



International Phenomenological Society
Philosophy and Phenomenological Research

The Virtues of Virtual Machines

Author(s): Shannon Densmore and Daniel Dennett

Source: *Philosophy and Phenomenological Research*, Vol. 59, No. 3 (Sep., 1999), pp. 747-761

Published by: [International Phenomenological Society](#)

Stable URL: <http://www.jstor.org/stable/2653793>

Accessed: 23-02-2016 16:38 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Wiley, International Phenomenological Society and *Philosophy and Phenomenological Research* are collaborating with JSTOR to digitize, preserve and extend access to *Philosophy and Phenomenological Research*.

<http://www.jstor.org>

The Virtues of Virtual Machines¹

SHANNON DENSMORE AND DANIEL DENNETT
Tufts University

Paul Churchland's book, *The Engine of Reason, The Seat of the Soul* (hereafter ER), is an entertaining and instructive advertisement for a "neuro-computational" vision of how the brain (and mind) works. While we agree with its general thrust, and commend its lucid pedagogy on a host of difficult topics, we note that such pedagogy often exploits artificially heightened contrast, and sometimes the result is a misleading caricature instead of a helpful simplification. In particular, Churchland is eager to contrast the explanation of consciousness that can be accomplished by his "aspiring new structural and dynamic cognitive prototype: recurrent PDP networks" (p. 266) with what strikes him as the retrograde introduction by Dennett of a virtual von Neumannesque machine—a "failed prototype"—as the key element in an explanation of human consciousness (in *Consciousness Explained*, 1991, hereafter, CE). We will try to show that by oversimplifying Dennett's alternative, he has taken a potential supplement to his own view—a much needed supplement—and transformed it in his imagination into a subversive threat. In part 1, we will expose and correct the mistaken contrasts. In part 2, we will compare the performance of the two views on Churchland's list of seven features of consciousness any theory must account for, showing that Dennett's account provides more than Churchland has recognized, and indeed offers answers to key questions that Churchland's account is powerless to address. At that point, Churchland's project and Dennett's could be seen to collaborate in a useful division of labor instead of being in mortal combat, were it not for what appears to be a fairly major disagreement about consciousness in non-human animals. Part 3 briefly examines this issue. It may be due to a misunderstanding, which when cleared up might restore the happy prospect of unification.

¹ This essay began as a seminar paper by the first author; it has been emended and amended by the second author, mainly to fit the format of this symposium, but also expanding on some of the points made in the original version.

1. Recurrent PDP networks and Virtual Machines

What are the prerequisites for conscious experience? In physicalist circles everyone agrees that what is called for is a brain of complex design. That much is plain. Disagreement ensues over just what sort of design turns a brain into a mind. In Churchland's view the mind/brain's design is that of a recurrent PDP network of a rather particular sort (to be described in due course). In contrast, Dennett argues that such networks may be adequate models of the brain's design (at a fairly high level), but they are insufficient to account for the mind. In his view, the mind is like a software program that is installed upon the parallel neural network of the brain.

Churchland's explains Dennett's idea that consciousness is a virtual machine as follows:

For example, we can both produce and understand the complex strings of symbols of a language; we can perform deductive operations on such strings with some facility and reliability; we can do recursive arithmetic operations such as addition, multiplication, division and so forth. When we do such things, according to Dennett, our underlying parallel neural architecture is realizing a 'virtual' computing machine, whose activities are now of the classical, discrete-state, rule-governed, serial kind (p. 264).

This is not so. The only feature of the von Neumann architecture that Dennett imputes to the "von Neumannesque" virtual machine he is discussing is its seriality, and even this feature is heavily qualified in the Multiple Drafts Model. And Dennett explicitly denies that this virtual machine is a "classical, discrete-state, rule-governed" machine. As Dennett puts it, "a virtual machine is a temporary set of highly structured regularities imposed on the underlying hardware by a program: a structured recipe..." but he goes on to insist that "In a von Neumann machine, you just 'load' the program off a disk into the main memory and the computer thereby gets an instant set of new habits; with brains, it takes *training*, This is, of course, a major disanalogy." (p. 218–19) This training, Dennett says (in terms one might expect Churchland to applaud), "is accomplished, we can surmise, by thousands or millions or billions of connection-strength settings between neurons, which all together in concert give the underlying hardware a new set of macrohabits, a new set of conditional regularities of behavior." (p. 218). These are not "rule-governed" or "discrete-state":

in place of the precise, systematic 'fetch-execute cycle' or 'instruction cycle' that brings each new instruction to the instruction register to be executed [in a classical von Neumann machine], we should look for imperfectly marshaled, somewhat wandering, far-from-logical transition 'rules,' where the brain's largely innate penchant for 'free association' is provided with longish association-chains to more or less ensure that the right sequences get tried out (p. 225).

