Part -I

Dennett's argument against us being brains in a vat rests on the impracticality and complexity of the situation. This does not really add anything new, I think-everybody would be in agreement on *that*. The question is that of principle. [The introduction of combinatorial explosions leads into an interesting discussion on hallucinations, though: Hallucinations avoid combinatorial explosions by borrowing from the real world.]

(Later on in the book, I see that the impossibility of things beyond a certain complexity to exist in a certain turns out to be a major theme, so perhaps I cannot dismiss this argument so lightly.)

Hints of Dennett's 'scattered-consciousness' theory are scattered throughout this part. Subconscious preparation for the mindset of the reader, perhaps.

"If doing things that matter depend on consciousness, mattering should depend on consciousness as well." Page 31, Chapter 2: A foreshadowing of how he will deal with qualia.

Dennett also briefly destroys dualism. (The basic issue is of how mind and matter may interact.)

Dennett says that the empiricist claim that we imagine a purple flying cow by taking purple from a grape, wings from an eagle, and cow from a cow, cannot be quite right, because "what enters the eye is electromagnetic radiation, and it does no thereupon become usable as various hues with which to paint imaginary cows". This is an oversimplification, though, and I do not see any contradiction between Dennett's vision of qualia and the empiricist claim. Dennett says the radiation is "transduced"-all we need to do would be to re-transduce, or inverse-transduce.

"We can come up with a perfectly sound biological account of why there should be pain…but we can know in advance that it won't satisfy anybody. Any biological account of pain leaves out the intrinsic awfulness of pain. For the time being we can note that any account of pain that left in the awfulness would be circular."

Dennett says that people will be unsatisfied with a purely biological account of, say, pain, because they would complain that the intrinsic awfulness, the "painness" has been left out. He then goes on to poke a hole in this and say that if pain had been explained in terms of painness, then we were back where we started, because nothing would be explained at all and it would be circular. This is a subtle shift made by Dennett because in fact what people are clamouring for is not an explanation of pain in those undecidable terms, but an explanation of both pain and that vague undecidable term which actually expresses its undecidability.

"My point is not that you have no privileged access to your conscious experience, but just that we should be alert to very tempting overconfidence on that score."

Dennett says that we should be less sure about our own conscious experience and belittles introspection, because what we see to be red, we may later say is green, etc. due to tricks played on our senses. While this is correct, there is another part to our conscious experience that we cannot be unsure about. Even if the colour we see is not actually red, the experience of seeing red is very much real! Even if a certain "phenom" is an illusion, there is no less mystery to it-it is an illusion that must be explained. The illusion would then be the mystery. Incidentally, this is something John Searle repeats on numerous occasions in his own book. Introspection, in this way, gives itself importance.

Dennett gives a short overview of the empirical methods he deems acceptable to use in order to unravel the mystery of consciousness. There is not much to remark regarding this, for it is a tried and tested Quinean study of how to build one's theory, but it is rather pleasing to read.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Part-II

Dennett introduces his theory of consciousness with a flourish.

Reminiscent of Quine's exposition on quality spaces, Dennett's claim is that there is spatiotemporal smearing across events that occur with extreme spatiotemporal proximity. However, while Quine's conclusion was that there are discrete qualities which we cannot distinguish between, Dennett's conclusion is that qualities lose their discreteness and individuality at such proximity. And since any one quality always has many qualities at such proximity, *no* quality possesses individuality. ('Quality' is a substitute for 'experience' here, so what Dennett is trying to say is that there is no privileged 'stream of consciousness'. I would like to note that this is *not* in blatant defiance of Carnap's (and Quine's) 'rational reconstructions', as it may initially seem to be. Carnap is still allowed to take qualitative stimulus as a building block, regardless of whether or not it turns out to be illusory.)

The quality/experience becomes a function of the spatiotemporal point at which probing occurs-not in the obvious sense, but in the sense that the same provided stimuli may elicit differing perceptions/experiences/quality if the probing occurs at slightly different times.

