

Dennett in a Nutshell

The following is a brief outline of Dan Dennett's positions regarding the "problem" of consciousness. For more detail see his popular expositions: "Consciousness Explained", "Freedom Evolves" and "Darwin's Dangerous Idea- Evolution and the Meanings of Life". See also the Dennett chapter summaries and discussion threads:

<http://forums.philosophyforums.com/thread/6031>

<http://forums.philosophyforums.com/thread/6331>

1. Naturalism: First and foremost Dennett will simply not entertain any appeals to magic. These ideas come under a variety of headings with a full spectrum of philosophies from New Age quantum consciousness to religious belief in the immaterial soul.

Therefore Dennett allows: no dualism (properties or substances), no vitalism, no soul, no epiphenomenism, no essentialism, no intrinsic supernatural or metaphysical forces or processes and most of all- NO SKYHOOKS!

What does Dennett mean by a "skyhook"? The concept of "skyhook" has an uncertain origin, though Dennett cites an anecdote of "...an aeroplane pilot commanded to remain in place (aloft) for another hour, who replies: 'the machine is not fitted with skyhooks' "... Dennett goes on to explain "The skyhook concept is perhaps a descendant of the deus ex machina of ancient Greek dramaturgy: when second-rate playwrights found their heroes into inescapable difficulties, they were often tempted to crank down a god onto the scene, like Superman, to save the situation supernaturally. ... Skyhooks would be wonderful things to have, great for lifting unwieldy objects out of difficult circumstances, and speeding up all sorts of construction projects. Sad to say, they are impossible." (DDI, p74). Ultimately, skyhooks are the Lockean "mind-first" processes that often disregard that apparent design is the result of mindless mechanism.

On the other hand a "crane", according to Dennett, acts as an adaptation or learned "good trick" which builds on an existing foundation in order to speed up design (however that is defined). Cranes take into consideration the mechanistic algorithmic (and heuristic) nature of life processes and build from the bottom up, and are examples of good reductionism. Dennett rightly points out that "where a mystic invokes skyhooks, a greedy reductionist tries to do without cranes."

"Darwin's dangerous idea", says Dennett- is a really good crane basically, "promising to unite and explain everything in one magnificent vision." But don't assume that "skyhooks" are just for the supernaturalists. As Dennett points out in his preface: "I have found not just lay people and religious thinkers, but secular philosophers, psychologists, physicists, and even biologists who would prefer, it seems, that Darwin were wrong.". Dennett's term "skyhooks" takes aim at those whose intuitions cry out against "mere" mechanism. In short, all those who cannot conceal their discomfort with Darwin's dangerous idea, ranging from nagging skepticism to outright hostility.

In any case, merely observing our current lack of knowledge in the study of the human mind and applying an intuitively appealing or magical name to what we don't understand does not constitute an explanation. Yes, God (or whatever) could very well be "in the details", but since we cannot, in principle, find evidence for or against such a position, there is no point in such unmotivatable speculation.

Religious convictions aside (and how is revelation different from introspection?) Dennett's embarrassment over some philosopher's infatuation with such ideas is clearly evident in this quote from "Get Real- Reply to 14 Essays" 1994.

<http://ase.tufts.edu/cogstud/papers/getreal.htm>

"It cannot have escaped philosophers' attention that our fellow academics in other fields-- especially in the sciences--often have difficulty suppressing their incredulous amusement when

such topics as Twin Earth, Swampman, and Blockheads are posed for apparently serious consideration. Are the scientists just being philistines, betraying their tin ears for the subtleties of philosophical investigation, or have the philosophers who indulge in these exercises lost their grip on reality?"

Ultimately Dennett simply asks that philosophical investigations be both informed by science and that they themselves inform science. By this he means that philosophers should try to make sure their speculations are compatible with what we already know about how the brain and mind operate (which is not inconsiderable at this point), and also by trying to create ideas that are at least testable in principle by science.