How did Churchland miss these qualifications? Perhaps because he does include in his list of examples cited above some vivid instances of a person becoming, for the nonce, a classical virtual machine: “addition, multiplication, division”. Suppose that Paul Churchland is engaged in an act of long division at noon. This is manifestly a serial, rule-governed, discrete-state activity, so there is indeed one level of explanation of the noontime phenomenon that treats Churchland as a hand-simulation of a classical virtual machine: the *long division machine*. We may agree, however, that his brain is still a massively parallel recurrent PDP machine at noon, engaged in all manner of vector completion and so forth, and his stream of consciousness at noon may include not just “27 goes into 94 thrice; three times 27 is 81, which subtracted from 94 leaves 18, bring down the 6 ...” but also various digressive musings, nagging itches, fleeting scraps of sexual fantasy, and who knows what else. Dennett would say that Churchland’s recurrent PDP machine is generating a (non-classical, non-rule-governed, non-discrete-state) von Neumannesque virtual machine which in turn is generating, more or less continuously (depending on how hard Churchland is concentrating), another virtual machine, a hand-simulation of the classical, discrete-state, rule-governed long-division machine. If you want to explain why Churchland gets the answers to the long division problems right so often, why certain problems take him longer than others, and why his pencil-pushing behavior produces the patterns of marks on the paper that it does, then the level to which you must ascend to explain is the level at which he is hand simulating the long-division machine. If instead what you want to explain are some other regularities in his behavior, such as his humming or whistling while he works, or his periodic lapses into dreamy-eyed grinning, or his muttered subvocalizations, then, according to Dennett, you had best descend to a somewhat lower level, but not—if you actually want to explain these patterns—all the way down to the level of the recurrent PDP networks, which are at least one level too low into the trees to permit us to see the woods we are interested in.

Churchland, on the other hand, sees the recurrent PDP network level as providing all the explanatory power needed (p. 266). In a trivial sense, that could be true, but in the same trivial sense, a computer scientist could insist that the von Neumann instruction cycle is the only level needed to explain all the phenomena exhibited by today’s computers. Churchland claims to account for the various contributions of Aristotle, Descartes, Newton and Einstein by saying they were “using their recurrent pathways” (p. 278). Of course they were. And in the same spirit, when Microsoft Word, Lotus 1-2-3, Myst, and Netscape do their different sorts of magic for you, they are just “using their instruction cycles” in the CPU of your laptop. The micro-code building-blocks of a von Neumann CPU can be assembled into indefinitely rich and competent higher-level systems, but if we insist on couching all our explana-

tions at the level of the building blocks, we won't explain what needs explaining. A parallel problem for Churchland becomes clear when he says that creative folks are "unusually skilled at such recurrent manipulation" (p. 279). We are entitled to ask: manipulation *by whom*? The manipulation of recurrent pathways is not a familiar, accessible, folk-psychological category of mental activity; our creativity may indeed be underwritten by "skillful" manipulation of recurrent pathways, but this is not something *we do* in the way we can frame a mental image on command, or conjure up a memory of an old flame, or count backwards silently from one hundred. Who or what is playing the role of the manipulator in Churchland's account? Not a homunculus, surely, but what, then?

Churchland needs a way of cashing out such manipulation-talk in terms of higher-level patterns of activity. What he needs is virtual demons, or something else at that intermediate level. In fact, his vivid impressionistic accounts of how scientists, moralists and others come to their insights via the manipulation of their prototypes might serve as "pseudo-code"—approximate first specifications of just the virtual machines he needs! He might try, as a relatively simple warm-up exercise, to design the system of manipulations required in his own brain to turn it temporarily into the classical long-division machine. (Contrast this proposed project in connectionist modeling with another: just training up a network to give correct answers to some unenlargeable subset of long division problems. The proposed project would be to train up some recurrent PDP networks to pursue the very paths of calculation revealed in Churchland's own protocols when he hand simulates the long-division machine. Saying it can be done is one thing—of course it can, in principle; doing it *without invoking at least one higher virtual machine level* is quite another.)