[A deep interdependence had arisen which was attempting to slip past us unnoticed. We receive stimulus and an experience of that stimulus requires a further stimulus (since conscious experience has become a function of probing). Where does this "Second-order stimulus" lie in our scheme of things? Dennett does not address this potential infinite regress anywhere.]

Dennett suggests two ways in which the brain may cause this difference in perception.

One possibility is that we experience said event with unaltered stimulus, and then the brain quickly alters our memory to make us remember it differently. (Orwellian.)

Another possibility is that the brain alters the stimulus before we "experience" it, and we experience only the altered stimulus. (Stalinesque.)

What content "crossed the finish line" first? Altered stimulus or unaltered stimulus?

Dennett shows pretty convincingly that there is no way to confirm any one of these over the other. (Reminiscent of Quine's indeterminacy of translation: Two equally correct theories.) His point is that one cannot "draw a line" as to when a person becomes conscious.

(However, just because it is impossible to empirically draw this line does not mean we can conclude that such a line does not exist. If one takes a peep at the further parts of his book, however, one can strengthen the argument by saying that since there are no dispositions relevant to drawing this line, and since Dennett's final posit is that qualia *is* dispositions, the conclusion would be that there are no qualia that can be said to draw the line.)

Dennett briefly shows some explanations of 'seeming' and 'experiencing' in terms of projecting qualities outwards onto a phenomenological world. He must have tread carefully here, for his own theory of the phenomena relies on combining internal mechanism with external. However, he wraps these explanations up by saying that positing a phenomenological world and entities needlessly multiplies entities, although he leaves us waiting for his own alternative for a good amount of time.

Dennett now, I feel, makes a first stab at the binding problem: If consciousness and perception in the present is so scattered, why and how do they, in the past, seem like such a unified stream?

To this end, he makes the point that just because an event A temporally precedes an event B does not imply that the brain must represent event A before event B for us to experience it as such.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

However, there seems to be something intuitively correct about temporally ordered events having temporally ordered representations. This intuition, says he, is backed by two valid arguments. To achieve the temporally correctly ordered experience, order of representation matters until the brain has decided the order of the events, after which it may relay the information at its leisure. The information of the temporal order of events is required for the brain to decide the order of events, and sometimes that information can only be received from the order of the representation.

The other is that, in the case of deadline-sensitive events, the representing must occur before the deadline has been crossed.

Dennett gives us the example of Libet's experiment, the results of which-he tries to show-seem to be pretty strongly in favour of this theory of temporal experience.

Another of Libet's experiments seems to indicate that conscious knowledge of an action occurs only nearly half a second after the brain has made the decision to perform the voluntary action. This, however, says Dennett, is the Stalinesque interpretation. The Orwellian alternative is that we make the decision consciously and forget about it. (Why would we forget? Dennett does not attempt to answer this, but Marvin Minsky has suggested that introspection is difficult due to the enormous amount of brainspace it takes. Similarly, conscious awareness of the decision may be erased in order to accommodate for the execution of the decision.)

From all this, the conclusion drawn by Dennett is that the attempt to pin down the temporal instant of consciousness is incoherent.

Dennett makes a neat observation: It is easier to predict the behaviour of a device one synthesizes than to analyse the external behaviour of a "black box" and figure out what must be going on inside. Although it is left as an offhand comment, it seems to me to have relevance in questions like why philosophical and mathematical questions are so easy to ask and philosophical and mathematical answers so hard to give: Philosophical questions are equivalent to devices we synthesize, while philosophical answers can be given only by doing something equivalent to analysing the external behaviour of black boxes.

In any case, Dennett thus proposes a trip through the process of evolution which gave rise to the beings we are now.

A very Dawkinsian lowdown of our evolution is given.

