Chapter Nine: The Architecture of The Human Mind

Part 1: Where Are We?

Though much ground has been covered, Dennett now attempts to deliver on some postponed promises by examining more critically and in detail the issues often raised by his opponents, but first gives a thumbnail sketch of his position that is worth quoting in its entirety:

> "There is no single, definitive "stream of consciousness," because there is no central Headquarters, no Cartesian Theater where "it all comes together" for the perusal of a Central Meaner. Instead of such a single stream (however wide), there are multiple channels in which specialist circuits try, in parallel pandemoniums to do their various things, creating Multiple Drafts as they go. Most of these fragmentary drafts of "narrative" play short-lived roles in the modulation of current activity but some get promoted to further functional roles, in swift succession, by the activity of a virtual machine in the brain. The seriality of this machine (its "von Neumannesque" character) is not a "hard-wired" design feature, but rather the upshot of a succession of coalitions of these specialists.
>
> The basic specialists are part of our animal heritage. They were not developed to perform peculiarly human actions, such as reading and writing, but ducking, predator-avoiding, face-recognizing, grasping throwing, berry-picking, and other essential tasks. They are often opportunistically enlisted in new roles, for which their native talents more or less suit them. The result is not bedlam only because the trends that are imposed on all this activity are themselves the product of design. Some of this design is innate, and is shared with other animals. But it is augmented, and sometimes even overwhelmed in importance, by microhabits of thought that are developed in the individual, partly idiosyncratic results of self-exploration and partly the predesigned gifts of culture. Thousands of memes, mostly borne by language, but also by wordless "images" and other data structures, take up residence in an individual brain, shaping its tendencies and thereby turning it into a mind."

Dennett goes on to say that while his theory draws equally from the fields of psychology, neuro-biology, Artificial Intelligence, anthropology and philosophy, many of these researchers in one field can't understand why he wastes his time with ideas from competing fields. As Dennett sees it, a typical problem that befalls some neuroscientists is the occupational hazard of thinking that consciousness is the "end of the line"- which is a bit like thinking that the end product of apple trees is apples, when it's really more apple trees. For example he presents the hypothesis by Crick and Koch that:

> "We have suggested that one of the functions of consciousness is to present the result of various underlying computation and that this involves an attentional mechanism that temporarily binds the relevant neurons together by synchronizing their spikes in 40 hz oscillations."

Dennett complains: "So a function of consciousness is to *present the results of underlying computations*- but to whom? The Queen? This kind of hypothesizing merely begs the question: "And then what happens?" and avoids the hard questions of how to explain "the tricky path from (presumed) consciousness to behavior, including, especially, introspective reports."

Hypotheses from cognitive science or AI almost never have this problem. They simply propose a "workspace" or "working memory" that replaces the Cartesian Theater and instead merely show how the results of computations eventually can guide behavior, inform verbal reports, etc. These investigations are so concerned with the "work" being done that they never get around to explaining the problem philosophers find so fascinating and that is the "sort of delectation of phenomenology that seems to play an important feature in human consciousness."

In this respect, Dennett says, some neuroscientists sometimes look somewhat like dualists in that once they have "presented" things, they still haven't answered the question: who is the "presentee?" On the other hand, cognitive scientists seem to look a little like zombists (automatists?), since they describe structures and processes unknown to neuroscientists and purport to show how all the work can get done without that intuitive narrative sense we really do seem to hold in our heads.

Part 2: Orienting Ourselves With The Thumbnail Sketch

Dennett now proposes to step through the above sketch and explain how a convincing theory of consciousness could be constructed out of these parts. Though he warns his reader not to expect confirmation at this stage where the field is so wide open as to not know what the right questions are, much less the right answers.

He starts by quoting Baars (1988) that there seems to be a "gathering consensus" that consciousness is accomplished by a "distributed society of specialists that is equipped with a working memory, called a global workspace, whose contents can be broadcast to the system as a whole." Let's look at each essential point of the "thumbnail sketch" and examine it in more detail:

> "There is no single, definitive "stream of consciousness," because there is no central Headquarters, no Cartesian Theater where "it all comes together" for the perusal of a Central Meaner…."

