8 Two Steps Closer on Consciousness

DANIEL C. DENNETT

For a solid quarter century Paul Churchland and I have been wheeling around in the space of work on consciousness, and though from up close it may appear that we've been rather vehemently opposed to each other's position, from the bird's eye view, we are moving in a rather tight spiral within the universe of contested views, both staunch materialists, interested in the same phenomena and the same empirical theories of those phenomena, but differing only over where the main chance lies for progress. Our purely philosophical disagreements are arguably just matters of emphasis: we agree that folk psychological assertions limn real patterns in the world (to put it my way) and that these are (only?) useful approximations. Are they truths-with-a-grain-of-salt (my glass of mild realism is half full) or intermittently useful falsehoods (Paul's glass of eliminativism is half empty). We agree that there is no good motivation for shoehorning these folk categories into neuroscientific pigeonholes via a strict type identity theory, and even a strict functionalism would require some Procrustean labors that might better be postponed indefinitely, since the domain on the left hand side of the equation – the folk categories – is composed of items that are just not up to the task. This is true of folk categories more generally, not just the familiar terms of folk psychology. We also don't need counter-example-proof functionalistic definitions of charisma, moxie, or bizazz, though these are real qualities I have always admired in Paul.

To some observers, such as those of various mysterian persuasions, Paul and I are scarcely distinguishable, both happily wallowing in one 'scientistic' or 'reductionistic' swamp or another, taking our cues from cognitive scientists and unwilling or unable to begrudge even a respectful hearing to their efforts to throw shadows on the proceedings. For those who can see no significant difference between us, this essay will try to sharpen a few remaining disagreements, while at the same time acknowledging that in fact we are approaching harmony on a number of heretofore contested topics. I will try to close the gap further, much as I have always enjoyed his loyal opposition.

I met Paul in 1977, at the University of Manitoba, where I gave the first version of a talk that was subsequently published in *Brainstorms* (Dennett, 1978): "Two Approaches to Mental Images." The published version owed a lot to Paul's discussions, even though he is not referenced therein (I now note, with a smidgin of chagrin). We were both enthusiastic about enlarging the imaginations of philosophers of mind by getting them to dwell on actual scientific models and explanations, and since we tended to know different fragments of the relevant sciences, we had a lot to teach each other. His first book, *Scientific realism and the plasticity of mind* (Churchland 1979) was driven more by epistemological concerns than issues in the philosophy of mind but his discussion of perception and the plasticity of introspection was a major source of insights for me, especially informing my thinking about how important it was to ask what I later called the Hard Question: And Then What Happens? (*Consciousness Explained* 1991: 255). To capture the contents of consciousness, you need to see what a person can do with that state. Paul recently put it this way:

> Specifically, we both seek an explanation of consciousness in the dynamical signature of a conscious creature's cognitive activities, rather than in the peculiar character or subject matter of the contents of that creature's cognitive states. Dennett may seek it in the dynamical features of a 'virtual' von-Neumann machine, and I may seek it in the dynamical features of a massively recurrent neural network, but we are both working the 'dynamical profile' side of the street, in substantial isolation from the rest of the profession. (Churchland, 2002: 65)

Theories that stop when they reach some scarcely imagined "presentation" process (in the Cartesian Theater) are self-disabling, since they leave an unanalyzed (but knowledgeable, appreciative) witness to confront a now bafflingly contentful state with apparently miraculous powers of self-intimation, self-interpretation, and so on. You have to break up the given, and the taking of the given, into more modest parts whose operation we can actually begin to understand. Since people have no privileged access into this machinery, we'll just have to set aside the traditional philosophical method of introspection and turn to third-person models of processes that might arguably have the necessary competences. As he notes in this passage, Paul's favorite hammer and anvil for this breaking job is connectionism and vectors in a multidimensional space of content. I am impressed by some of

the work these tools can do, but don't view them as obviating the need for other perspectives, other levels of modeling.