Churchland's desire to distance himself from the "classical" tradition in AI also tempts him to overstate the case for the "nonalgorithmicity" of his alternative. In rejecting Roger Penrose's vision, he says:

One need not look so far afield as the quantum realm to find a rich domain of nonalgorithmic processes. The processes taking place within a hardware [emphasis added] neural network are typically nonalgorithmic, and they constitute the bulk of the computational activity going on inside our heads. They are nonalgorithmic in the blunt sense that they do not consist in a series of discrete physical states serially traversed under the instructions of a stored set of symbol-manipulating rules (pp. 247–48).

Notice the insertion of the word "hardware" here. Without it, what Churchland says would be false. In fact all the results he discusses—NETTalk, Jeff Elman's grammar-learning networks, Cottrell and Metcalfe's EMPATH, and all the others—were produced not by "hardware neural networks" but by virtual neural networks simulated on von Neumann machines. And so, at a low level, every one of these demonstrations *did* "consist in a series of dis-

crete physical states serially traversed under the instructions of a stored set of symbol-manipulating rules.” This is not the level at which to explain their power, of course, but it is an algorithmic level, and nothing these programs do transcends the limits of Turing computability. Now it is unlikely in the extreme that Churchland would want to claim that the very models he discusses so favorably *fail* to exhibit the powers that he thinks are crucial to the explanation of mentality—because they are not themselves nonalgorithmic. But then his claim that *hardware* neural networks are nonalgorithmic, even if true, would not play any role in explaining the powers they exhibit—since algorithmic approximations thereof have all the necessary powers. Churchland would do better to join Dennett in the conclusion that the powers of virtual machines, whether they are virtual parallel recurrent PDP networks realized on von Neumann machines, or virtual von Neumannesque machines realized on parallel recurrent PDP networks, are best explained at the virtual machine level.

2. The Magnificent Seven

Churchland compares his own account of consciousness to Dennett’s by assaying their respective performance on a benchmark catalogue of seven features of consciousness:

- (1) it displays steerable attention;
- (2) it is independent of sensory inputs;
- (3) it has the capacity for alternative interpretations;
- (4) it involves short-term memory;
- (5) it disappears in deep sleep;
- (6) it reappears in dreaming; and
- (7) it harbors the contents of the several basic sensory modalities within a single unified experience (ER, p. 213–14).

It is important to note that Churchland’s claim is not that Dennett cannot account for these features, but rather that he does not do so (ER, p. 269). Since it is Churchland’s list, that might only reveal a difference in emphasis. But in any event, let’s consider Churchland’s claims.

1. *Steerable Attention*

Churchland and Dennett agree that the ability to direct one’s attention to particular facets of the environment to the exclusion of others is a fundamental feature of conscious experience. Churchland dubs this ability “steerable attention” and Dennett argues that it is an essential means of control. Since the brain’s function is to “produce future,” Dennett has much to say about the

origin and design of its control powers. In his evolutionary account, steerable attention arises out of a need for internally driven higher level control. “[W]ith increased functional plasticity, and increased availability of ‘centralized’ information from all the various specialists, the problem of what to do next spawned a meta-problem: what to *think about* next.” (CE, p. 188) The solution to the “what to think about next?” problem originated, Dennett claims, with autostimulation, and the crowning stages of the human solution to this meta-problem arose out of language, which permitted talking to oneself, diagramming to oneself, and other self-manipulations from which indefinitely many further systems of representation have developed.