The following is an example of what needs to be done more by philosophers if they want to get "in on the action" (again from "Get Real- Reply to 14 Essays"):

"...the answer is that to a surprising degree, the visual part of your brain is more like a *camera obscura* than you might have thought. On the last page (468) of *CE*, I described an experiment with eye-trackers that had not been done, and predicted the result. The experiment has since been done, by John Grimes (forthcoming) at the Beckmann Institute in Champaign Urbana, and the results were much more powerful than I had dared hope. I had inserted lots of safety nets (I was worried about luminance boundaries and the like--an entirely gratuitous worry as it turns out).

Grimes showed subjects high-resolution color photographs on a computer screen, and told the subjects to study them carefully, since they would be tested on the details. (The subjects were hence highly motivated, like Betsy, to notice, detect, discriminate, or judge whatever it was they were seeing.) They were also told that there *might* be a change in the pictures while they were studying them (for ten seconds each). If they ever saw (yes, "saw," the ordinary word) a change, they were to press the button in front of them--even if they could not say (or judge, or discriminate) what the change was. So the subjects were even alerted to be on the lookout for sudden changes. Then when the experiment began, an eyetracker monitored their eye movements, and during a randomly chosen saccade changed some large and obvious feature in each picture.

(Some people think I must be saying that this feature was changed, and then *changed back*, during the saccade. No. The change is accomplished during the saccade, and the picture remains changed thereafter.) Did the subjects press the button, indicating they had seen a change? Usually not; it depended on how large the change was. Grimes, like me, had expected the effect to be rather weak, so he began with minor, discreet changes in the background. Nobody ever pressed the button, so he began getting more and more outrageous. For instance, in a picture of two cowboys sitting on a bench, Grimes exchanges their heads during the saccade and still, most subjects don't press the button! In an aerial photograph of a bright blue crater lake, the lake suddenly turns jet black--and half the subjects are oblivious to the change, in spite of the fact that this is a portrait of the lake. (What about the half that did notice the change? They had apparently done what Betsy did when she saw the thimble in the epistemic sense: noted, judged, identified, the lake as blue.)

What Dennett is asking is that at some point we might need to abandon our intuitively informed philosophical conceptions and rely on reality motivated methods that actually yield results. I'll grant you, sure, it seems to us all that we see our whole visual field in color, but a simple experiment shows that we are simply wrong in this sincere conviction. Our heartfelt introspective "experiences" are no more (or less) reliable than our heartfelt perceptions. Dennett says of this (again from "Get Real"):

"I'm all for open-mindedness and scientific optimism, but surely Rey, Flanagan, and the others would agree that there are *some* occasions when the jig is up--when it is just silly to hold out hope for such a scientific revelation. Consider Einsteinian physics. Einstein noted that it is impossible to distinguish by local observation between a gravitational field and an accelerated frame of reference. This led him to postulate the equivalence that is at the heart of relativity theory. Now insert the "realist," who says "Oh just because you can't distinguish the two doesn't mean they aren't different! There might be a difference that is indistinguishable by any current test! Never say never!" Yeah, there might be, but in

the meantime, tremendous progress is made by concluding that there isn't (Cf. Field, 1974, 1975). I am proposing similar simplifications: since you can't distinguish between the Orwellian and Stalinesque models of meta-contrast, or between a zombie that acts just as if its conscious and a conscious being, they are equivalent. “

- 2. Appropriate Levels of Explanation: related to the above is Dennett's position that explanation must be reductionist in principle, but non-reductionist in practice. This is what is meant by “normal” science. For example, even though we can't in practice calculate the “wetness” of water from quantum principles (the calculation is simply too large on existing computers), we still can agree that ultimately, quantum mechanics does provide a complete explanation for the observable properties of H₂O.**

Dennett often refers to “greedy reductionism” which would be, for example, the case were one to claim that we must be able to provide a full quantum description of biology to say that it had really been “explained”. That is, if we fail to provide a completely reductionist explanation for biology down to the atomic level or further- we've somehow “left something out” (like what? vital substances, chi, prana, water “memory”, etc).