Virtual machines versus recurrent neural networks – the opposition almost dissolves on closer inspection. I daresay that the virtual machines I like to talk about are implemented by the massively recurrent neural networks Paul likes to talk about. What else could do it? I entirely agree that there are massively recurrent neural networks churning away in our brains, and they are, as he has insisted for years, the key to understanding the neurocomputational perspective. So far, then, we differ only about whether in order to make sense of the powers of these networks we also need to describe their activity at a somewhat higher level, a virtual machine level. Should we take virtual machines in the brain seriously? That is the first question that divides us, and it leads directly to two others: Should we take memes seriously as what these virtual machines are 'made of'? and Is human consciousness a virtual machine (made of memes)? Note that the first two questions could get positive answers, and yet my most startling and revolutionary claim – answering yes to the third question – could be rejected. I suspect that part of Paul's motivation in dragging his feet so strenuously on the first two is that he wants to give himself plenty of room to maneuver in denying the third. Be that as it may, let me take them in order, and try to close the gaps. The history of our disagreement on this topic has had five steps:

1. CE. My 1991 book, *Consciousness Explained*, puts forward the virtual machine idea.
2. ER. In *The Engine of Reason, the Seat of the Soul* (1995) Paul includes a brief critique of my idea along with some delightful diagrams to illustrate it (264–69).
3. VV. In "The Virtues of Virtual Machines," (1999) Shannon Densmore and I portray step 2 as marred by a caricature of the idea of a virtual machine in the brain, and propose the division of labor suggested above: two levels of explanation, compatible and complementary.
4. VMC. In "Densmore and Dennett on Virtual Machines and Consciousness," his reply in the same issue of Phil and Phenom Research, Paul rejects our proposal quite vigorously (calling it "self-deceptive and uncomprehending"!) and rejects as well the language of virtual machines: "Those metaphors do not need to be qualified, they need to be junked" (763).

5. CC. In yet another essay, "Catching Consciousness in a Recurrent Net" in Brook and Ross (2002) Paul adds a further reading of the points of disagreement. It is clearly my turn to take a step – indeed I am one move behind his pace – so count this essay as two steps, taking his two prior steps as my springboard.

1. THE BASIC CONCEPT OF VIRTUAL MACHINES

Here I think the main obstacle to agreement is that Paul is still fixated on the wrong stereotype of virtual machines. He should clasp them to his bosom, not dismiss them scornfully. He needs them; they are just the right gadgets to complete his tool box. He got off on the wrong foot with an unflattering portrait of a virtual machine in his original critique in ER: "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" (264). But as was made clear in VV, of the four adjectives in this list, I endorse only one: serial – and even that one gets highly qualified in my Multiple Drafts Model. In VMC, Paul acknowledged that I had never said my virtual machines were classical or discrete-state or rule-governed and indeed had quite explicitly cautioned against those interpretations, but, he claimed, this made matters even worse: my altered and metaphorical use of the concept of virtual machines was Pickwickian at best: by disavowing the very features that explain the power of classical virtual machines, I was creating an illusion of explanation where none existed. By Paul's lights, a virtual machine only makes sense when you have explicit source code (I had somewhat jocularly characterized a virtual machine as a "machine made of rules" and these are the "rules") implemented on a digital (discrete-state), serial machine. He says this, but he offers no reasons, and for the life of me, I don't see why he thinks so. To me, this is like insisting that you can implement a virtual machine only on an Intel chip. Why should other architectures, even non-digital architectures, be ruled out? His explanation doesn't help much:

> The network isn't following fuzzy rules, or imperfectly marshaled rules, or virtual rules; it isn't following rules at all. What we should look for, in explanation of the network's behavior, is the acquired dynamical landscape of its global activation space. That, plus its current activation state, is what dictates its behavior. Rules play no causal role at all, and neither do 'rules'. To use that term in scare quotes, as Dennett does, is just to undermine the primary negative point here, and to set up explanatory hopes and expectations

> (concern 'virtual machines' and 'design level explanations') that are doomed to go unsatisfied (765).

Strong talk, but there is a problem of levels lurking here, which we can make vivid by imagining an electrical engineer or chip designer making the parallel claims about a von Neumann machine:

> ...it isn't following rules at all. What we should look for, in explanation of [the von Neumann machine's] behavior, is the [temporary] dynamical landscape of its global activation space. That, plus its current activation state, is what dictates its behavior....

Programs in memory are, after all, just large fields of varying voltages that determine the dynamic sequence of voltage changes racing through circuit boards. Rules play no causal role in them either! Once you compile the source code, the "rules" evaporate. When you get right down to it, all the causal work is done at the level of flip flops and logic gates, and when a logic gate responds to its input "It is no more following rules than is the water of the Mississippi following rules in order to meander its way down a literal landscape from the northwest plains to the Gulf" (VMC: 764).