Dennett claims that language permits our human brains to be parasitized by cultural units called memes, and some of these well-designed cultural products, when they “infect” a brain, can be seen as the installation of a (more or less) serial software program.² (The long-division machine is just a particularly simple, serial, rule-governed example of such a meme-machine. The game of hide-and-seek, the maxim “Look before you leap!”, and the very idea of talking to yourself much of the time are examples of memes that are less rigid sorts of installable cultural software.) Unlike our ancestors who were at the mercy of coalitions of possibly archaic (evolution-designed) “specialists” vying for control in the parallel architecture of the brain, our internally driven methods of autostimulation set new agendas for the meta-problem. The result, in effect, is the successive nomination of coalitions of specialists, including experts imported (with all their expertise) from the culture, not provided as part of our biological heritage. Just how is this solution to the problem of what to think about next supposed to account for steerable attention? Steerable attention emerges as the ability to set the agenda for the brain. The ability to track particular features of the world comes with the development of systems of representation. The ability to focus in on and rank representations depends upon the methods and results of autostimulation. And the ability to plan around a feature of the world is a function of the various serial programs installed by memes.

To see this aspect of Dennett’s theory in action let’s apply it to Churchland’s example of the third baseman. Churchland writes, “[t]he third baseman focuses on the batter’s swing, determined to recognize immediately and accurately how and where the ball will be launched back into the infield.”³ The thought processes going through the third baseman’s mind we might imagine to be something like the following: “O.K. it’s the eighth inning, there’s one man on first base, we’re up by three, no outs, Jones, who is renowned for bunting, is up at bat, ...” All of this information (and much more) participates in the process of anticipating just what the batter is going

² Dennett, p. 209–26.

³ Churchland, p. 217.

to do and what the third baseman should do in response. "If he bunts, throw to second base. If he bats it into left field, turn around and get ready to receive it from the left fielder. If he bats it into right field, the right fielder may send it my way depending upon how far the guy on first gets around the bases." In order for the third baseman to perform these tasks, he must be able to represent these features of his environment one way or another. In addition, he must be able to manipulate and rank his thought processes in order to focus on particular things to the exclusion of others, and he must also be able to process all of the relations between the stage of the game, what the batter might do and what he should do. A conscientious novice typically accomplishes this by saying the words to himself, even out loud, while in a professional baseball player, perhaps all of this sequential processing gets done "automatically" and without internal speeches of self-admonition. We may suppose that it was at one time painstaking and deliberately worked through, but once he has mastered the range of options associated with his position, and overlearned his habits, he has, in effect, a software baseball-playing program stored in his brain. Subsequently, the routine is performed effortlessly, aided mainly by autostimulation to keep him focused and on track for the duration of the game.

For Churchland steerable attention is to be understood as the pre-activation of a particular prototype vector. In recurrent networks the nodes are arranged so that the information at hidden layer nodes cycles back to the nodes that precede them. The effect is a bias in the input layers that favors the activation of a particular prototype vector. If the input layers receive information that fits into or is related to the pre-selected vector, then that vector will be activated. Although Churchland's account appears to be relatively straightforward when compared to Dennett's account, it leaves out an explanation of just how pre-activation occurs. That is, Churchland doesn't tell us what the operative mechanism is. How is it that the desired prototype is activated instead of one that is irrelevant to the situation? In a parallel distributed system how does one prototype get selected to the exclusion of others? In Dennett's terms, how is the meta-problem of what to think about next solved? These may not be insurmountable problems for Churchland, although there is wide agreement that this is a problem area for connectionist theories (in contrast with classical AI theories, which have invested heavily in models of planning). If a pure connectionist network is to be a viable alternative to classical approaches, we need an account or at least an inkling of how the higher-level control problems are to be solved.

2. Independence of Sensory inputs

Churchland argues that consciousness is independent of sensory inputs. What this means is that in a recurrent network, information can travel from the

nodes of the hidden layers to nodes closer to the input layer by means of descending pathways. The activation vectors generated in this way are internally precipitated and thus independent of sensory inputs. Since these vectors are not direct results of sensory inputs, he recommends that they be described as daydreams or fantasies (ER, p. 217). Dennett's claim that all mental activity is a matter of "interpretation and elaboration of *sensory inputs* [emphasis added]" (CE, p. 111) may look at first like a denial of Churchland's claim that consciousness is independent of sensory inputs. To see why this is the wrong way to view his statement we need to review some features of the Multiple Drafts Model and clarify what Churchland is asserting when he makes his claim.