Dennett says that the dividing line between active learning and evolutionary development is irrelevant for our purposes. It seems to me, however, that this line is consciousness itself, unless he is using the words in a different sense than I think he is, which is what I suspect may be happening.

Regardless, he ploughs on. "Postnatal design-fixing", says he, must be accomplished by a process very similar to the one which fixes prenatal design: In other words, natural selection *within* the individual. Operant conditioning is a great example.

Dennett speaks of the Baldwin effect, wherein organisms are able to imbibe 'Good Tricks' from those in their vicinity due to their plasticity. Therefore, plasticity is not only a product of genetic evolution by natural selection, it can also be used to reflect back on genetic evolution and speed it up. He suggests this has had a part to play in human evolution.

He resorts to something very much like Minsky's society of agents to resolve the problems of higher-level control that come with plasticity.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

(Also, looking at the kind of examples that have been given, it seems to me that the importance of language can be justified by the fact that it is indispensable to our plasticity.)

Dennett exposits briefly on autostimulation, following which he introduces the Dawkinsian concept of memes. He wraps it up by making the interesting observation that it is senseless to speak of 'protecting' ourselves from memes, for it is memes themselves which have played a major role in making us who we are. It is not genes vs memes: If a meme is a good replicator, what other eternal objective ideal is there to judge it against? He concludes by saying that memes and the Baldwin effect/plasticity have accelerated genetic evolution and in turn accelerated the advent of what we call 'consciousness'.

(This chapter (apart from the last upcoming segment) is my least favourite in the book for the simple reason that, although I see how the background may be required in such a book, it felt rather out of place to me when looked at in the context of the rest of the book. I think it has to do with the fact that it is a lone unoriginal interlude in a book brimming with such great originality.)

Dennett asserts in an offhand manner that we will probably never be able to pin down functional differences to the neurological level. This seems to me not to be in line with his philosophical views- it directly implies that consciousness is an epiphenomenon of some sort, something he violently rejects later on.

A higher-level description which draws an analogy to computer software is suggested.

The concept of virtual machines is introduced.

Computers have a great degree of plasticity. "A virtual machine is what you get when you impose a particular pattern of rules on all that plasticity." Different programs endow computers with different details.

Turing breaks the working of a computer with a serial architecture down to five basic components: *A serial process in a severely restricted workplace to which both data and instructions are brought from an inert but reliable memory there to be operated on by a finite set of primitive operations.* (Dennett lists out the primitive operations.)

"This sequence of actions could…. incorporate all rational thought and perhaps all irrational thought as well." This is acceptable to me.

"What we have to understand is how a Joycean/von Neumannesque serial phenomenon can come to exist….in the parallel hubbub of the brain." And so we are back at the binding problem.

Dennett makes his second stab at it.

We know by introspection that there must be something von Neumannesque going on in ourselves. That is, in fact, why the binding problem is a problem.

Now, Turing's greatest achievement is proving that a von Neumman machine can, in principle, do anything a parallel one can do. The only difference is the time taken for the computation. From this it follows that the brain can be simulated by a serial von Neumman machine.

Dennett's suggestion is that the human brain is just such a serial von Neumman machine "on the parallel hardware evolution has provided for us." Just like how parallel machines are simulated by serial ones, our brain is a parallel machine simulating a serial one.

[Simulating for whom? Did Dennett himself accidentally slip into the Cartesian theatre?]

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Dennett admits that there are many “disanalogies” with this comparison, but I think it is a very good one overall.

Now comes the ace in the hole. While a von Neumann machine’s workings aren’t visible to itself, the workings of our Joycean machine is visible to itself. This makes it an object of its own perceptual system!

[Simulating for itself, then. The theatre has been avoided.]

It is not surprising to me that Dennett, having worked closely with Hofstadter, ended up incorporating a strange loop in his thesis. This bootstrapping is not a vicious one, however.