Dennett would argue that each level of complexity has an appropriate level of explanation. To attempt to go beyond that is fruitless, even if it is possible in principle. As he has said, “...sometimes an impossibility in fact, is theoretically more interesting than a possibility in principle”.

Dennett's explanations all resolve philosophically back to evolution (which Dennett calls “The best idea anyone ever had”), but different levels of explanation are critical if we are to see the forest in spite of the trees. But more importantly is the starting point for our explanations. We can't start with skyhooks, we have to start (and for that matter, end) with cranes. Yet this reduction to mechanism, no matter how complex, integrated and emergent, finds enormous emotional resistance in many of us even to this day. For example this comment from his APA Presidential Address in 2000:

<http://ase.tufts.edu/cogstud/papers/apapresadd.htm>

“Valéry's “*Variation sur Descartes*” excellently evokes the vanishing act that has haunted philosophy ever since Darwin overturned the Cartesian tradition. If *my body* is composed of nothing but a team of a few trillion robotic cells, mindlessly interacting to produce all the large-scale patterns that tradition would attribute to the non-mechanical workings of my mind, there seems to be nothing left over to be *me*. Lurking in Darwin's shadow there is a bugbear: the incredible Disappearing Self. One of Darwin's earliest critics saw what was coming and could scarcely contain his outrage:

In the theory with which we have to deal, Absolute Ignorance is the artificer; so that we may enunciate as the fundamental principle of the whole system, that, IN ORDER TO MAKE A PERFECT AND BEAUTIFUL MACHINE, IT IS NOT REQUISITE TO KNOW HOW TO MAKE IT. This proposition will be found, on careful examination, to express, in condensed form, the essential purport of the Theory, and to express in a few words all Mr. Darwin's meaning; who, by a strange inversion of reasoning, seems to think Absolute Ignorance fully qualified to take the place of Absolute Wisdom in all the achievements of creative skill.[3]

This “strange inversion of reasoning” promises—or threatens—to dissolve the Cartesian *res cogitans* as the wellspring of creativity, and then where will we be? Nowhere, it seems. It *seems* that if creativity gets “reduced” to “mere mechanism” we will be shown not to exist at all. Or, we will exist, but we won't be thinkers, we won't manifest genuine “Wisdom in all the achievements of creative skill.” The individual as Author of works and deeds will be demoted: a person, it seems, is a barely salient nexus, a mere slub in the fabric of causation. “

Today, no educated sane person suggests, just because we cannot calculate to an exact atomic determination the actual molecular working processes of a heart muscle cell, that means there is more to the functioning of that heart muscle cell than mere bio-chemical physical mechanism. Again, at the next level of explanation, no one suggests, because we cannot exactly and completely describe the functioning of the heart at the cellular level, that we are “leaving something out” of our anatomical explanation the heart as a device (heart transplant patients still seem to be human). The same is true of the human mind and brain. But because the human brain is many orders of magnitude more complex than the heart, we need to proceed to a level beyond the anatomical. Dennett offers several levels of explanation, the highest level being the “intentional” level.

This level of explanation is entirely reductionist in principle, yet like most appropriately reductionist science, in practice it is not. For Dennett the “intentional stance”, as he calls it, is a way of describing physical, observable or behavioral outcomes that are “merely” but *not* insignificantly apparent. Just as the apparent “design” of the human eye is explained as “merely” an algorithmic adaptation for improved survival, yet not an insignificant one. That is to say, “intentionality” is not a truly “intrinsic” or “original” essence, but rather a language description we apply to certain modes of behavior by organisms (or devices) that appear (and in principle, ultimately from a Darwinian perspective) to have real purpose.

