CS-TR-3232 March 18, 1994

Consciousness and Complexity: The Cognitive Quest

Don Perlis

Institute for Advanced Computer Studies
Computer Science Department
AV Williams Bldg
University of Maryland
College Park, MD 20742

perlis@cs.umd.edu

Abstract

Some implications of the view that mind is a suitably complex kind of process are investigated in various contexts. The underlying theme is that the behavior of complex systems cannot be adequately judged by that of simple systems. I first present a personal exploration of the mechanistic account of mind in terms of non-technical considerations; then I present and criticize some ideas of Kripke, Nagel, and Jackson that challenge the mechanistic view. Next I turn to a brief synopsis of some of Dennett's recent ideas. Finally I offer some critical comments on Dennett's views and suggest possible modifications.

Note: This paper will appear in a Festschrift in honor of Professor Jack Minker, as a special issue of the Annals of Mathematics and Artificial Intelligence.

Consciousness and Complexity: the Cognitive Quest*

Donald Perlis
Department of Computer Science
University of Maryland
College Park, MD 20742
(301) 405-2685
perlis@cs.umd.edu

Abstract

Some implications of the view that mind is a suitably complex kind of process are investigated in various contexts. The underlying theme is that the behavior of complex systems cannot be adequately judged by that of simple systems. I first present a personal exploration of the mechanistic account of mind in terms of non-technical considerations; then I present and criticize some ideas of Kripke, Nagel, and Jackson that challenge the mechanistic view. Next I turn to a brief synopsis of some of Dennett's recent ideas. Finally I offer some critical comments on Dennett's views and suggest possible modifications.

1 Introduction

My topic is the thesis that the mind is a machine, or at least that it consists of certain aspects of the working of a suitable machine (such as the brain).[1] Among other things,

*It is a great pleasure to me, to be able to dedicate this essay to Jack Minker in celebration of his 65th birthday and of the many contributions he has made in research, in human rights, and in education (of students and colleagues alike). May he continue with us at the University of Maryland for many more productive years! I would like to say, on a personal note, that in my early years here as a naive floundering young assistant professor (well, not so young, but certainly naive and floundering) I was—in Jack's colorful phrasing—a "dead duck"; until, that is, Jack took me under his wing and educated me in how to carry on a vigorous research program. He got me interested in circumscription; co-authored papers with me; showed me how to isolate open problems; write grants; take bad reviews constructively; and eventually he weaned me into a live duck! Jack: thanks, very much.

[1] We can try to make this a little more concrete by characterizing a machine as a complex of simple parts interacting according to simple laws. Then mind, if it is a machine, arises from the complexities of the particular overall architecture of the interacting parts.

I will discuss some ideas in the important new book by Daniel Dennett, *Consciousness Explained* [1991]. To glance ahead, Dennett suggests that consciousness should be regarded as a special kind of report, what we might call a narrative self-history, produced out of a highly complex, parallel and fluctuating conglomerate of processes.[2] I begin with some very general thoughts on the cognitive science enterprise, and then move to some objections, addressing in particular ideas of Kripke, Nagel, and Jackson; and then I point out some aspects of Dennett's approach.

2 The cognitive quest: some general considerations

Will we ever understand the mind? Can the mind possibly be intelligible in terms of physical processes, as a kind of machine? I will discuss some specific claims and counterclaims in the following sections. First, I want to explore uninhibitedly some thoughts related to the prospect of *understanding* ourselves and in so doing changing our very concept of ourselves.

We may get to the point of knowing enough about ourselves to be able to make major changes in what we are, we may control our own future, our evolution. We already do this in small ways today, e.g., by body-building, by education, by medicine. But we may someday be able to choose much more profound aspects of our being, and, even more exciting, we may thereby come to see entire ways of being that today totally elude us! We may find that we are merely a rather minor form of consciousness in a vast spectrum of possibilities waiting for us to grow into them.