And so Dennett concludes (slightly paraphrasing):

“Human consciousness is a huge complex of meme-effects best understood as the operation of a virtual machine whose powers are vastly enhanced by its organic hardware. Moreover, many of its most curious features and limitations can be explained as the by-products of the kludges that make possible this reuse of an organ.”

Additionally, he suggests that many of the kludges may be found in autostimulation.

Most of this exposition is rather scientific in nature, concerned with the physical aspect of how consciousness happens, and thus laying a sound, scientific foundation to what had been said regarding spatiotemporal distortion and what will be said regarding qualia. This is a solid foundation to work with, although there is some work yet left explaining exactly how our von Neumann-like machine works, complete with all its spatiotemporal perceptual distortions. He comes back to exactly how this mechanism “gets anything done” in chapter 9.

“If there isn’t a Central Meaner, where does meaning come from?” Again, the binding problem.

Theories of speech production/language comprehension which rely upon a Central Conceptualizer run into an infinite regress, for the Conceptualizer must give directions to the agents in some language of the brain-and where does the Conceptualizer get the blueprint for *this* language?

We do not encounter this issue when speaking of the comprehension processes of programs in computers because we, the designers, give them the first blueprint, thus avoiding the regress.

What is at the other end of this bureaucratic extreme which gives such ominous power to the Conceptualizer? Dennett shows us the Pandemonium alternative: Our mind throws at us all kinds of verbal concoctions, ranging from ‘REEEEEEE!’ to ‘Hello, how are you?’ to ‘That chair is a human book!’

‘The problem with the Pandemonium alternative is that we need to find a way in which sources of content can *influence* the creative energies of the word-demons without *dictating* them.’

Unfortunately, all Dennett does is point a finger towards Minsky’s society of agents for a satisfactory model for representing such phenomena, and although he has convinced me, at least, on the *logical* possibility of this alternative, the lack of a concrete model from Dennett means that we have made no notable progress on the matter of the binding problem.

Where exactly our brain lies between bureaucracy and pandemonium is an empirical matter, concludes Dennett.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Now, in a pandemonium scenario: Why would our brain conjure up all these garbled phrases? In fact, one may think: Why this garbled phrase? Why not this one? What makes our brain throw any phrase towards us in the first place?

Dennett suggests that it is not our brain doing this job, but that *the phrase itself* may be seen to have an active goal of wanting to be said. He gives the example of how a catchphrase spreads like wildfire. This is not very dissimilar to how genes may be said to be actively wanting to survive-phrases are, after all, memes.

Our brain does not go hunting around for material-the material itself waits for opportunities to be said.

What does it mean to speak of a phrase trying to be spoken, though? No light is shed over here. Before this, the problem was regarding how the brain selects phrases. Now, the problem is regarding how the phrase selects itself/how the phrase tries to get itself spoken. (It is not very difficult to see that many plausible mechanisms can be devised for this, including autostimulatory ones. As mentioned before, it may be much like a gene surviving its own self. If the phrase is one that (mechanically) enhances its own capability to be selected, much like a gene that (mechanically) does something to increase its own survival chances, it will be selected more often.)

“The most accessible or available words and phrases could actually change the content of the experience.”

Dennett gives a few examples of a person simultaneously performing contradictory speech acts in order to illustrate the lack of a Central Meaner and add empirical evidence to a pandemonium-oriented theory. “We can note that the brain’s machinery is quite able to construct apparent speech acts in the absence of any coherent direction from on high.”

Dennett directly mentions the binding problem only in an offhand manner, saying that the problem is misconstrued because it works under the assumption that everything comes together at a point. I think this is a misunderstanding regarding what the problem is. The real problem, according to me, need not the binding itself, but the *appearance* of binding. If consciousness is so scattered, why does it *appear* so unified? Dennett has made similar statements previously and I gave the same comments.