This appropriate level view fits in well with Damasio’s related theory of the organism/consciousness as levels of biological explanation. At the lowest level are basic life regulations consisting of simple patterns of response such as “metabolic regulation, reflexes and the biological machinery behind what will become pain and pleasure, drives and motivations.” Above this comes the “emotions”, more complex patterns of response, which he divides into secondary, primary and background emotions. Closer to the line of consciousness are the “feelings” which are sensory patterns signaling pain, pleasure and emotions become images. Finally, having crossed an arbitrary line of “consciousness”, we have “complex, flexible and customized plans of response are formulated in conscious images and may be executed as behavior”.

One very interesting result of Damasio’s and other neurologists clinical work is the observation that in patients where the limbic (emotional) system has been physically damaged in certain ways, the patient’s abilities to make rational decisions is completely destroyed, though the patients are “normal” in most other ways. Apparently we actually need (biochemically induced) emotions to produce (reasoned) intentional behaviors. A finding from science that philosophers ought to consider carefully.

3. Observable outcomes: That the various levels of description, “physical”, “design” and “intentional” come about merely as observable outcomes in the environmental and social interactions of organisms.

Continuing with the previous line of thought is the idea that if something in nature looks designed, it really was designed, even though most design (including ourselves) is the result of myriads of mindless selective interactions between the organism and its environment. That there is no “intrinsic” or “original” design in the apparent “perfection” of the human eye or the hummingbird’s beak is just as true as there is no “intrinsic” or “original” design in the apparent “evil” of parasitic wasps or population control through starvation.

But does the appearance of thought indicate thought or just “simulation” of thought? Does Deep Blue actually “know” how to play chess? That is to say, who actually beat Kasparov? Deep Blue or his team of programmers?

From all appearances that computer “knows” how to play chess better than any human being on the planet. Does Deep Blue use the same algorithms or “intuitions” that a human player uses? No, of course not. Does Deep Blue “know” that it “knows” how to play chess? No. It was not given the ability of create a narrative stream that could describe it’s internal deliberations. Dennett says (also from his APA address):

“Blue designs better chess games than any of them [the programming team] can design. None of them can author a winning game against Kasparov. Deep Blue can. Yes, but. Yes, but. I am sure many of you are tempted to insist at this point that when Deep Blue beats Kasparov at chess, its

brute force search methods are *entirely* unlike the exploratory processes that Kasparov uses when he conjures up his chess moves. But that is simply not so—or at least it is not so in the only way that could make a difference to the context of this debate about the universality of the Darwinian perspective on creativity.

Kasparov's brain is made of organic materials, and has an architecture importantly unlike that of Deep Blue, but it is still, so far as we know, a massively parallel search engine which has built up, over time, an outstanding array of heuristic pruning techniques that keep it from wasting time on unlikely branches. There is no doubt that the investment in R and D has a different profile in the two cases; Kasparov has methods of extracting good design principles from past games, so that he can recognize, and know enough to ignore, huge portions of the game space that Deep Blue must still patiently canvass *seriatim*. Kasparov's "insight" dramatically changes the shape of the search he engages in, but it does not constitute "an *entirely* different" means of creation. Whenever Deep Blue's exhaustive searches close off a *type* of avenue that it has some means of recognizing (a difficult, but not impossible task), it can re-use that R and D whenever it is appropriate, just as Kasparov does. Much of this analytical work has been done for Deep Blue by its designers, and given as an innate endowment, but Kasparov has likewise benefited from hundreds of thousands of person-years of chess exploration transmitted to him by players, coaches and books.

The fact is that the search space for chess is too big for even Deep Blue to explore exhaustively in real time, so like Kasparov, it prunes its search trees by taking calculated risks, and like Kasparov, it often gets these risks pre-calculated. Both presumably do massive amounts of "brute force" computation on their very different architectures. After all, what do neurons know about chess? Any work *they* do must be brute force work of one sort or another.