The dull robots of today's factories are a very poor image for this prospect, and it invites confusion to use the same words (robot, computer, machine) for these as for the future scenario of ourselves I am suggesting. But in another sense the words are appropriate, just as chemistry is an appropriate term for describing the basis for life. It is chemistry, but a rather fantastic version of it not suspected until rather recently, one that puts 19th century chemistry to shame as a dull dead affair. In the same sense I think that the possibilities for machines will turn out to be even more dramatic and we will see that we are not only chemical in nature but also mechanical—yet this will not detract from our wonderfulness at all, any more than chemistry has. We will find that we are wonderful soft machines, and this will attest to the amazing qualities machines can have: they can be as wonderful as us: they *are* us! It is just that so far we have not come to realize the incredible range of things machines can be, we have at present only rather pitiful examples, just as people of the past had only pitiful examples of what chemistry could be.

One aspect of this is the study of complex systems, something we have little

[2] I submit that we have an excellent example in our guest of honor, Professor Jack Minker. For he is clearly conscious, bafflingly complex, and possesses a distinguished 65-year narrative self-history!

experience with so far. Few would contest that the mind is a highly complex system. Yet we know little of how complex systems work; from a description of the (workings of the) parts an understanding of the whole is not at all immediate.

It may be fair, then, to characterize artificial intelligence as a branch of the general science of complex systems. Yet complex systems, as we are discovering in biology and now in so-called *chaotic* systems, are far richer than we had previously suspected. The mind may well prove to be the most complex (and correspondingly rich) of all.

3 Some technical objections considered

Standard objections to mind-as-machine typically fail to take account of the implicit notion that any such machine will be incredibly complex, and consequently may have features that cannot be foreseen merely by analogy with comparatively simple systems. This is what we shall look into in the remainder of this paper.

3.1 Pains and brains

Toward the end of his book *Naming and Necessity* [1980], Saul Kripke presents an argument against the mind-brain identity thesis, i.e., he argues that brain process in itself cannot be mind, nor in particular can it constitute experiential phenomena such as pain. I will try to summarize what I think is the gist of his argument, calling attention to features that are not completely explicit in his presentation.

Let us suppose that an experiential phenomenon, say a felt pain, were in fact nothing more than a kind of brain process. Now, the brain is largely composed of neurons, which either fire or do not at any given time. Suppose that a particular set of neurons in an organism's brain—the "pain" neurons—are the ones whose firings amount precisely to the phenomenon of felt pain, as experienced by that organism.[3] But the set of pain neurons could in principle fire in isolation, detached from a whole brain or anything else, in which case there'd be no feeler to experience the pain, hence there is no felt pain. Contradiction.[4]

But this argument has two hypotheses: a general one (the identity thesis) and a specific one (about a specific set of "pain" neurons). The latter one is the one being refuted, and the former is left for us to feel suspicious of. Let us restate Kripke's

[3]Kripke speaks of C-fibers. In fact, C-fibers are not in the brain at all but are slow-carriers (along with fast carriers, A-delta-fibers) of 'pain signals' from portions of the body to the spinal cord. The spinal cord relays these signals to the brain, especially to the thalamus region, where some have speculated that actual pain sensation may occur. The thalamus of course is made of fibers too, including billions of neurons. See Dennett [1978] for philosophically weird facts about the pain mechanism and Nichols et al [1992] for some more technical background.

[4]Kripke says (p. 153–154): "to make the C-fiber stimulation correspond to [felt] pain, God must do something in addition to the mere creation of C-fiber stimulation...the stimulation could exist without the pain."

argument in terms suitable to our analysis. There are two implicit assumptions we need to draw out.

First, the argument tacitly assumes the set of pain neurons is a rather small portion of the brain, so that it is implausible that they alone constitute a kind of brain able to have experiences. That is, their firing must be a very isolable phenomenon; and we can agree I think that no mere dozen or hundred neurons will suffice to constitute a kind of personhood able to feel pain.[5]

Second, the argument assumes that if felt pains are brain states, then they are firings of certain fibers, since the brain is made up of fibers (neurons and other tissue). But Kripke shows that this (together with the first assumption) is incoherent, and concludes then that brain states cannot be pain-feeling.