There is no rule-following in the hardware – either neural or silicon – but it doesn't follow from this that there is no rule-following (or "rule"-following – the scare quotes are needed in both cases) at a higher level. You simply cannot make sense of the versatile powers of a von Neumann machine without ascending to the program level, the virtual machine level, and at that level the regularities to be discovered, while based on or implemented via fundamental physical microregularities (the province of the gate designer) cannot be accounted for at the level of physics. To take a vivid example, consider the visible regularities of click-and-drag in the desktop user interface virtual machine. The icon on the desktop changes color and becomes somewhat translucent when it's being moved, and reverts to its original color when the cursor lets go of it. These regularities are not curious reflections of the implementation physics ("Hmm, could it be heating up due to some friction in the medium?") but regularities imposed by patterns in the implementation, and these regularities can be concocted ad lib, depending on features of the world outside the hardware, features tracked, or honored, or represented. The "physics" of the virtual machine is whatever the designers want it to be; it is virtual physics.

The same sorts of regularities are ubiquitous in minds. Consider, for instance, the regularities of people's reactions to a Stroop test, in which color words such as "red" and "green" and "blue" are written in differently colored inks and people are asked to name the colors of the inks, not the words written. People who are illiterate have no difficulty following the instructions and naming the colors; people who can read find it very hard to follow the instructions. Why is it harder for some people? Is it physically more difficult? Well yes, of course, in a sense. All difficulties are physical difficulties in the end. But the shape or pattern of the difficulties may need another level to describe. Consider Paul's claim: "What we should look for, in explanation of the network's behavior, is the acquired dynamical landscape of its global activation space. That, plus its current activation state, is what dictates its behavior." Yes, but the only explanatory way to describe that acquired dynamical landscape is in terms of the virtual machine thereby implemented. In this instance, what matters is whether there is an English-reading machine installed. In another instance, a much shorter-lived virtual machine might be responsible for a predictable effect. (As in the old trap questions: What kind of music did Woodie Guthrie sing? Folk. Who was President during the California Gold Rush? Polk. What do you call the white of an egg? Yolk. No, you dummy; albumin!) These are tiny toy examples to illustrate the phenomenon; when they are compounded into much more complex and highly articulated structures, the utility of the virtual machine perspective is undeniable. Cognitive psychology abounds in confirmed hypotheses about these machines, the conditions under which they are invoked and the circumstances under which they can be provoked into malfunction.

Perhaps Paul's longstanding distaste for the terminology of virtual machines should be catered to here, and we should let him treat himself to an alternative vocabulary for talking about the highly structured dispositions imposable (with a little practice or training) on the underlying "global activation space," just so long as he recognized that many of the highly salient regularities at one level will be inscrutable at his favored lower level, and that these regularities are mostly physically arbitrary in just the way the changing color of the dragged icon is physically arbitrary (from the point of view of the underlying machinery). Then there would be only a terminological preference separating us: what I and others (e.g., Metzinger 2003) insist on calling virtual machines, he would insist on calling something else. But I continue to urge him to chill out and recognize the tremendous utility, the predictive fecundity, the practical necessity of speaking of these higher levels as virtual machines. As a parade case, I commend Ray

Jackendoff's recent book, *Foundations of Language* (2002) which is a tour de force of (speculative, but highly informed, and deeply constrained) modeling of the virtual machine levels of neural implementation of language. The details matter, and I challenge anybody to say how they might recast all the insights in, say, Chapter 6, "Lexical Storage versus Online Construction," and Chapter 7, "Implications for Processing," in terms of the underlying recurrent neural networks. (See also pp. 22–3 for Jackendoff's reflections on this issue of the level of modeling.)

In CC, Paul's most recent step, he perseveres in his campaign against virtual machines, in a most curious way. First he notes that I am "postulating that, at some point in the past, at least one human brain lucked/stumbled into a global configuration of synaptic connections that embodied an importantly new style of information processing, a style that involved, at least occasionally, the sequential, temporally structured, rule-respecting kinds of activities seen in a typical vN [von Neumann] machine" (70). Yes, that's one way of putting it, and Paul goes on to acknowledge that indeed this possibility has been demonstrated in artificial recurrent networks. For instance, Cottrell and Tsung have trained networks to add individual pairs of n-digit numbers and distinguish grammatical from ungrammatical sentences in simplified formal languages.

> But are these suitably trained networks 'virtual' adders and 'virtual' parsers? No. They are literal adders and parsers. The language of 'virtual machines' is not strictly appropriate here, because these are not cases of a special purpose 'software machine' running, qua program, on a vN-style universal Turing machine (71).