It may seem that I am begging the question in favor of a computational, AI approach by describing the work done by Kasparov's brain in this way, but the work has to be done somehow, and no *other* way of getting the work done has ever been articulated. It won't do to say that Kasparov uses "insight" or "intuition" since that just means that Kasparov himself has no privileged access, no insight, into how the good results come to him. So, since nobody knows how Kasparov's brain does it—least of all Kasparov—there is not yet any evidence at all to support the claim that Kasparov's means are "entirely unlike" the means exploited by Deep Blue. One should remember this when tempted to insist that "of course" Kasparov's methods are hugely different. What on earth could provoke one to go out on a limb like that? Wishful thinking? Fear? "

Dennett would argue that any organism or device that exhibits a complex behavior, actually has that complex behavior. The thought experiment advanced by Pylyshyn that if one had one's neurons slowly replaced, one at a time by silicon chips with the same functionality, would there come a point in the process when one is no longer "truly" conscious, seems to point to the need for a "substrate neutral" approach to behavior and thought.

But chess is just chess of course. What about real creativity? Those still holding out for intrinsic creativity should consider the computer composer program EMI, which Dennett describes:

"This is where David Cope's EMI comes into play. Cope set out to create a mere efficiency-enhancer, a composer's aid to help him over the blockades of composition any creator confronts, a high-tech extension of the traditional search vehicles (the piano, staff paper, the tape recorder, etc.). As EMI grew in competence, it promoted itself into a whole composer, incorporating more and more of the generate-and-test process. When EMI is fed music by Bach, it responds by generating musical compositions in the style of Bach. When given Mozart, or Schubert, or Puccini, or Scott Joplin, it readily analyzes their styles and composes new music in their styles, better pastiches than Cope himself—or almost any human composer—can compose. When fed music by two composers, it can promptly compose pieces that eerily unite their styles, and when fed, all at once (with no clearing of the palate, you might say) all these styles at once, it proceeds to write music based on the totality of its musical experience. The compositions that result can then also be fed back into it, over and over, along with whatever other music comes along in MIDI format, and

the result is EMI's own "personal" musical style, a style that candidly reveals its debts to the masters, while being an unquestionably idiosyncratic integration of all this "experience."

EMI can now compose not just two-part inventions and art songs but whole symphonies—and has composed over a thousand, when last I heard. They are good enough to fool experts (composers and professors of music) and I can personally attest to the fact that an EMI-Puccini aria brought a lump to my throat—but then, I'm on a hair trigger when it comes to Puccini, and this was a good enough imitation to fool me. David Cope can no more claim to be the composer of EMI's symphonies and motets and art songs than Murray Campbell can claim to have beaten Kasparov in chess. “

4. No Cartesian Theater: there is no "I" or "Me" inside my brain. "I" or "Me" is my brain (to a greater or lesser degree).

Dennett strongly feels that any theory of consciousness has to resist the allure of the apparent place in our brain where "I" become aware of "my" experiences. This is an obsession that haunts philosophers but also many cognitive scientists because it is apparently how our brain was designed by evolution to think about itself. Dennett summarizes his position against Joseph Levine's position "that conscious experiences themselves, not merely our verbal judgments about them, are the primary data to which a theory must answer." (again from "Get Real"):

“But since this [prior account] is to be an account of *first-person* epistemic access, the place in question must be none other than the place *where I reside*, the Cartesian Theater. There is no such place. Any theory which postulates such a place is still in the grip of Cartesian materialism. What (and where) is this I? It is not an organ, a subfaculty, a place in the brain, a medium--or *medium* (Dennett, 1993)--into which information gets transduced. My attack on the Cartesian Theater is among other things an attack on the very practice--illustrated here in an otherwise remarkably surefooted performance--of positing an unanalyzed "I" or "we" or "self" or "subject" who "has access" to *x* or *y*, as if we could take this as a primitive of our theorizing. Any sane account of the mechanisms of consciousness must begin with the denial of Cartesian materialism; and that leads irresistibly to the view that the "me" has to be constructed out of the interactions, not vice versa. This is the point of what I sometimes think is the most important, and under-appreciated, passage in *CE*:

How do I get to know all about this? How come I can tell you all about what was going on in my head? The answer to the puzzle is simple: *Because that is what I am. Because a knower and reporter of such things in such terms is what is me. My existence is explained by the fact that there are these capacities in this body.* (p410)”

This also relates to Dennett's claim that the best method to investigate the mind is "normal" 3rd person science which he calls "heterophenomenology" or the scientific investigation of how it seems to us without giving epistemic privilege to those "observations. Dennett explains the position (from his APA address also):

“You defenders of the first-person point of view are not entitled to this complaint about the "primary data" of heterophenomenology, since *by your own lights*, you should prefer its treatment of the primary data to any other. Why? Because it manifestly does justice to both possible sources of non-overlap. On the one hand, if some of your conscious experiences occur unbeknownst to you (they are experiences about which you have no beliefs, and hence can make no "verbal judgments"), then they are just as inaccessible to your first-person point of view as they are to heterophenomenology. *Ex hypothesi*, you don't even suspect you have them--if you did, you could verbally express those suspicions. So heterophenomenology's list of primary data doesn't leave out any conscious experiences you know of, or have any first-person inklings about.

On the other hand, unless you claim not just reliability or normal incorrigibility, but outright infallibility, you should admit that some--just some--of your beliefs (or verbal judgments) about your

conscious experiences might be wrong; in all such cases, however rare they are, what has to be explained by theory is not the conscious experience, but your belief in it (or your sincere verbal judgment, etc). So heterophenomenology doesn't include any spurious "primary data" either, but plays it safe in a way you should approve."

Even Levine is apparently starting to see the point as Dennett says that "Levine notes, correctly I daresay, that if I am right when I say 'there is no difference between being of the heartfelt opinion that something seems pink to you, and something *really seeming* pink to you,' then there is 'nothing left about which to argue.'"

5. Summary: Evolution provides an explanatory and even testable scientific framework, but not a "greedy" reductionist account of consciousness.

Dennett would like to see explanations of consciousness, creativity and "how it seems to us" posited at the appropriate level of explanation without excesses from either camp. He says (again from the APA address):

"There is a persistent problem of imagination management in the debates surrounding this issue: people on both sides have a tendency to underestimate the resources of Darwinism, imagining simplistic alternatives that do not exhaust the space of possibilities. Darwinians are notoriously quick to find (or invent) differences in *genetic fitness* to go with every difference they observe, for instance. Meanwhile, anti-Darwinians, noting the huge distance between a beehive and the *St. Matthew Passion* as created objects, are apt to suppose that anybody who proposes to explain both creative processes with a single set of principles must be guilty of one reductionist fantasy or another: "Bach had a gene for writing baroque counterpoint just like the bees' gene for forming wax hexagons" or "Bach was just a mindless trial-and-error mutator and selector of the musical memes that already flourished in his cultural environment." Both of these alternatives are nonsense, of course, but pointing out their flaws does nothing to support the idea that ("therefore") there must be irreducibly *non-Darwinian* principles at work in any account of Bach's creativity.

In place of this dimly imagined chasm with "Darwinian phenomena" on one side and "non-Darwinian phenomena" on the other side, we need to learn to see the space between bee and Bach as populated with all manner of mixed cases, differing from their nearest neighbors in barely perceptible ways, replacing the chasm with a traversable gradient of non-minds, protominds, hemi-demi-semi minds, magpie minds, copycat minds, aping minds, clever-pastiche minds, "path-finding" minds, "ground-breaking" minds, and eventually, genius minds. And the individual minds, of each caliber, will themselves be composed of different sorts of parts, including, surely, some special-purpose "modules" adapted to various new tricks and tasks, as well as a cascade of higher-order reflection devices, capable of generating ever more rarefied and delimited searches through pre-selected regions of the Vast space of possible designs."

I'd say that "there is a grandeur in this view of life"