On Dennett's account, daydreams and fantasies fall out of two features of the brain's parallel hardware: first, pandemonium style processing, and second, the fact that information is under continuous "editorial" revision. Dennett agrees with Churchland that the brain is a parallel distributed network; when information comes in through the senses it is processed and revised for an indefinite amount of time by a myriad of nodes or specialists (CE, p. 111). We do not experience the editorial processes themselves, which continue indefinitely, letting various "drafts" dominate for awhile, only to be succeeded by other drafts. What thus dominates, and hence is actually experienced, will in some cases be so distantly and indirectly related to long-past sensory inputs as to be practically independent of sensation. The balance between endogenous "epistemic hunger" and current sensory satisfiers of that hunger can swing between pure hallucination and dreaming at one extreme (CE, pp. 10–16, and see below for more details), through deliberate self-stimulation by mental imagery of various sorts (CE, pp. 285–303), to—at the other extreme—veridical, "data-driven" perception.

For Churchland, the computations performed at the hidden layers are a product of training on sensory inputs. Thus there is a sense in which the information that travels from these nodes to the input layer nodes is dependent upon sensory inputs—without the initial training there simply is no information to be processed. Thus, the issue is not whether consciousness is independent of sensory inputs altogether, but whether sensory inputs are necessary for consciousness at particular times. In Dennett's account consciousness is independent in the sense that the editorial process can continue indefinitely; because this process continues indefinitely sensory inputs are not necessary for conscious experience at particular times. Thus the two views are in agreement on this point.

3. Alternative Interpretations

The complement to steerable attention, Churchland says, is the capacity for alternative interpretations, especially of ambiguous data. Instead of keeping a

frame of mind constant in the face of changing inputs as was the task for steerable attention, alternative interpretation involves changing the frame of mind while keeping the input constant. In Churchland's view, this feature of consciousness is unique to recurrent networks—at least if contrasted only to feedforward networks; recurrent networks have descending pathways that circulate information to earlier processing stages. The descending pathways have the effect of biasing earlier stages of processing so that an ambiguous input triggers one vector rather than another. If the information received via the descending pathways had been different it might have caused an alternative interpretation of the input. Since the biasing information is provided internally, the network's capacity for alternative interpretations depends largely upon its antecedent cognitive state, collateral information and its educational background.

It would be surprising if Dennett, author of the Multiple Drafts Model of consciousness didn't offer an account of the capacity of consciousness to entertain alternative interpretations, and in fact he accepts and endorses Churchland's assessment of the factors that contribute to it: "[f]or instance, a discrimination of a picture of a dog might create a 'perceptual set'—making it temporarily easier to see dogs (or even just animals) in other pictures— or it might activate a particular semantic domain, making it temporarily more likely that you read the word 'bark' as a sound, not a covering for tree trunks" (CE, p. 135). On his view, how an ambiguous input is interpreted will depend upon which "specialists" or pattern recognition mechanisms are in control or dominant at a particular time, and on the nature of the information they have most recently processed. The question of whether this domination is accomplished by the specific mechanism of recurrent PDP networks is not addressed by Dennett, but nothing he says conflicts with the claim.

4. Short-term Memory

In Churchland's view the process that allows for steerable attention and alternative interpretation is also responsible for short-term memory. That process, you will recall, involves information travelling via descending pathways to earlier stages of processing. In the case of short-term memory the information provided by the hidden nodes was input information that has been processed by the intermediate nodes between the input layer and the hidden layer. Thus the input layer receives information not only from the senses but also from higher layers of processing. In Churchland's words, "[r]ecurrent pathways thus sustain a rudimentary form of short-term memory. They make the creature's immediate cognitive past continually available to it for processing together with incoming sensory information about the present. Information that passed by layer 2 just a split second ago can be brought back to layer 2, usually in modified form, to be added into the current mix." (ER, p. 100)