This leaves me gasping. Paul, having just acknowledged that I am claiming that there is a perfectly good counterpart to classical virtual machines in the world of parallel machines, and having offered just the sort of example I would have chosen to illustrate it, pulls the definitional plug on me. This is not "strictly" appropriate use of the term "virtual machine" he says, because it isn't running on a vN machine! This begs the question. The Cottrell and Tsung machine is a special purpose software machine running, qua program, on a parallel machine. That very same 'hardware' recurrent network could have been trained up to do something else, after all. It was trained up to be, at least for a while, an adder or a parser. That's what a virtual machine is. A virtual machine does the very thing ("literally") a hardware machine does; it doesn't just approximate the task.[1] You can't retrain a hardware adder. If Paul thinks these trained neural networks are literal adders and parsers, I wonder what on earth he would call a virtual adder or parser.

Pursuing this definitional curiosity further, Paul sees an irony:

> if we do look to recurrent neural networks – which brains most assuredly are – in order to purchase something like the functional properties of a vN machine, we no longer need to 'download' any epigenetically supplied meme or program, because the sheer hardware configuration of a recurrent network already delivers the desired capacity for recognizing, manipulating, and generating serial structures in time, right out of the box (71).

This remark baffled me for some time. The underlying and untrained potential for recognizing, manipulating and generating serial structures in time is – must be – there, but saying that that capacity gives recurrent neural networks the functional architecture of a vN machine is like selling somebody a laptop without even an operating system and calling it a word processor. A randomly weighted recurrent neural net "right out of the box" is no serial vN machine. Precisely what we do need is the installation from outside of some highly designed system of regularities.

Sometimes we do the design work ourselves, laboriously, and sometimes we get a relatively easy download of largely predesigned systems. A natural language, as Chomskians are famous for telling us, installs itself in jig time in just about everybody, while sound probabilistic thinking is an unnatural act indeed, seldom successfully implemented in neural tissue. Several decades ago, I mastered the Rubik's cube, and got quite deft at spinning it into order. The fad expired; twenty years of disuse, like the similar hiatus in my use of German and French, have taken their toll, and a few months ago it took me quite a few hours to reinvent and re-optimize my cubist competence. (I guess I just needed to waste some precious time! During the obsessional phase, I couldn't stop imagining the subroutines and problems. Thank goodness I soon got over it.) If I don't rehearse my Rubik routines often in the months ahead, they will soon slip away again. What is this thing that can be problematically preserved in the connection strengths in my recurrent neural networks? It has structure that would be practically invisible to anyone intent on studying my neural networks and their dynamic properties, but readily describable as a sort of program that I have installed in myself and can run almost as mindlessly now as I usually run my English parser.

I think the case has been made for the appropriateness of virtual machine talk in cognitive neuroscience, and not just by me, and I look forward to the day when Paul retires from this dubious battle. I also anticipate a bounty of insights to flow from Paul when he exorcizes another bee in his bonnet: his mistrust of memes.

2. MEMES

I'll be brief about this, since Paul is brief and I have had a lot to say in defense of memes elsewhere (Dennett 1995, 2001a–c, 2002, forthcoming). Part of his problem with memes stems from his decision to take theories, probably the largest, rarest, hardest-to-transmit, most unwieldy of all cultural objects, and use them as his examples of choice.

"An individual virus is an individual physical thing, locatable in space and time. An individual theory is no such thing" (Churchland 2002: 66). True, but an expression or representation of an individual theory is an individual physical thing, and if we take the gene/meme parallel seriously, we recognize that a gene, too, is the information, not the vehicle of the information, which is always an individual physical thing. To see this vividly: ask yourself the following question. What if people in the future decided to forego sex and reproduce thus: Al and Barb both have their genomes sequenced, whereupon a meiosis program randomly composes unique Al-gamete and Barb-gamete specifications from their respective genomes and joins them into a zygote specification – a computer file that specifies the genome of an offspring. This specification is sent to a lab that thereupon hand-assembles that very genome out of materials taken from other biological sources, and creates an implantable "fertilized" embryo, which (for good measure) is then implanted in a surrogate mother, not Barb. Are not Al and Barb the "biological" father and mother of the resulting child? It's the information that counts. So genes are like theories in this regard: "abstract patterns of some kind imposed on preexisting physical structures . . . " (66).