While everyone agrees that there is no single point in the brain "where it all comes together" there are still many that fall into the habit of "trying to find a representational space in the brain (smaller than the whole brain) where the results of various discriminations are placed in registration…". Such as "marrying the sound track to the film", "coloring in the shapes" and "filling in the blanks." Continuing with the sketch:

> "…Instead of such a single stream (however wide), there are multiple channels in which specialist circuits try, in parallel pandemoniums to do their various things, creating Multiple Drafts as they go. Most of these fragmentary drafts of "narrative" play short-lived roles in the modulation of current activity…"

Many in AI have long stressed that an understanding of the importance of "narrativelike sequences" is critical for explaining consciousness. That is, snapshots that are not just sequenced, but structures that instead are specific to temporal sequence types directly. Again, continuing with the sketch:

> "but some get promoted to further functional roles, in swift succession, by the activity of a virtual machine in the brain. The seriality of this machine (its "von Neumannesque" character) is not a "hard-wired" design feature, but rather the upshot of a succession of coalitions of these specialists…"

Noting the relatively slow, and awkward pace of conscious mental activity (compared to reaction to stimuli), this suggests that the brain is simply not designed, that is not hard-wired, for such activity. Instead consciousness might be a sort of serial virtual machine implemented on the parallel hardware of the brain. The point being that the underlying neural architecture is far from a "blank slate" at birth, but is utilized as a substrate upon which the cognitive functions are built on. Dennett goes on with the sketch:

> "The basic specialists are part of our animal heritage. They were not developed to perform peculiarly human actions, such as reading and writing, but ducking, predator-avoiding, face-recognizing, grasping throwing, berry-picking, and other essential tasks…"

That these and other specialists exist in the brain is well supported by the evidence although their size, roles and organization are hotly contested. Dennett describes how biologists have been able to explain almost all animal behavior with "quilts" of Innate Releasing Mechanisms (IRMs) and Fixed Action Patterns (FAPs) and that this is likely the basis for the evolution of more general purpose minds. One interesting point in "multiple systems" of this sort is that they buy one tolerance for noisy images of the kind that one encounters in the real world.

How these various modules interact and cooperate is the central problem in consciousness, but it ought to be clear that any appeal to a "central facility" is Cartesian doom. The point being that positing ever more stupid "homunculi", demons, agents, units, whatever you call them, is fine so long as there is a mechanism for them to interact and amplify coalitions or hierarchies. This is the basis for all functionalist models of consciousness. As Dennett sketch concludes:

> "They are often opportunistically enlisted in new roles, for which their native talents more or less suit them. The result is not bedlam only because the trends that are imposed on all this activity are themselves the product of design. Some of this design is innate, and is shared with other animals. But it is augmented, and sometimes even overwhelmed in importance, by microhabits of thought that are developed in the

> individual, partly idiosyncratic results of self-exploration and partly the predesigned gifts of culture. Thousands of memes, mostly borne by language, but also by wordless "images" and other data structures, take up residence in an individual brain, shaping its tendencies and thereby turning it into a mind."

Well this is certainly bold enough. Now it is time to fish or cut bait. How do the homunculi interact to accomplish anything? Let's look at some models that might explain these ideas.

Part 3: And Then What Happens?

We've already seen how the von Neumann architecture works by a serial fetch-execute cycle that relies on a program of instructions with an obvious bottle neck. But AI model builders revised all this by expanding the "workspace" and replacing the rigid fetch-execute cycle with a more flexible "blackboard" where various demons could write messages for all other demons to read, which in turn provokes another cycle of writing and reading.

ACT* is one such model based on a working memory where the basic "production" actions occur which are pattern recognition mechanisms tuned to fire on "if-then" operators. Consider the orders given a human sentry: " IF you see something that looks unfamiliar to you, AND further investigation does not resolve the issue OR you have residual doubts, THEN sound the alarm." We can make complex behavior by utilizing such simple mechanisms especially if you incorporate "fuzzy-edged" IF-clauses.

But how does such a system deal with internal conflict resolution? Of course, first remember that all these production type models have some basics, a workspace for demon interaction, and a memory for innate and accumulated information. The Hard Question of "and then what happens" does have some candidates for conflict resolution:

1. Degree of match- a better match has higher priority
2. Production strength- recently successful productions have a higher priority
3. Data refractoriness- the same production cannot match the same data more then once (to prevent infinite loops and similar, if less drastic ruts)
4. Specificity- if two productions match the data, the more specific IF-clause wins.
5. Goal dominance- among the items that productions can deposit in working memory are "goals", there can only be one active goal at a time in working memory and any production whose output matches the active goal has priority

To other models such as SOAR, impasses are the basic production opportunities in the system. Conflicts are not dealt with using a rule based system but non-automatically by creating a "problem space" in which the impasse is the problem to be solved. This can go on potentially forever but in all models so far, the impasse is resolved after a few layers, dissolving the proliferation of "problem spaces" after making a non-trivial exploration of the possibilities. With the "newly minted" production on hand, future similar impasses are automatically solved.