Before we take this as an account of the kind of short-term memory we are familiar with and rely upon every day, consider just how short this kind of short-term memory is. Churchland says, “[t]his allows the creature to represent its current situation in a way that takes into account the situation that *immediately* [emphasis added] preceded it.” It is important to remember that the kind of short-term memory this account describes is memory that “extends at least a few fractions of a second into the Extended Past”. When we talk about short-term memory, we usually have in mind something of somewhat greater duration: like a telephone number we have just looked up, or where we parked the car at the mall, or even what we ate for breakfast this morning—things that we will not likely remember a day or a week from now but will remember for a minute or maybe a few hours. There are a host of relatively short-term memory phenomena well studied by psychologists, ranging from the iconic sensory memory that last only a few milliseconds and cannot be rehearsed, to the feats of mnemonists using elaborate virtual machines of memory-enhancement. Churchland’s account does not (yet) provide any account of most of these forms of short-term memory. He is surely right that recurrence within neural networks provides a fundamental architectural feature underlying at least some of these memory phenomena, but his perspective is so limited by his attention to this low-level phenomenon that it is unclear how he could expand upon it to account for the variety of short-term memory phenomena. It is probably wise to hold off declaring that he has an account of short-term memory.⁴

Does Dennett’s account of short-term memory fare any better? Consider how he would handle the case of recalling where one parked the car, for instance. The information about where the car is parked is duly embodied, we may suppose with Churchland, in the connection strengths of recurrent PDP networks, but now it must be retrieved. Confronting the parking lot, one may *ask oneself* “where did I park the car?,” thereby manipulating (as Churchland himself would say) the relevant recurrent pathways. By such techniques of autostimulation, one may succeed in precipitating what Dennett calls a probe, retrieving the right piece of knowledge. Or one may not; nothing may come to mind (as one says) when one asks oneself. Scanning the parking lot visually may then be the next strategy of autostimulation, allowing external

⁴ Ramsey, Stich and Garon (1991) discuss a second problem about memory and connectionist networks that Churchland has hardly addressed: the reliance by networks upon repetitive training for long-term memory. As they point out, it is difficult to see how the human ability to hear something once and commit it to memory is to be accounted for in connectionist theories. Dennett (1991) raises still further problems, noting that many feats of human memory (such as our ability to answer “directly” and “without thinking” such questions as “Have you ever danced with a movie star?”) demand an explanation in terms of meta-levels of inference based on recognition of the failure of memory to respond to the question with a recollected instance. Once again, these are phenomena that call for higher levels of explanation.

cues to boost the process of retrieving stored information. Sometimes one cannot remember where one parked the car, try as one might. In Dennett's view, these cases and others like them fall into several different categories. They may be instances in which knowledge retrieval strategies don't work at the right time, or they may be instances in which the sought information has decayed or been overwritten by later contents.

5, 6, and 7. Deep Sleep, Dreaming, and The Unity of Consciousness

Our experience of the world is unified—sights, sounds and smells are experienced together, not in isolation—and yet the information that makes for experience is obtained by five distinct sensory modalities. In order to explain this phenomenon, Churchland argues that there must be an information “bottleneck” (ER, p. 215), a place where all the information comes together. He claims that the intralaminar nucleus of the thalamus is the bottleneck and that the representations in that recurrent network are polymodal; the information obtained by each of the five senses has a unique form in which it arrives as input at the intralaminar nuclei, there to be processed as a multimodal representation. Moreover, studies show that bilateral damage to the intralaminar nuclei produces an irreversible coma (ER, p. 221), so Churchland decides, plausibly, that their activity is necessary for consciousness. He then accounts for the absence of consciousness in deep sleep and its reappearance in dreaming by citing the role of intralaminar nuclei, which are active in states of waking and dreaming, but inactive in deep sleep. Churchland proposes that the contents of dreams are vectors activated by descending pathways that are then represented in multimodal form at the intralaminar nucleus.

It is noteworthy that in Churchland's view, dream sequences are rather “mundane and prototypical in character” (ER, p. 222). Dennett, in contrast, stresses the bizarre and illogical character of dream content. We suspect that Churchland's commitment to the mundane here is due to the difficulty he would have accounting for anything else as the result of operations in the sorts of recurrent networks he has discussed, in which the extant activation vectors must be the product of training and thus familiar, routine, the opposite of unprecedented. Since on his view the hidden nodes must supply the content for dreams via descending pathways, the hidden nodes will have to send bizarre information in order to trigger strange vectors, but if the hidden nodes send rather ordinary information the dream is going to have to be mundane in character. Thus, it seems likely that Churchland argues that dreams are mundane and prototypical in character because he can't (yet) account for hidden nodes sending strange information and, more importantly, he has no account of why or how this might occur in dreams but not waking.