"Furthermore," Paul goes on, "a theory has no internal mechanism that effects a literal self-replication. . . . " Neither does a virus, of course. It travels light and is artfully designed (by Mother Nature) to mindlessly commandeer the copying machinery in the cell it invades. A virus can be considered a string of DNA with attitude, but once again, it is the information that counts. Prions bring this out even more clearly, as Szathmary (1999) shows. Similarly a meme invades a body and gets itself copied, again and again, in a brain, But the physical token doesn't enter the body literally. A written word, for instance, does not enter the body (unless you're in the habit of eating your words!); rather, it produces an offspring on your retina, which then gets replicated again and again and again in your brain. Not so, says Paul. "It is that there is no such mechanism for theory-tokens" (67). I beg to differ, not just about individual words, and other individual vehicle-copies that get perceived, but even about "theories" large and small. This is what we call rehearsal or review, and it happens all the time. I just gave the vivid example

of my involuntary rehearsal of Rubik's cube memes, betokening themselves thousands of times in my poor brain, building ever stronger, better traces. What was being held constant while all the connection-strengths were being adjusted? The information.

Whole theories are unwieldy memes. Consider a much better example: a word. A grade school teacher of mine used to admonish "Say a word three times and it's yours!" and while the advice was largely gratuitous, the principle was right on target. Repetition is close to being a necessary condition for memorization, especially when we acknowledge that involuntary repetition (and unconscious repetition, which probably is ubiquitous) may do most of the work. If my Rubik's cube memes don't have offspring in the weeks to come, the lineage may well go extinct. What needs to be resurrected in me is not so different from woolly mammoth DNA after all. It lies unusable and unreplicable in the Vast state-space of my brain's parallel recurrent networks unless it gets regular cycles of reproduction. Paul (2002: 67) notes that "the 'replication story' needed, on the Dawkinsean view, must be nothing short of an entire theory of how the brain learns. No simple 'cookie-cutter' story of replication will do for the dubious 'replicants' at this abstract level." Exactly. Now where's the problem? Nobody ever said that a meme had to replicate by invading a single neuron.

It is curious that Paul ignores this perspective, since he has written hymns glorifying the repetitive power of recurrent neural circuits and their role in any remotely plausible theory of learning. The habit of rehearsal is a potent habit indeed, and it is required – Paul says as much in his discussion of the difficulties of internalizing a theory – to drive a theory into the network. How do you get to Carnegie Hall? Practice practice practice. But of course a lot of the rehearsal is not only not difficult; a lot of it is well nigh impossible to shut down. Rehearsal is itself a habit that is ubiquitous in our phenomenology – and it's just the tip of the iceberg! So here I'll just help myself to Paul's hymns to recurrence, for right there is the hardware that underlies the software, the built-in proto-rehearsal machinery that makes copying one's memes such an irresistibly easy step to take. The differential replication of memes within an individual brain is the underlying competitive mechanism of learning. And here a well-known evolutionary trade-off confronting parasites – should they specialize in the within-host competition against other strains of resident parasites (the path to virulence) or should they specialize on the competition to get from one host to the next (which leads to a-virulence, so that hosts can be up and about and in position to infect others)? – finds a parallel in the evolution of memes: getting a mnemonically potent phenotype that will get obsessively

rehearsed in one brain is part of the battle: getting transmitted favorably to another brain is a quite different venture. (I'll never forget John Perry's amusing bumper sticker: Another Family for Situation Semantics. John and a few colleagues and students had replicated the novel memes of situation semantics in uncounted rehearsals, but was anybody else ever going to be infected? John was not above trying the Madison Avenue approach.)

3. THE JOYCEAN MACHINE

But even if the virtual machine idea is worth pursuing, and even if the meme idea has some attractions, is there any hope for the preposterous claim that consciousness – consciousness! – is the activity of a virtual machine that only human beings implement, a virtual machine that depends on culture in general and language in particular? Surely this is just crazy! Many think so. Some of the wisest (and least conservative) heads in cognitive science think so. Paul thinks so.

> Instead, I shall argue, the phenomenon of consciousness is the result of the brain's basic hardware structures, structures that are widely shared throughout the animal kingdom, structures that produce consciousness in meme-free and von-Neumann-innocent animals just as surely and just as vividly as they produce consciousness in us (CC: 65).

This is a factual disagreement, not necessarily a philosophical disagreement of any sort, and he may be right. Or he may not. The point I want to make here is that his grounds for his belief are not anywhere near as strong as he thinks. I grant that we share a large part of our neurocomputational architecture with other animals, and that this shared architecture is sufficient to explain a great deal of the integrated, coherent, subtle behavior that both we and other animals exhibit, but I want to resist the further supposition, popular though it undoubtedly is, that this shared architecture (at the 'hardware' level) gives animals the sort of subjectivity, the sort of stream of consciousness, the point of view that we human beings all know that we share.