Now, Dennett returns to develop his von Neumann analogy a bit more, asking how exactly it “gets anything done”. He indulges in an interesting discussion on whether a “demon” can have a dual specialized as well as generalized role, eventually referring to Minsky once again (whose book really does have some solid ideas on these matters) for the only good theory of the sort that he has seen.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Part-III

Dennett offers an alternative to the hypothesis that thinking involves real images in a Cartesian theatre, making notes of our brain's inefficiencies and shortcuts regarding visualization as he does so.

The alternative offered was an empirical one. Dennett suggests that *abstract structures* "lay down some of the tracks on which 'thoughts' can then travel". (This was strongly hinted at when he previously drew the analogy between virtual machines and brains.) (One ought to note that although a satisfactory medium for thought has been posited, no explication has (yet) been supplied for what thought itself is.)

(I consider 'abstract structures' here to be interchangeable with *languages*.)

Dennett now constructs a paradox out of us via the Cartesian theatre: To express a *belief* (underlying dispositional state), we need to have a *thought* (occurrent/episodic states/transient events) about the belief. However, we now need to express this *thought*, and to be conscious of it, we need to have another thought…

The paradox is destroyed only by stepping out of the theatre and asserting that we are, in fact, (typically) not directly conscious of the first-order thought. (We are usually conscious of the belief we wish to express, and not of expression itself.) (Shades of Minsky speculating about the difficulty of introspection.)

"What distinguishes a conscious state from a non-conscious state…is not some inexplicable intrinsic property, but the straightforward property of having a higher-order accompanying thought that is about the state in question", says folk psychology.

There are issues with that view of folk psychology. If a robot could know that it was in a certain "unconscious" mental state, does that mean it is conscious? Dennett asserts that this means *it* thinks it is conscious.

Now, what's wrong with this picture?

It has it all backwards, says Dennett. We do not first become aware of our belief in a thought and then express the thought (usually without being aware of the thought itself)-rather, the thought itself is synthesized alongside our report of the belief. "The only way for a human brain to get itself into something like a higher-order belief state…is to engage in the process rather like reporting first-order states itself."

This manages to avoid the zombie paradox in the previous section.

Dennett claims that it also circumvents the errors possible between the belief and its expression, preserving 'subjective intimacy'. He points out an infinite regress between the two, with all the errors adding up. However, basic calculus tells us that the sum of an infinite number of terms may well be finite. As far as errors in reporting are concerned, then, I do not see any improvement precipitated by Dennett's theory.

Dennett introduces the phenomena of blindsight: An occurrence in which a person is able to relay events without 'consciously' experiencing them.

He gives examples of phenomena happening right in front of a person without the person 'realizing' or 'perceiving' its occurrence, and suggests a pandemonium-style explanation for it, wherein only a select few of the multitude of phenomena occurring in front of the person get perceived. This does

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

not, however, explain the missing "phenomenal qualities" between realization and the lack of it in a blindsighted person. Dennett delays this point to the next chapter, although he briefly suggests that the rate of information flow may have a part to play in it.

Dennett suggests an alternative to the typical notion of 'filling in' phenomenal qualities in blind spots and the like. In the conventional view, our brain will, say, 'fill in' the blind spot with blue if we are seeing something blue. He says that the brain doesn't have to 'fill in' the blind spot, and unless "the brain received contradictory evidence from some region", it just concludes that in the blind spot, there is 'more of the same'. 'Filling in' consists not of the brain providing something but rather of it ignoring something-ignoring the blind spot due to a lack of epistemic hunger due to the death of the homunculi who are supposed to care about sense-data from that region. "The discontinuity of consciousness is striking because of the apparent continuity of consciousness."

We now move onto the very heart of Dennett's theory of consciousness and the segment which I consider to be the most crucial in his work.

I have been pointing out Dennett's lack of explanation regarding phenomenal qualities and the binding problem throughout. This is where he gets to it…by denying its existence.