Churchland correctly states that in *Consciousness Explained* Dennett does not offer an explanation of either the absence of consciousness in deep sleep

or its reappearance in dreaming. However, in his discussion of hallucinations, Dennett does suggest that the composition of dreams is to be accounted for along the same lines. Dennett takes the fact that a dreamer is cut off from sensory inputs as a significant fact that may serve to explain the bizarre nature of dreams; hallucinations are often experienced in cases of sensory deprivation and dreaming likewise occurs in a state of sensory deprivation. In his view, the content of a dream or a hallucination is supplied by something like a hypothesis generator that receives random confirmation of its hypotheses. That is, because the data-driven side of the confirmation cycle is in a state of sensory deprivation, it is not receiving the inputs that it is accustomed to receiving, and so lowers its threshold for activation. The result is random confirmation of hypotheses, which readily elaborate into bizarre sequences: “So the farmhouse in Vermont is now suddenly revealed to be a bank in Puerto Rico, and the horse I was riding is now a car, no a speedboat, and my companion began the ride as my grandmother but has become the Pope. These things happen.” (CE, p. 14). The dreamer receives enough information to satisfy her epistemic hungers. If a strange turn of events goes unquestioned, or is not questioned in detail, it goes unchallenged.

Dennett calls his discussion of hallucinations and dreams a “metaphorical theory sketch”. He does not take it to be a definitive account of either phenomenon and since he uses it to open his book, he acknowledges that it “does not even address the problem of our *consciousness* of dreams and hallucinations” (CE, p. 15). In order to make it more than a theory sketch we need to know more about the mechanism or process of hypothesis generation. How, and by what mechanism, are hypotheses generated? Is consciousness in fact a prerequisite for dreaming, as Churchland simply assumes? Or might there be narrative-generating phenomena with content indistinguishable from that of dreams in the absence of all the other marks of consciousness? That hypothesis was floated by Dennett, 1976, and it still requires careful consideration, since manifestly some of the standard symptoms of consciousness are notably lacking during dreaming. The suggestion that intralaminar nuclei activity settles the issue depends on the *independent* identification of such activity as not just the *sine qua non* of consciousness, but also as the guarantor of consciousness when other conditions (which other conditions?) are met. Churchland offers an explanation of the absence of consciousness in deep sleep and criticizes Dennett for not doing the same, but his own account falls far short of delivering what he claims. There is still lots of work to be done by both theorists, and no reason to be found in this quarter for supposing that their views, once elaborated, will be in conflict. As before, Churchland emphasizes neuroanatomical details while minimizing the scope of the tasks that the brain must accomplish, while Dennett emphasizes features of those tasks that apparently will require higher levels of explanation while postponing speculations about neuroanatomy.

In one regard, however, Dennett has been forthright about neuroanatomy, in his denial that there is *any* place in the brain where “it all comes together” for consciousness. Is Churchland declaring that the intralaminar nuclei compose what Dennett calls the “Cartesian Theater”? No, precisely because he does *not* declare that arrival at the intralaminar nuclei suffices for consciousness. On his view, those nuclei are a coordination and distribution center only, with consciousness emerging from elaborate further recurrent interactions distributed in space *and time* in the brain. Although Churchland does not discuss this important point, his view of the role of the intralaminar nuclei apparently avoids the traps that Dennett has described. In particular, the order of arrival of particular contents at the intralaminar nuclei need not have any bearing on the subjective order in conscious experience of those contents, and the time of arrival is not to be confused with the time of the “onset of consciousness” of the content in question. (For a further point of triangulation on this issue, see Jeffrey Gray (1995), another intrepid theorist of the neuroanatomy of consciousness, who singles out the hippocampus for a similar bottleneck role, and Dennett’s commentary, “Overworking the Hippocampus,” (1995).)