Paul is willing to grant that an uncultured, untutored, languageless mind is a relatively barren mind, perhaps even drab and boring in comparison to a normal (noninfantile) human mind:

> I do not hesitate to concede to Dennett that cultural evolution – the Hegelian unfolding we both celebrate – has succeed in 'raising' human

> consciousness profoundly. It has raised it in the sense that the contents of human consciousness – especially in its intellectual, political, artistic, scientific and technological elites – have been changed dramatically.... Readers of my 1979 book (see especially Chapters 2 and 3) will not be surprised to hear me suggesting still that the great bulk and most dramatic increments of consciousness-raising lie in our future, not in our past.
>
> But raising the contents of our consciousness is one thing – and, so far, a purely cultural thing. Creating consciousness in the first place, by contrast, is as firmly neurobiological thing, and that must have happened a very long time ago. For the dynamical cognitive profile that constitutes consciousness has been the possession of terrestrial creatures since at least the early Jurassic. James Joyce and John von Neumann were simply not needed (CC: 79).

That could not be clearer. I particularly applaud his allusion to Chapters 2 and 3 of his 1979 book, which remain, for me, my favorite bits of Churchlandiana. And as I say, he may be right. But until I am proved wrong, I am going to defend a more abstemious and minimalist view, one that resists the easy and popular course of supposing, with tradition, that our furry friends (and, if Paul is right, our feathered friends and even many of our scaly friends) have streams of conscious much like our own. I consider it telling that when Paul disparages this outrageous view of mine in ER, he shows a diagram of a grumpy-faced chimp (contrasted with a smiling member of H. sapiens) and goes on to say that "Dennett's account of consciousness is... unfair to animals" (269). The moral dimension is thus lurking not far beneath the surface, and we should all recognize that part of what is repugnant (to many) in my view is that it seems destined to license a shocking callousness with regard to animals (who are not really conscious, just as Descartes said, that evil man!). Recognizing that this is, or ought to be, an irrelevant consideration insofar as we want to know the scientific truth, and recognizing moreover that it nevertheless plays a potent role in biasing people against any hint of such a position, we ought to go out of our way to consider whether or not it might be true. That is why I continue to push my shocking view: because I see no good reason has been offered for not counting it as a serious candidate.

It might well seem that the disagreement between Paul and me here is just a special case of our earlier disagreement about whether a recurrent neural network counts as a serial architecture. He says yes, and I say no: the settings of the connections make all the difference, since they are what fix the truly remarkable powers of some such recurrent networks – by

programming them, in effect. Similarly, he says that animals are conscious, and I say that they are not, since what they are conscious of, the settings, if you will, that flavor their consciousness, do not do enough good work to count. But if that were all that divided us, it wouldn't be much of a disagreement. I could lament the fact that you just can't teach a chimp to solve the Rubik's cube, and so, you see, the chimp has such a paltry stream of consciousness that it hardly counts as conscious at all, and Paul could insist, on the contrary, that dim though a chimp's stream of consciousness is, it still counts as a stream of consciousness. But I am envisaging a more radical difference between the chimp and us. I am supposing that nothing like a stream of consciousness occurs in a chimp brain precisely because what kindles and sustains such a stream of consciousness in us is a family of microhabits of self-stimulation that have to be installed by culture. Without the cultural inculcation, we would never get around to having a stream of consciousness, though, of course, we would be capable of some sort of animalian activity. I am not denying that there are crucial architectural differences between chimp brains and ours. If it weren't for these, chimps could be enculturated and given human languages of some kind – manual sign languages most likely. But the differences might be quite subtle (see Deacon 1997 for an insightful account of the possibilities.) Deaf human infants, for instance, are intensely curious about human communication in spite of the absence of auditory input, while chimps that can hear perfectly well have to be heavily bribed with rewards to pay attention to human efforts at communication. Our brains are in some regards genetically designed to download cultural software, and chimps' brains are apparently not so designed.

Interestingly, Paul himself draws attention to this in a passage that is meant to cast doubt on the meme/virus parallel: "A mature cell that is completely free of viruses is just a normal, functioning cell. A mature brain that is completely free of theories or conceptual frameworks is an utterly dysfunctional system, barely a brain at all" (CC: 67). There are several points to make about these claims. First, it is not the case in general that normal cells can function without any viruses or other endosymbionts. After all, the mitochondria that are the prerequisite for eukaryotic life started out as cellular parasites, and more and more of the standard intracellular machinery turns out to have begun its career as software downloads of a sort. This is still going on, and it is well known that many cells cannot perform their current functions without the aid of "foreign" visitors of one sort or another. As in computer science, software development often precedes hardware development. More important for the present point, I agree with Paul that a mature human brain free of culture is utterly dysfunctional.