However, it may be that pain is not local to a small set of fibers, and also it may be that no particular set of fiber-firings constitute pain; it may be temporal patterns of neuronal behavior, rather than any particular set of neurons. Such patterns could be at any level of organization at or above the neuronal level. It may be—and Kripke's argument seems to lend support to this—that pain and other mental states can only occur at a very high organizational level, maybe even involving most of the brain, and that at this level pain is not isolable, i.e., it cannot happen apart from the personality that it is an aspect of. The feeling of pain may be a process over time, not a single instantaneous battery of firings, and constituting the recognition of something as happening to one's self. For an analogy, we cannot isolate the surface tension from the ocean, nor the primality of 5 from the natural numbers. Pain may not be so much "where" or "which" neurons fire, as "how" they fire. Certain variations in the firing patterns across billions of neurons constituting our consciousness may constitute the presence of pain in that consciousness.[6]

This is precisely the complex-systems contention. A highly complex system may have all sorts of unexpected properties, and we should not assume in advance that there are no surprises in store. But there is a valuable insight to be gained from Kripke's argument: Pain (and mind in general) cannot be a small system.[7]

Maybe the whole brain does constitute a feeler, a conscious agent. Indeed that is what the identity thesis claims. In fact, by not recognizing the essential assumption

[5]The previous footnoted quote from Kripke I think clearly indicates that he takes the stimulation of pain neurons to be far less than the full normal stimulation of the entire brain. For simply re-read the quote with "full normal brain stimulation" replacing "C-fiber stimulation"; more on this below.

[6]To speak of those variations isolated from underlying patterns is meaningless. And to deny that the whole brain's neuronal firing patterns can be consciousness is simply to beg the question of the identity thesis rather than to make an argument against it.

[7]This is not to say that there is no localization of function in the brain. Far from it: many processes are highly localized, but apparently those that constitute consciousness are less so. Thus much early (low level) visual processing occurs in various stages between the retina, lateral geniculate nucleus, and back of the brain (area V1). But these local processes do not in themselves constitute sight: an individual having only those portions of brain (and the brainstem, necessary to support autonomic functions) will not be conscious at all, let alone have visual awareness.

of *smallness* in his fiber-argument, Kripke does leap to the conclusion that no brain structure at all can constitute mind. But his argument itself, without this unjustified leap, in no way undermines the identity thesis. Rather it gives us new information about what must follow from it: that mind (and feeling) is a global property of the brain, and highly local events in themselves cannot constitute conscious (cognitive or mental) events. (This is not to say that the entire brain must be involved in any one feeling; but it does say that a sufficiently large and complex part to produce a personal subject is needed.)

The tendency to make a claim about a system as a whole, based only on an individual accounting of isolated—local—parts, leads often to an extrapolation fallacy: assuming that what can be seen to hold about small numbers of parts must also be seen about large numbers of parts. Complex systems may admit higher-order phenomena not readily described in lower-level terms.

What then about the thalamus and pain-fibers in the brain? According to our analysis, the actual experience of pain—felt pain—can occur only as a state of a feeling being. Thus the thalamus in and of itself in a certain state cannot be what pain is unless the thalamus itself constitutes a feeling agent, something no one at this point seems prepared to say. However, the thalamus is constituted of billions of fibers (neurons), so that this is not out of the question.

3.2 Brain, knowledge, and experience

Thomas Nagel and Frank Jackson have argued that science does not provide us with certain kinds of information, namely knowledge of what subjective experience is like. Nagel argues, for instance, that all our detailed examinations of a bat's brain will still leave us in the dark as to what it is like to be a bat.

He may be right, in that a bat's brain is very complex, far too complex for an unaided human brain to keep track of all that is going on; there will probably be very many important patterns of activity that we cannot see, even in principle, without sophisticated detection and memory aids. But perhaps with such aids, which may not yet have been invented, we would indeed be able, in effect, to relive the bat's experiences and thus come to know what it is like to be a bat. Nagel wants to deny this.