3. Language and Animal Consciousness

In the previous section we looked at Churchland’s criticisms of Dennett’s theory and saw that contrary to Churchland’s claim, Dennett does offer at least partial explanations of the features Churchland highlights. We also saw that the accounts offered by both Dennett and Churchland are incomplete or deficient, in largely complementary respects. Should their accounts be merged? Not if, as Churchland claims, Dennett’s account relies too heavily on the “failed prototype” of a von Neumannesque virtual machine, which Churchland characterizes as having a “broadly linguistic stream of activity” (ER, p. 263). Churchland says “broadly” because he recognizes that Dennett himself insists that much of what happens in what he calls the Joycean machine is not linguistic at all. It remains true, however, that Dennett sees language as playing an important, and perhaps even necessary, role in the *acquisition* of this virtual machine. The *sort* of consciousness human beings enjoy, Dennett claims, is—thanks in large measure to language—so different from that of any other species that to call the other varieties consciousness is to court confusion. This is well captured in Churchland’s cartoon figure 10.1 (ER, p. 265), except for the caption declaring that the Joycean machine is a “discrete-state” machine.

According to Churchland’s “neurocomputational” account, what one needs for consciousness instead is a “*suitably* [emphasis added] recurrent network” (see Figure 10.2, p. 268). One might pause to wonder if what makes some recurrent networks “suitably recurrent” is that this recurrence supports the running of a virtual machine, but this is not the direction in which Church-

land turns to cash out his term. When we sort out the unsuitable recurrent networks from the suitable, we find that “the higher animals are just as conscious as we are, at least when they are awake. For most of those animals have multilayered cortex, viscerally-connected parietal representations, and widespread recurrent connections between their thalamus and cortex, much as humans do.” (ER, p. 268) So, he insists, Dennett’s view is “unfair to animals,” since “the social institution of language has nothing to do with the genesis of consciousness” (ER, p. 269). Dennett has more recently addressed the question of non-human animal consciousness (1996), and clarified and elaborated his reasons for remaining skeptical about the similarity of the varieties of animal consciousness to ours. Churchland has not had the opportunity to address those reasons, nor does he develop his own claims about animal consciousness beyond the general declaration quoted above. He needs to consider what he should say about the many animals that lack “multilayered cortex and viscerally connected parietal representations.” Are they simply unconscious? Are there just two varieties of animal mind, then, conscious and unconscious? Which phenomena of human consciousness are also found in the minds of other species? A broader and more detailed survey of the kinds of animals minds, and the sorts of mental phenomena they can and can’t support, might convince Churchland that his view is just as much in danger of being unfair to animals—by positing an oversimplified view of consciousness that admits or rejects species from the charmed circle on the basis of neuroanatomical markers of variable relevance to the phenomena that matter.

Both Churchland and Dennett offer oversimplified sketches of theories of consciousness. Both sketches need much further elaboration before they can properly be counted as confirmable theories, much less confirmed theories, and since the weaknesses of each match the strengths of the other, Churchland’s attempt to compare Dennett’s theory unfavorably with his own is largely misdirected. When both theorists address the same set of questions, their positions may not be so different after all.

Bibliography

- Churchland, Paul, 1995, *The Engine of Reason, The Seat of the Soul*, MIT Press.
- Dennett, Daniel, 1991, *Consciousness Explained*, Little, Brown and Co.
- Dennett, Daniel, 1976, “Are Dreams Experiences?” *Philosophical Review*, LXXXV, 151–71.
- Dennett, Daniel, 1991, “Two Contrasts: Folk Craft Versus Folk Science, and Belief versus Opinion,” in J. D. Greenwood, ed., *The Future of Folk Psychology: Intentionality and Cognitive Science*, Cambridge University Press.
- Dennett, Daniel, 1995, “Overworking the Hippocampus,” *Behavioral and Brain Sciences*, **18**, pp. 677–78.

- Dennett, Daniel. 1996, *Kinds of Minds*, New York: Basic Books.
- Jeffrey Gray, 1995, "The Contents of Consciousness: A Neuropsychological Conjecture," *Behavioral and Brain Sciences*, **18**, pp. 659–76.
- Ramsey, William; Stich, Steven; and Garon, Joseph, 1991, "Connectionism, Eliminativism, and the Future of Folk Psychology," in William Ramsey, David Rumelhart, and Stephen Stich, eds. *Philosophy and Connectionist Theory*. Lawrence Erlbaum Associates, Inc., 1991, Chapter 9, pp. 199–228.