That's my point. But a chimp brain free of theories or conceptual frameworks – depending on what we mean by that – is not so obviously abnormal or dysfunctinoal. Animals learn from their own experience, by trial and error and general exploratory behavior, and Avital and Jablonka (2000) draw attention to the evidence that much of what has been standardly deemed to be "instinctual" knowhow transmitted through the genes is better considered animal "tradition" and can in fact be imparted by parent-offspring interactions and other social learning situations. (See also my review in *Journal of Evolutionary Biology*, Dennett 2002b). But no nonhuman animal species has a brain that is as adapted for massive cultural downloading as ours is, and hence no nonhuman animal is as handicapped by being denied its conspecific culture as we would be. Given these undeniably huge differences in both potential and dependence, the assumption that animal brains are architecturally enough like ours to sustain something properly called a stream of consciousness owes more to cultural habit than scientific insight.

It is worth noting that as primatologists and animal psychologists learn more and more about the minds of chimpanzees and bonobos (and dolphins and orangs and other favored species), they discover more and more surprisingly blank walls of incomprehension. The idea that these creatures are, in some regards, sleepwalking through life, to put it crudely and misleadingly, is not so easy to shake. I have been monitoring and occasionally contributing to the experimental literature on animal intelligence – especially higher-order "theory of mind" intelligence – for several decades, and to me the striking fact is that for every gratifying instance of (apparent) comprehension in one species or another, there are more instances of frustrating stupidity, unmasked tropism, and hard-to-delineate density that is hard to reconcile with the standard presumption that these creatures are confronting a world of experience pretty much the same way we are. Yes, they can be seen to be ignoring some things and attending to others, but the attention they can pay doesn't seem to enlighten them in many of the ways ours does. In short, they don't show much sign of thinking at all.

"But still, they are conscious!" Oh yes, of course, if all you mean is that they are awake, and taking in perceptual information, and coordinating their behavior on its basis in relatively felicitous fashion. But if that is all that you mean by asserting that they are conscious, you shouldn't stop at mammals, or vertebrates. Insects are conscious in that sense. Molluscs are too, especially the cephalopods. That is not what I am skeptical about. I am skeptical about what I have called the Beatrix Potter syndrome: the imaginative furnishing of animal minds with any sort of subjective appreciation, of fearful

anticipation and grateful relief, of any capacity to dwell on an item of interest, or recall an episodic memory, or foresee an eventuality. Animals can "learn from experience," but this kind of learning doesn't require episodic memory, for instance. When we see a dog digging up a buried bone it is quite natural for us to imagine that the dog is happily recalling the burying, eagerly anticipating the treasure to be recovered just as he remembered it, thinking just what we would if we were digging up something we had earlier buried, but in fact there is not yet any good evidence in favor of this delightful presumption. The dog may not have a clue why he is so eagerly digging in that spot. (For the current state of the evidence of "episodic-like" memory in food-caching birds and other animals, see Clayton and Griffiths 2002). And animals can benefit from forming a "forward model" of action that doesn't require the ability to foresee "consciously"; we ourselves are seldom conscious of our forward models until they trip up on an anomaly. Once we have stripped the animal stream of consciousness of these familiar human features, it is, I claim, no longer importantly different from a stream of unconsciousness! That is, it is a temporal flow of control processing, with interrupts (pains, etc.) and plenty of biasing factors, but it otherwise shows few if any of the sorts of contentful events that we associate with our own streams of consciousness. I think we need to set aside the urge to err on the side of morality when we imagine animals' minds; this attitude has its role in making policy decisions about how to treat animals, but should not be hardened into an unchallengeable "intuition" when we ask what is special about consciousness.

