
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., 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.
Franklin, S., and L. McCaulley. Feelings and Emotions as Motivators and Learning Facilitators. (Architectures for Modeling Emotions. AAAI Spring Symposia Technical Series Technical Reports; SS-04-02)
Franklin, S., B. J. Baars, U. Ramamurthy, and M. Ventura. in review. The Role of Consciousness in Memory. .
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. The Wave Packet: An Action Potential for the 21st Century. *Journal of Integrative Neuroscience* 2:3–30.
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.
Glenberg, A. M. 1997. What memory is for. *Behavioral and Brain Sciences* 20:1–19.
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.
Kanerva, P. 1988. *Sparse Distributed Memory*. Cambridge MA: The MIT Press.
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.

Maes, P. 1994. Agents that Reduce Work and Information Overload. *Communications of the ACM* 37:31–40.

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.

Meador, K. J., P. G. Ray, J. R. Echauz, D. W. Loring, and G. J. Vachtsevanos. 2002. Gamma coherence and conscious perception. *Neurology* 59:847–854.

Minsky, M. 1985. *The Society of Mind*. New York: Simon and Schuster.

Negatu, A., and S. Franklin. 2002. An action selection mechanism for 'conscious' software agents. *Cognitive Science Quarterly* 2:363–386.

Negatu, A., T. L. McCauley, and S. Franklin. in review. Automatization for Software Agents.

Newell, A. 1990. *Unified Theories of Cognition*. Cambridge MA: Harvard University Press.

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

Panksepp, J. 1998. *Affective Neuroscience*. Oxford: Oxford University Press.

Ramamurthy, U., S. D'Mello, and S. Franklin. 2004 to appear. Modified Sparse Distributed Memory as Transient Episodic Memory for Cognitive Software Agents. In *Proceedings of the International Conference on Sstems, Man and Cybernetics*. Piscataway, NJ: IEEE.

Rensink, R. A. 2000. The dynamic representation of scenes. *Visual Cognition* 7:17–42.

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.

Sloman, A. 1999. What Sort of Architecture is Required for a Human-like Agent? In *Foundations of Rational Agency*, ed. M. Wooldridge, and A. S. Rao. Dordrecht, Netherlands: Kluwer Academic Publishers.

Sloman, A., and R. Chrisley. 2003. Virtual Machines and Consciousness. *Journal of Consciousness Studies* 10:133–172.

Squire, L. R. 1987. *Memory an Brain*. Oxford: Oxford University Press.

Taylor, J. G. 1999. *The Race for Consciousness*. Cambridge, MA: MIT Press.

VanRullen, R., and C. Koch. 2003. Is perception discrete or continuous? *Trends in Cognitive Science* 7:207–213.and clarity of presentation.

[EDITOR-TEXT BOX]

The IDA Cognitive Cycle

1. PERCEPTION. Sensory stimuli, external or internal, are received and interpreted by perception assigning 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 the slipnet (Perceptual Memory), a semantic net with activation. (For example, a codelet finding the string `Norfolk` would activate a **Norfolk** node.) {Plausible neural correlates— Early visual and auditory cortex}
    b. CHUNK PERCEPTION: Activation passes from node to node in the slipnet. (for example from **Norfolk** to **Location**.) 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. (For example, **Preference for a Particular Location**.) {Plausable neural correlates— Visual/auditory object and word recognition areas of cortex, reentering widely via gamma coherence. Perirhinal Cortex (Davachi et al. 2003)}

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. These buffers in humans may involve visuospatial, phonological, and other kinds of information. (For example, in the computational IDA, a percept might include a nine-digit number tagged as social security number or a text string tagged as a location, or the recognition of a stated preference for a particular location.) {Plausible neural correlates— Temporo-parietal and frontal lobes (Baddeley 2001)}

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 (Taylor 1999, Donald 2001, Conway 2002 page 137) and from long-term declarative memory. The contents of the preconscious buffers together with the retrieved local associations form transient episodic memory and long-term associative memory, together, roughly correspond to Ericsson and Kintsch's long-term working memory (Ericsson & Kintsch 1995) and to Baddeley's episodic buffer (Baddeley 1993). {Plausible neural correlates—Hippocampal System (Panksepp 1998 page 129Shastri 2001, 2002)}

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. (For example, in the computational IDA, an attention codelet may gather into a coalition an information codelet carrying the name and rate `AS1 Kevin Adams`, another carrying the location `Norfolk`, and yet another the idea `Preference for a Particular Location).` 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 aren't picked up by any attention codelet. {Plausible neural correlates— First stable re-entrant organization of perceptual and immediate association areas.}

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. (For example, IDA may become "conscious" of AS1 Kevin Adams' preference for being stationed in Norfolk.) This broadcast is hypothesized to correspond to phenomenal consciousness[4].

   The current contents of consciousness also determine reinforcement to procedural, serve to update perceptual memory, and are 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 (Shastri 2001, 2002). Transient episodic memory is an associative memory with a relatively fast decay rate (Conway 2002, Ramamurthy et al. 2004 to appear). It is to be distinguished from autobiographical memory, a part of long-term declarative memory.[5] {Plausible neural correlates— Correlated firing from sensory projection areas to target areas: parietal, frontolimbic and medial-temporal cortex, hippocampus and basal ganglia. Gamma coherence (Meador et al. 2002). Thalamocortical core (Edelman & Tononi 2000)}

[4] We make no claim that IDA is phenomenally conscious.

[5] Christof Koch (personal correspondence) pointed out the similarity of the contents of consciousness derived from a "metastable neuronal coalition" (of codelets) to proto-objects from Coherence Theory (Rensink 2000). He went on to ask whether conscious contents, such as street images streaming past while driving, when newly written to transient episodic memory also tend to quickly dissipate according to the IDA model. They certainly quickly dissipate from consciousness. Something that comes streaming by may come to consciousness only during a very few cycles and so make only a slight impression on transient episodic memory. It may decay away (dissipate) during the next very short time period (Franklin et al. in review).

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.{Plausible neural correlates— Re-entry between target areas and sensory cortex}

7. SETTING GOAL CONTEXT HIERARCHY. Some responding behavior codelets instantiate an appropriate behavior stream, if a suitable one is not already in place. Using information from the conscious broadcast they also bind variables and send activation to behaviors. (In our running example, a behavior stream to find jobs to offer Kevin Adams might be instantiated. His preference for Norfolk might be bound to a variable.) 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[6]. {Plausible neural correlates—Frontolimbic re-entrant processing to prepare action}

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 goals), by the current situation, external and internal conditions, by the relationship between the behaviors, and by the activation values of various behaviors. (In our example IDA would likely choose to begin extracting useful data from Kevin Adam's personnel record in the Navy's database.) {Plausible neural correlates—prefrontal cortex (Fuster et al. 2000)}

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. {Plausible neural correlates—Motor efference from motor/premotor cortex.

[END TEXT-BOX]

[6] Though part of the IDA conceptual model, the material on non-routine problem solving has not yet been published.