Now suppose, as Nagel permits us to do, that we find out what bat feelings are, in ordinary scientific terms. Then we do not necessarily know what those feelings feel like. But I want to argue that we might, the long way around. It may be a little like empathizing with an unusual character in a well-written novel, where the author has to go to some lengths to draw the appropriate picture for us.[8] We of course have to be able to follow it, and this may tax our memory and attention and so on, to the point that we may need mechanical memory aids. It may even, in the case of

[8]An anonymous referee has suggested novels by men about women, and *vice versa*, as related examples.

bats, require us to use batty memories temporarily, to be able to hold onto the large number (suppose) of sounds all at once.

This is not cheating. The same holds for many things we understand or know. Cloud formation, for instance, is only understood by us in a vague and approximate way, but not because of any profound ontological features; rather, simply because clouds are so complex as to defy detailed analysis by us. So we use simplifying mathematical assumptions instead, and point to that. If we want to grasp the process of cloud formation as it really is, we would need to enlarge our short-term memories enormously to allow us to account for each droplet of water vapor all at once, and all higher-level cloud structures too. Swirls of mist, pressure gradients, and so on, are real enough, but hard to think of without special training. Maybe we need to formalize a very fancy language for the study of complex systems, something a bit like the powerful set theories of mathematical logic.

Scientific understanding is a function of how rich, how expressive a language the scientists use, or are able to use. Perhaps all "elementary" events can be described in terms of addition and subtraction alone, and thus in some very real sense a science based on these operations alone would be able in principle to describe everything. Yet such a science would leave out any kind of direct description of multiplication, prime numbers, logarithms, calculus, clouds, wave motion, non-linearity; these concepts are at a higher level of description, requiring a richer, higher-order language. Indeed, scientific progress is often progress in language, in conceptual vocabulary. And no doubt our brains have limits here. But this need not be an in-principle limitation.

A related illustration is visio-numerical apprehension ("subitization"). Most people can judge the number of a pile of up to four objects without explicit counting. The visual system seems to do this automatically. But only a very few can do it for 10 or more objects. What does it feel like to have the latter ability? We may someday understand in total detail just what the ability is, in terms of brain structure. But this will not necessarily let us (mere 4-apprehenders) have the feeling.

This sounds to be in support of Nagel. In part, yes. I claim that a typical fact about feeling like something (a bat, or a 10-apprehender) is very complex, too complex to grasp in its entirety without special purpose equipment. That is, understanding is itself constrained by the brain's hardware. We rely on approximate or fuzzy versions of complex things when we try to understand them. To feel like a bat one needs the relevant aspects of a bat's hardware. But this is not to say that science cannot give us this knowledge (this feeling). It can, if we are prepared to use crutches to aid our brains. And why not? Why should we view science as limited to what the unaided human brain can accomplish?

Consider an example due to Frank Jackson: Mary lives in a black-&-white room and sees no colors but has access to complete scientific knowledge, including knowledge about wavelengths of light and about people's eyes and brains outside her room, and can figure out all facts (so the story goes) expressible in the language of science. Jackson claims that Mary will not know what it is like to see color. But this precisely begs the question, for how are we to know that *all the facts* do not include color sensation facts?—that is, if we do not limit Mary to just *elementary* facts, rather than high-level conceptual constructs built out of those facts on the basis of her (assumed perfect) ability to conceptualize, abstract, synthesize, etc.

Perhaps a suitably detailed account of the neurophysiology and cognitive phenomenology of seeing red, would produce in Mary the experience of seeing red the long way around [Perlis 1978]: instead of it coming to her through her eyes and visual cortex it might come through other avenues, but might be color experience nonetheless.

3.3 No way to say it

Here is a perhaps more familiar way to think about the previous section. It also leads into a consideration of Dennett's work in the next section.