Paul's firm insistence that of course animals are conscious, and that human consciousness is just richer, is to me like the claim that five-year-olds write novels. Here's Billy's: Tom hit Sam. The End. It's a novel if you say so. But why are you so eager to say it's a novel? I am as eager as Paul is to support the humane treatment of animals, but I don't believe that the right way to do it is to saddle myself uncritically with the folk concept of (animal) consciousness that makes us happy to imagine that there is some (nice, warm, familiar, unified) sort of inner "show" in animal's brains, "the way there is in ours." When you look hard at the conditions for the "show" going on in ours, you begin to see that it is probably heavily dependent on a great deal of activity that is specific to our species. If you doubt this, ask yourself the following weird questions: What is it like to be an ant colony? What is it like to be a brace of oxen? The immediate, "intuitive" answer is that it is not like anything to be either one of these things, because these things, impressively coordinated though they may be in many regards, are not sufficiently unified, somehow, to support a single

(conscious) point of view. But just putting that ant colony inside a skull wouldn't automatically unify their activities the extra amount needed, would it? Tremendous feats of coordination are possible in control structures that nevertheless do not conspire to create the sort of user-illusion that we human beings call consciousness. If animal brains could do what ant colonies can do, why would animal brains bother doing all this further work? I have offered a sketch of an evolutionary explanation about why our brains, the brains of a linguistically communicating social species, would go to this extra work. I grant that there may well be an explanation for why the architectural features Paul finds shared in most if not all animal brains should be seen to yield enough of the human features to persuade us to call the processes that run so effectively therein conscious processes. But that is work still to be done, not presupposed. This is an empirical issue, not a philosophical one, except insofar as there are residual unclarities or misapprehensions about the meanings of the claims being advanced. And to echo the note from Paul with which I began this essay, for all our disagreements, Paul and I are united in thinking that these are scientific questions that cannot be solved, or even much advanced, by the intuition-mongering of armchair philosophy.

Note

1. It is worth remembering that today almost all hardware machines are designed and exhaustively tested as virtual machines long before the first hardware version is built. It is also possible, of course, for a virtual machine to be designed to approximate the task of a hardware machine. Virtual machines are the ultimate modeling clay – you can make just about anything out of rules.

Works Cited

Avital, E. and Jablonka, E. (2000). *Animal Traditions: Behavioural Inheritance in Evolution*. Cambridge, Cambridge University Press.

Brook, A. and Ross, D. eds. (2002). *Daniel Dennett*. Cambridge, Cambridge University Press.

Churchland, P. (1995). *The Engine of Reason, the Seat of the Soul*. Cambridge, MIT Press.

Churchland, P. (2002). "Catching Consciousness in a Recurrent Net." *Daniel Dennett*. Brook and Ross, 64–81.

Churchland, P. (1999). "Densmore and Dennett on Virtual Machines and Consciousness" *Philosophy and Phenomenological Research* **59** (3): 763–7.

Clayton, N. S. & Griffiths, D. P. (2002). Testing episodic-like memory in animals. In Squire, L. and Schacter, D. (eds.) *The Neuropsychology of Memory*, Third Edition. Chapter 38, New York, Guilford, 492–507.

Deacon, T. W. (1997). *The Symbolic Species*. New York, Norton.

Densmore, S. and Dennett, D. (1999). "The Virtues of Virtual Machines" in *Philosophy and Phenomenological Research* **59** (3): 747–61.

Dennett, D. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MIT Press.

Dennett, D. (2001a). "The Evolution of Culture," *The Monist* **84** (3): 305–24.

Dennett, D. (2001b). "Memes: Myths, Misgivings, Misunderstandings," Chapel Hill Colloquium, October 15, 1998, University Chapel Hill, North Carolina, translated into Portugese and published in *Revista de Pop*, No. 30, 2001.

Dennett, D. (2001c). "The evolution of evaluators." In *The Evolution of Economic Diversity*, Antonio Nicita and Ugo Pagano, eds., New York, Routledge, 66–81.

Dennett, D. (2002a). "The New Replicators," In *The Encyclopedia of Evolution*, volume 1, Mark Pagel, ed., Oxford, Oxford University Press, E83–E92.

Dennett, D. (2002b). "Tarbutniks rule. Review of Eytan Avital and Eva Jablonka, Animal Traditions: Behavioural Inheritance in Evolution, 2000." *Journal of Evolutionary Biology* **15**: 329–34

Dennett, D. forthcoming, "From Typo to Thinko: When Evolution Graduated to Semantic Norms," In S. Levinson & P. Jaisson (Eds.), *Culture and evolution*. Cambridge, MA, MIT Press.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford, Oxford University Press.

Metzinger, T. (2003). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MIT Press.

Szathmary, E. (1999). "Chemes, Genes, Memes: A Revised Classification of Replicators." *Lectures in Mathematics in the Life Sciences*, **26**: American Mathematical Society, 1–10.