Dennett presents these different models and variations (explored in much greater detail in the book) not to compare the models so much as to give the reader an idea of the sorts of issues that need to be dealt with. His feeling is that all current models are still too simple, but that the solution lies in more complicated models that will also be tested by "building them and running them."

As Dennett points out, the real philosophical question is could ANY of these models explain consciousness, but it would be premature to pin our hopes on any one model at this time. Though he goes out on an empirical limb in Appendix B where he outlines some experiments that could falsify his ideas on consciousness. Of the experiments that have been described, only the "change-blindness" prediction has been tested that I know, but as counter-intuitive as is seemed at the time it was proposed, it has been strongly upheld.

Regardless, Dennett goes on to show that philosophical ideas can benefit from utilizing what we already know about how mere mechanistic subsystems can convincingly simulate biological mental activity at some levels. In any case, however the shift has occurred from von Neumann to neural connectionist architectures, at the heart of the most volatile pattern-recognition system (connectionist or not), "lies a von Neumann engine chugging along computing a computable function." Now that AI has created these non-rigid, fuzzy, holistic systems, the critics of AI have to either declare that these connectionist systems were what they had in mind all along or raise the stakes and declare that no connectionist system could he “holistic” or “intuitive” enough for them. Two of the best known critics of AI have split on this issue. Berkeley philosopher Hubert Dreyfus has sided with the connectionists while his colleague John Searle insists that no connectionist computer could exhibit real mentality.

So although some skeptics are in retreat, huge problems remain. The largest one according to Dennett is the idea of the globalness of the work space. All these boxes and diagrams in these diagrams, all share the same anatomical space in the brain. How could this be?

> “Suppose you learn to make cornbread, or learn what “phenotypic” means. Somehow the cortex must be a medium in which stable connection patterns can quite permanently anchor these design amendments to the brain you were born with. Suppose you are suddenly reminded of your dentist appointment, and it drives away all the pleasure you were deriving from the music you were listening to. Somehow the cortex must be a medium in which unstable connection patterns can rapidly alter these transient contents of the whole “space”- without of course, erasing long term memory in the process.”

Of course one can imagine functionally distinct networks that interpenetrate, like telephone and power lines, but the deeper issue is how individual specialist demons, recruit others in a larger scale enterprise. We already have models for how specialists get recruited for specialist talents, but what if the specialists are themselves sometimes recruited as generalists, contributing functionality in which their individual talents have no significant role?

It is tempting as Dennett says to suppose that informational content can arise in some specialist sense, and can then “be propagated across cortical regions exploiting the variability in these regions without engaging the specialized semantics of the units residing there.” But this idea has not been successfully modeled yet (as of 1991).

Yet, this is not surprising in a way. Human engineers with “imperfect foresight train themselves to design systems in which each element plays a single role, carefully insulated from interference.” This is done to minimize unforeseen side effects of operation. But Mother Nature, has no such worries, and can capitalize on unforeseen side effects when they arise and if they are useful (once in a blue moon). Humans, not used to assigning multiple roles to available elements, have conceptual difficulties with this.

Of course some philosophers (mysterians) will say that this is exactly why the brain can simply not understand itself, but Dennett believes that it is not impossible, just fiendishly difficult for the above reasons. In fact neurophysiologists have identified mechanisms in neurons which are plausible candidates to play the role of rapid modulators of connectivity between cells. These gates might permit swift formation of such “transient” coalitions, which could be superimposed on networks without requiring any alteration of long term synaptic strengths associated with long term memory.

The anatomical map is filling in. It is now also known that the reticular formation and the thalamus play a crucial role in arousing the brain from sleep or emergency while the frontal lobes are now known to be responsible for long term control and sequencing of behavior. New information is constantly coming to light. But we will never find a central “boss” in the brain.