Just because qualia seem to exist, says he, does not imply that they do exist. Granted, but once again, it does not matter whether or not they are illusory-they must still be explained, for we undeniably experience them first-hand. (I think it is interesting to bring up Minsky's speculation here-that the 'soul/self' may merely be abstractions invented as a by-product of evolution for survival purposes.)

"There seem to be qualia, because it really does seem as if science has shown us that the colours can't be out there, and hence must be in here." Dennett follows this up by saying that there is no *functional* need to posit qualia (and I agree with him on this).

The essential thesis Dennett presents to us next is this: Qualia and dispositions are inseparable. One exists on the basis of the other. Our aversion to red is inseparable from the phenomenological quality of red. And so, concludes he, one cannot speak of the existence of qualia independent of everything else, for they are inescapably bound to our dispositions.

A lot of the rebuke Dennett has received ('Consciousness Explained Away', etc.) seems now to me to be a tad unfair and misunderstood. Dennett is not outright denying qualia; they can be spoken of, but only alongside dispositions.

This is a very functional attitude, with assumptions similar to the ones Quine adopted in his arguments against radical translation. Penrose concluded via Godelian arguments that consciousness could not possibly be simulated on a machine. However, Dennett's claim is that one need not simulate consciousness at all; if one merely has all the outward dispositions down, the machine can be said to be conscious!

This is the kind of rebuttal John Searle gave Penrose. He argued that in principle a machine could be constructed whose blind external behaviour is the same as that of a human. (Penrose was not talking about such blatant behavioural copying, of course.) However, Searle takes it as axiomatic that such a machine cannot be conscious-as axiomatic, it follows, that Dennett is wrong, for this is exactly what Dennett is trying to argue against.

Perhaps we can arrive at a Dennett-centric agreement between these three.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Why did I dismiss Searle's rebuttal of Penrose? Penrose meant not only external dispositions but also trains of thought and other such 'mental' activities, which Searle does not consider. Now, in Dennett's sense, the word 'disposition' certainly includes all these 'mental' predilections. We can now consistently maintain that a simulation of every single human disposition implies consciousness, while at the same time claiming that this simulation cannot be performed algorithmically in principle-in my reading, Dennett did not raise any serious objections towards this. There is a difference between humans and machines, but the difference becomes merely one of dispositions.

We need to be careful with how we use the word 'disposition', it has become of paramount importance. The experience of seeing red is not a disposition; the aversion that follows is. Dispositions are reactions. An external beginning is required.

But wait-if not all dispositions are algorithmic, where do these 'non-algorithmic' dispositions lie in our scheme of things?

All the old problems come crashing back down upon us if we try to divide dispositions into algorithmic and non-algorithmic ones, but Penrose's argument holds good. What can we do now?

I am unable to think of any way to resolve this lone extremely major problem, so I must let is pass for now.

Turning towards Dennett's methodology: While he impressed upon the reader the virtues of using an empirical third-person perspective, this last, most important leg has been based almost entirely on introspective methods. Introspection is where the problem of qualia was born! Why should we trust this method now?

It works in a way because introspection is not used here to prove an assertion but, rather, to show the limitations of introspection itself; to show that another assertion is unfounded, and to show us that the content of this assertion (qualia) is unspeakable of-and finally, to show its lack of individual and independent existence due to this.

So what is hot, cold, colour, spice, sour, pain? A collection of dispositions. The way we feel about something is the way we react towards something.

Dennett returns to the theme of a unified self being 'invented' for evolutionary survival purposes. He comments on Nagel's arguments about the impossibility of thinking in certain ways, and says that it must be false, for have we not succeeded in freeing ourselves from the traditional ways of thinking?

Certainly, but one could easily interpret Nagel as speaking of manners of thinking which are beyond even Dennett's radical approach. I believe it certainly is impossible to construe of certain things in certain manners, although it need not necessarily affect our concerns. He attacks Nagel's main thesis ('What It Is Like To Be A Bat') using his conclusions regarding qualia and dispositions (in this case, speaking of those of a bat).

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.