Sometimes one hears statements to the effect that a given expression E in, say, English, simply cannot be said in French; there's no word or even concept in French that corresponds to the content of E. This is probably false, at least if we are broad enough in our perspective. For natural languages are not static things, in fact they contain their own means for enlargements, new words, new concepts. And it takes a competent speaker time to adapt to such changes. But upon successful adaptation, the new concept no longer seems top-heavy with definitional baggage. What might begin in French as a very long explanation of E, including verbal and ostensive examples, ends up—after much practice—as a simple F: it has been assimilated. Then F itself is part of the material the speaker can draw on for making reports to others—or indeed to herself.

Much the same sort of thing might be the case for Mary, for subitization, and for bat-mimicry. After much top-heavy explanation, one might start to form the appropriate connections to eventually get it "right." Or, one might not. The view propounded here is that this will depend solely on the mental material, the brain capacity to form those connections. That is, it will depend on the ability to process a sufficiently rich structure of interrelated facts of the matter, into a new conceptual structure that one can draw on. It does not depend on access to a non-scientific, non-rational realm.

Of course, to argue like this is not to prove, nor even to amass evidence. At best, it is to provide encouragement to look further, harder, in this, the decade of the brain, and on into the coming century of the mind.

4 Dennett's contribution

It is impossible here to adequately summarize Dennett's enormous contribution in his new book, *Consciousness Explained.* He incorporates in one coherent view—and painstakingly analyzes—a wealth of psychological, philosophical, and computational data. I will just single out a few highlights.

One key datum Dennett cites is the variation on apparent-motion experiments due to Kolers and von Grunau. Here a subject is presented first with, say, a red light for 150 milliseconds and then after 50 milliseconds of no light, with a second green light, say to the left of the first light. The subject reports seeing not two lights flashing in succession but one moving light, which changes color halfway along its motion from right to left. On being questioned, the subject insists on not having seen stationary flashing lights.

Does the subject first experience a stationary light and then re-experience it as moving—forgetting the original experience—or simply experience it as moving from the first? Dennett calls these alternative explanations Orwellian and Stalinesque, respectively:

Orwellian: Event E is experienced and then all memory of E is replaced by a false revision E′, so that the original E is no longer part of the agent's awareness or recall.

Stalinesque: Event E occurs but before it can be experienced, E′ interferes and is experienced instead. This false revision is the only one that reaches consciousness.

He argues persuasively that no ingenious variation on the experiment could distinguish between these competing explanations. He then urges us to accept the view that there is no deeper matter of fact about this: that experience is not a real phenomenon apart from the reports the experimenter captures from the subject. This approach he calls *heterophenomenological* methodology, which I paraphrase loosely here: we should focus on what subjects report, and take that as an operational definition of experience. Don't ask about something "real" underlying the report, for if we do all we will find is a sea of multiple drafts of reports in progress, with no distinguished one that has any claim to being more conscious than others.[9] There is no single real train of thought, at least not until an actual external report is fashioned. Until this report is done, there are merely competing processes.

On the other hand, he also speaks of a self-model the subject has, and hints that on the basis of such the subject may internally report her own narrative history, what we might call a "narrative self-history". Indeed, this is what the experimenter's report often sounds like, and how we often describe ourselves: as embodying single thought-streams.[10] Moreover, Dennett argues that out of the sea of multiple drafts

[9] In this, Dennett echoes Minsky [1985].

[10] On self-models, it is of interest that the cortex of the human (mammal) brain has a "sensori-motor" strip, consisting of roughly one billion neurons, that is hardwired to much of the body's muscles and skin, in just two of several relatively permanent bodily self-maps. Thus it appears

of work-in-progress, arises a *virtual mind* (in the computer sense), but that where to place the boundaries on the mind (what is mind and what is not, even within an individual's manifold drafts) is largely a matter of no real import: you can draw the line where you like, and there are no (and can be no) empirical grounds for choosing.