Part 4: Powers of The Joycean Machine

So, based on our sketch so far, we have a competition (internal natural selection of sorts?) among many contentful events in the brain, and in which, a select subset of such events gain global influence by forming sympathetic coalitions. Some events, might unite with language production demons and eventually results in verbal sayings (both aloud and internally), others might lend content to other subsequent self-stimulation such as diagramming to oneself. The rest die out almost immediately, leaving only faint traces. But how does getting to the next round of self-stimulation (this charmed circle) give us consciousness or more specifically a Joycean stream of consciousness?

(The following is painfully brief without much of the supporting argument, so a complete reading of chapter 9 in CE is recommended for those that want to really grasp the ideas presented.)

The first point Dennett tries to make is that there is no sharp dividing line between being “in consciousness” from events that always stay “beneath or outside consciousness.” See Allport, 1988 for further discussion of this. In any event, if Dennett’s theory of the Joycean machine is going to be convincing it had better make “the next round of self-stimulation” something remarkable, because intuitively it sure seems remarkable!

Next we should ask if these proposed mechanisms can explain that conscious function, whatever it is eventually determined to be. Being mindful that consciousness might have multiple functions, functions that are not well served by existing features, let’s see what we have:

How do we get “self-control” over the proliferation of concurrently active specialists? Clearly simple or rote tasks can be performed without the enlistment of additional forces and hence unconsciously, but difficult or unpleasant tasks require “concentration,” something we accomplish with the help of self-admonition and various other mnemonic tricks, rehearsals and other self-manipulations. We might talk out loud, a throwback to the crude but effective strategies of which our private thoughts are sleek descendants.

Such self-control strategies govern our perceptual processes. For example, although our visual systems are natively designed to detect some sorts of things- the sort of things that just “pop out” at us when we “just look”, other sorts of things we can only identify if we deliberately set up a policy by an act of self-representation. A red spot among a slew of green spots will stick out like a sore thumb (actually like a ripe berry among some leaves), but finding a red spot among a set of multi-colored ones is a task that requires serial searching with extended self-control. So representing things to ourselves permits us to set up serial tasks that other animals do not even approach.

> “We can work out policies well in advance, thanks to our capacity for hypothetical thinking and scenario-spinning, we can stiffen our own resolve to engage in unpleasant or long-term projects by habits of self-reminding, and by rehearsing the expected benefits and costs of the policies we have adopted.”

And learn from these efforts by remembering rehearsed strategies and what went wrong by recollection - “well I mustn’t do that again!” But as Dennett goes on to describe, this memory-loading habit is only one of many valuable effects that occurs in the brain. Broadcasting effects allow any of the things one has learned to make a contribution to the current workspace. These events provide context to events “in consciousness” that themselves are not conscious.

Part 5: But is this a theory of Consciousness?

Dennett admits that he’s been coy until now. He hasn’t yet claimed that a Joycean machine is conscious or that any particular state of such a virtual machine is a conscious state, because he wanted to avoid arguing over what consciousness is until he could show that many of the presumed powers of consciousness can be explained by the powers of a Joycean machine. Whether or not one believes that such a machine endows its host with consciousness.

But a la zombie, “couldn’t there be an unconscious being with an internal global workspace in which demons broadcast messages to [other] demons, forming global coalitions and all the rest? If so, then the stunning human power of swift, versatile adjustment of mental state in response to almost any contingency, however novel, owes nothing to consciousness itself, but just to the computational architecture that makes this intercommunication possible.” As Dennett puts it “If consciousness is something over and above the Joycean machine, I have not provided a theory of consciousness at all, even if other puzzling questions have been answered.”

But Dennett is willing grab the bull by the horns and declare that anything that has such a virtual machine as its control system, is conscious in the fullest sense, simply because it has such a virtual machine.

The next chapter will deal with the usual objections, zombies and qualia. The “but I just know that I can imagine an entity with all that, those functional processes, and it still won’t have consciousness” sort of objections, to which Dennett responds “Oh, can you? How do you know? How do you know that you’ve imagined “all that” in sufficient detail, and with sufficient attention to all the implications?”

He asks us to consider a similar objection by some present-day Vitalist:

> “That’s all very well, all that stuff about DNA and proteins and such, but I can just imagine discovering an entity that looked and acted just like a cat, right down to the blood in its veins and DNA in its “cells,” but was not really alive.”

Why is this not a good argument? Because the effort of imagination doesn’t count against the account of life presented by contemporary biology. Dennett agrees that cognitive science today is in a position to shift the burden of proof to the mysterians. Our intuitive convictions only count for so much and it’s simply not enough to say one can imagine otherwise.

By Probeman