I think that although he has come up with some wonderful, powerful, and illuminating tools for the study of the mind, he goes a bit too far in this latter claim of arbitrariness of boundary. For the very narrative self-history that he holds up as the hallmark of the mental seems to me to offer some real possibility of empirical grounding. What if neuroscientists discover, for instance, a brain mechanism that forms a narrative self-history, and then determine that in the apparent-motion experiments in fact there is a "fixed flashing light" report fully formed in the narrative self-history before the second light flashes, only to be erased totally after the second flash. This is indeed distinct from the alternate scenario in which no such initial report arrives in the narrative self-history and the first such is the one with a moving light. The first scenario supports the Orwellian position, the second the Stalinesque. (A most pleasing outcome might be the Stalinesque one, in a discovery that it takes more time to form a narrative self-history than can be done in the duration of an Orwellian preamble; then at least we would not be fooled about our own experiences.) All that suffers in this account is the very heterophenomenological methodology that Dennett has introduced. It may give way to neuroscience. And once neuroscientists start to find structures and processes along the lines of narrative self-histories, I think we will be well on our way to a robust science of cognition.

Perhaps the narrative self-history itself might appropriately form the boundary of consciousness—Dennett almost says as much himself, but then pulls away with the disclaimer that the self is after all only a fiction. But what if it is not a fiction? What if there really is a process that does exactly what the narrative self-history is supposed to do? (We are talking computations now, not qualia.) And then either a stationary flashing light gets into the history or it does not. There need not be multiple drafts at this level. There is a final draft, namely the one that, in ideal conditions, would become part of the heterophenomenological record.

Indeed, water waves and operating systems are not merely useful fictions, they are real, even though they are not readily described in "reduced" terms (water molecules or bits). So why should narrative self-reports be any less real? If the brain "really" provides such, then it is real.

Suppose agent g is attached to a "cerebroscope" that reveals whatever scientists want to know about the workings of g's brain. Suppose a process "NSH" is found in g's brain, that plays the role of narrative self-history: it monitors g's behavior in many ways, it is accessed by other processes in g's brain that then give rise to speech production, etc. Suppose further that we have worked out a reliable intentionality for elements in NSH, so that we can associate external content with those elements. Now

---

that the brain is built in such a way as to treat the self as special, not merely another part of the environment. See Perlis [1978,87,90,91, and forthcoming] for uses of self-models in AI.

we see on the cerebroscope that an element with content P has just formed in NSH. We ask g what she was just thinking, and she tells us "P"—or at least something we recognize as expressing the content P. At some point, we might start to simply trust the cerebroscope's reporting on the status of the NSH, and stop questioning the agent in addition. In fact, if the agent is for some reason unable to report to us in any customary way, so that we must rely on the cerebroscope, mightn't we be fairly comfortable in the assessment that g nevertheless really did think "P"?—especially if some time later g regains speech and assures us that she had indeed been thinking just that!

The virtual informational level of the narrative self-report is very real indeed. We do not delude ourselves—as Dennett suggests—that we have minds. We really do have minds, and Dennett has taken important steps to show us how.

In several later papers Dennett (e.g., [1993a]) elaborates further, along lines perhaps closer to the view espoused here, and offers the following helpful analogy: there is no precise point in time when the British Empire first knew of the signing of the treaty in the War of 1812. The analogy is that an empire is a bit like a mind: made up of multiple and conflicting parts and tendencies. Thus parts of the Empire knew of the treaty before other parts; it is not clear what would count as *The Empire* "really" coming to know something. In the same way, it may not be clear what counts as "really" seeing a red dot, as opposed to processing red-dot retinal inputs in numerous ways. Yet it seems that over time—and on occasion—even an Empire can get its collective act together and arrive at a collective salience or consensus. Might not a similar neural salience occur in the brain and amount to what we call consciousness? Dennett [1993b] calls such a salience "cerebral celebrity" but still disparages the thought that this can have a precise temporal boundary.

It may not be useful to quibble over milliseconds of precision in a boundary. Maybe there is an interim period of, say, 50 milliseconds during which an NSH record is in the process of forming, and whether or not we regard it as truly in conscious experience in that period is moot: maybe we can only report such an interim experience in cases where we have it afterwards, consistent with both Orwellian and Stalinesque interpretations. But maybe not; maybe clever cerebroscopic technique can tell the difference. Nevertheless, precise boundary or not, NSH may constitute real consciousness.

A caveat: our lesson from Kripke suggests that NSH cannot be confined to any small piece of brain. It is better thought of as a process involving large portions of brain: it is the constantly self-updating structure of internal cognition-reports, though whatever neural form that may take is still very much unknown. It is beyond the pandemonium[11] of multiple drafts, and in the realm where, for instance, decisions can be made; the latter require commitment by the whole organism and thus have the requisite salience.

[11] See Selfridge [1959]; this is an early precursor of the multiple draft idea.

5 Conclusions and a formal approach

Has Dennett solved the problem of mind? No, certainly no more than the early biologists who began to take a chemical basis for heredity seriously had solved the problem of life. But they did discover a crucially important route to investigate, which did lead substantially to a solution (still being worked out today). Perhaps Dennett has helped us toward a similarly important key regarding complex systems. And just as there turned out to be highly complex molecules with utterly unsuspected properties (e.g., self-reproduction), there may turn out to be highly complex self-reporting systems, with properties we still do not appreciate.

In the attempt to explain the phenomenon of life, the very concept evolved and eventually gave way to a rather different view involving peculiar and previously unsuspected molecules with properties no chemist had dreamed of prior to 1920 or so. These molecules allowed a new conception of the biological realm, one in which the usual notions of living things (as involving heredity, reproduction, survival, metabolism, development, and so on) retained important roles but now as consequences of the new, deeper view. And as a result it became less compelling to look for a definitive definition of the term "life" since it was seen that this term was too vague to do justice to the complex and detailed new view that was unfolding.

Similarly, the notions of mind and consciousness may give way to a deeper view, in which familiar cognitive attributes (feelings, thoughts, desires, self, etc.) may still retain roles but derivative of (and less precise than) the new view. It is not clear that this is what Dennett thinks. Nevertheless, I think Dennett may well have taken us a key step along the route toward a robust science of cognition. Perhaps his suggestion of mind as a virtual self composed of a narrative report evolving out of multiple "drafts" of information processing, when explored with the sharp tools of neuroscience as well as those of cognitive psychology, will reveal new and exciting worlds as yet undreamed of. Although others have made related suggestions, Dennett has been the first to offer a sustained and coherent analysis of such a view, and tied it firmly to research findings in the various cognitive sciences.

So, where are we with respect to complex systems? We now have computers with thousands of processors. But the human brain has upwards of 100 billion neurons—some even say as many as one trillion! Even the humble praying mantis has on the order of 100,000 neurons.[12] Moreover, mind does not come in small isolable parts; nor is it describable in isolable terms. It is a whole, in the sense of a complex system, and we so far have only limited experience in understanding such systems. But computer systems do provide some experience here, especially in the notion of a virtual level of

[12]This varies according to the species of mantis, and according to sex: mantis society is highly sexually-dimorphic, and female-dominated. The mantis has sophisticated visual, motor and auditory systems, allowing her to stalk prey and avoid predators. It is of interest that the female mantis is cannibalistic: she finds mantises tasty; and yet she never munches on herself, indicating a robust self-model. Whether the mantis brain has sensori-motor self-maps as do mammals is not known, but it is a good guess.

explanation. (See [Cole 1991] for more on virtual systems and mind.)

In sum, many of the arguments against the mind-brain complex-systems thesis fail to take proper account of the unexpected behaviors that may occur in a complex system. Indeed, there seems to be no way to take account of these behaviors except by actually discovering them: by doing the hard work of figuring out the details. However, plausibility arguments can be given, and Dennett has made a valiant and very worthwhile effort here.

I have avoided use of a plethora of "isms" in the above discussion, but it might be worth pointing out that they include some traditional ones and some new ones; dualism, isolationism, reductionism, indescriptivism, eliminativism, virtualism, and fictionalism are some of the themes we have seen here. But I think that a proper appreciation of the wonders of complex systems will show that good old reductionist physicalism/realism is as strong as ever. In fact, Dennett's fictionalism I think really paves the way for a revitalized brain realism. But it will be a virtual or emergent brain realism, a new kid on the block. Exciting times lie ahead!

What can we learn from this? How can it help us reach an understanding of cognition, from below (neuroscience) or above (formal commonsense reasoning)? One obvious suggestion is to look for narrative self-histories, whether in the brain or in formal programs. The latter effort is already at least implicit in non-monotonic reasoning [Ginsberg 1987], where a kind of meta-inference is made to the effect that, since $\neg P$ has not been proven, then we conclude as a default that P. Here the observation or inference that we have not proven $\neg P$ is a kind of introspection of self in a very rarified sense: the "self" is simply the logic.

Recent work [Elgot-Drapkin and Perlis 1990, Kraus et al 1990, Miller and Perlis 1993] has pursued this idea further, by introducing "active" time into formalisms, to allow for the evolution of an explicit self-history. This in turn has allowed us to solve problems not even expressible in classical logics.[13] In [Miller 1993] particular active- (or step-) logics are identified in which an inconsistency appearing at one step can be dissipated at a later step. Ongoing work involves a "network" of information that gives rise to beliefs but itself is not constituted of beliefs, reminiscent of Dennett's multiple drafts. The long-range goal of this effort is the creation of an embodied inference mechanism, with an ongoing narrative self-history that it can reason about, including its own body and actions. The fact that its very reasoning adds to the history—and therefore increases time as measured in that history—is key, and is the chief distinguishing feature of active-logics to date.

[13] That is, if the theorems of the logic are identified with the beliefs of the reasoning agent being modeled.

REFERENCES

- Cole, D. [1991] Artificial intelligence and personal identity. *Synthese*, 88, 399–417.
- Dennett, D. [1978] Why you can't make a computer that feels pain. In D. Dennett, *Brainstorms.* Bradford Books.
- Dennett, D. [1991] *Consciousness Explained.* Little, Brown.
- Dennett, D. [1993a] Living on the edge. *Inquiry*, 36, 135–59.
- Dennett, D. [1993b] The message is: there is no medium. *Philosophy and Phenomenological Research*, 53, 889–931.
- Elgot-Drapkin, J., and Perlis, D. [1990] Reasoning situated in time I: basic concepts. *J. of Experimental and Theoretical Artificial Intelligence*, 2, 75–98.
- Ginsberg, M. [1987] *Readings in Nonmonotonic Reasoning.* Morgan-Kaufmann.
- Jackson, F. [1982] Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–136.
- Kraus, S., Nirkhe, M., and Perlis, D. [1990] Planning and acting in deadline situations. *AAAI-90 Workshop on Planning in Complex Domains.*
- Kripke, S. [1980] *Naming and Necessity.* Harvard.
- Miller, M. [1993] *A view of one's past and other aspects of reasoned change in belief.* PhD dissertation, Dept of Computer Science, Univ of Maryland, College Park MD.
- Miller, M. and Perlis, D. [1993] Presentations and this and that: logic in action. *Proceedings, Cognitive Science Conference, Boulder CO.*
- Minsky, M. [1985] *The Society of Mind.* Simon and Schuster.
- Nagel, T. [1974] What is it like to be a bat? *Philosophical Review*, 83, 435–50.
- Nichols, J., Martin, A., and Wallace, B. [1992] *From Neuron to Brain*, 3rd edition. Sinauer Associates.
- Perlis, D. [1978] Consciousness as account-taking. Manuscript, U of Rochester, CS Dept.
- Perlis, D. [1987] How can a program mean? IJCAI-87, Milan.
- Perlis, D. [1990] Intentionality and defaults. *Intl J. of Expert Systems*, 3, 345–354.
- Perlis, D. [1991] Putting one's foot in one's head–Part I: Why. *Noûs*, 25, 435–455.

- Perlis, D. [forthcoming] Putting one's foot in one's head–Part II: How. To appear in E. Dietrich (ed.), *From Thinking Machines to Virtual Persons: Essays on the Intentionality of Computers*. Academic Press.
- Selfridge, O. [1959] Pandemonium: a paradigm for learning. *Symposium on the Mechanization of Thought Processes*, London: HM Stationery Office.