Adam Conkey
WORD COUNT = 14,873

The visual processing system is divided into several pathways and, most interestingly, into different stages in which incoming signals from the retinal map are processed. This makes me believe that practitioners in machine learning were seriously on to something when the notion of “Deep Learning” was conceived of, and lends to the ongoing popularity of the sub-field. Deep Learning creates an ANN having multiple hidden layers (instead of the traditional one), each hidden layer computing a non-linear transformation of the activation levels from the preceding level. In the ANN the idea is that each layer learns a way to re-represent the data, developing more complex representations as activation spreads to deeper layers. The first level might detect edge orientations or dark/light spots, the second level might detect longer lines and contours, the third level might start to recognize shapes and outlines, the fourth level might start differentiating foreground/background (depth), the fifth level might learn complete objects, the sixth level might start classifying the objects, and so forth. There can also be feedback connections (recurrent neural networks) for error correction or for representing time-dependencies, which could allow the system to recognize motion.

The fact that deep ANNs are so successful in making accurate classifications of very complex concepts and images, to my mind, suggests that the visual pathways in the brain are structured as they are to accomplish the same sort of thing, i.e. each layer in a pathway is re-representing the incoming retinal signals and performing non-linear transformations thereon. I believe the deep ANN design got its inspiration from the structure of the brain (research this to verify), so it would make good sense that this is the case. But the ANNs are drastically easier for us to manipulate and experiment with - you just input some parameters and training data and let it loose. And now we are seeing and understanding what can be achieved with this layered approach in an artificial setting. I think these results will be able to be generalized to the biological domain. That is, when we construct certain network structures that end up performing accomplishing some task or instantiating some ability, and if we find similar structures/organization in the brain, we will be able to conclude that the brain is organized in such a way to accomplish the same task or to instantiate the same ability. We might be able to sidestep a lot of experimentation on the physical brain by instead experimenting aggressively with ANNs and then verifying those results in the brain. Until more advanced and less-invasive imaging techniques for the brain are developed, I think this is a promising course of research to probe the capabilities of the human brain cheaply and efficiently.

==============

From my understanding, computer architecture has not changed very drastically since its inception. The Von Neumann architecture was one of the first serious designs for a digital computer and that design has more or less prevailed to this day. Digital computers are of course orders of magnitude more powerful and smaller and cheaper today, and adopt newer technologies like multi-core processors and parallelization, but the principle is the same: addressed memory, a comparatively tiny number of registers, and a hierarchy of storage, where instructions to be executed are located in storage (wherever in the hierarchy it may be) and swapped onto the CPU for processing. Multi-tasking is achieved by potentially having multiple cores or multiple processors, and maybe even distributing the load onto several distinct machines, but primarily by having a process schedule and rapidly swapping processes in and out on the limited processing resources available.

I am taking special notice of what I see as the future of computing. These are systems that deviate from the traditional architecture described above and are ones that instead look quite a lot like artificial human brains (which is where they get their inspiration from). I am thinking of neuromorphic computers, cognitive computing, and various other system paradigms that rely on phenomena such swarm behavior/intelligence and self-organization. These systems are certainly gaining traction (IBM is pioneering mainstream cognitive computers and has released brain-like chips, DARPA heavily funds research in all of these areas) and it is clear to me that once the technology matures, it will revolutionize how we engineer robots and artificial intelligence generally. I think now it is not as clear to people how they will be as useful as the traditional architecture, because the new systems will be less under our programmatic control. Instead the new systems will have the basic architecture and programming to learn from experience, and we will have to teach them to do everything we wish them to do.

To achieve this we will need to focus more on the hardware than the software, i.e. in a way there will be no high-level imperative code to be executed, there will only be hardware that is programmed to be adaptive. In a way, this what humans get at the start. We are born with the infrastructure to be human and the rest is left to experience. We also come equipped with a lot of behavioral responses, knowledge embedded in our being that results from millennia of evolutionary experience, and that undoubtedly can get us pretty far in the world. But a lot of our motor dexterity, our language abilities, the ability to reason and think critically, the ability to use and manage our emotions, all of these things have to be learned through living. It seems absurd that we should expect to build intelligent machines by telling them everything outright. This will not work because they are going to be expected to function in an ever-changing world. From a survivability standpoint, a dynamic environment requires dynamic inhabitants. So a much more sensible route is to build learning machines, and I think the most sensible way to do that is make them learning machines at the core, not just simulated ones, i.e. provide them with architecture suitable for learning, not just a digital computer that can simulate the abilities of such an architecture.

I think this can be seen as general purpose architecture, in that it can learn whatever you decide to teach it, but once teaching commences the architecture will become specialized and the more teaching it receives, the more specialized it will become. And by teaching I don’t necessarily mean a human has to sit there and tell it things or, if it has a body like a robot, to move its joints manually or show it what it needs to do. A lot of human learning is self-initiated - we arguably figure out how to walk on our own, we pick up a lot of language just by overhearing it, and we develop motor skills and navigational abilities just by moving about in our environment generating action and receiving perceptual feedback. The learning machine can do the same sort of thing. This, I believe, is the proper path to the singularity. For here you will have machines whose learning abilities are unbounded, restricted only by the architecture we provide them with. Theoretically, if we built a sufficiently complex and robust architecture, it could very quickly surpass human knowledge and understanding and there you have initiated the singularity.

I speak pretty loosely about these things but this is just where I see things headed. The pursuit of AI will at some point become primarily a bottom-up endeavor, a project centered around building brain-like architecture suitable for adaptation and learning, and will move away from trying to hard-code many of the high-level attributes that human beings instantiate. There is also the matter of artificial consciousness. I am of the opinion that hard-coding AI will never, ever, lead to conscious states. But an artificial brain built from the bottom up, that just might. How that will happen, I do not know, but if we're placing bets I am all-in on neuromorphic engineering, cognitive computing, synergetic computing, self-organization, etc..

================

Even if a digital computer is augmented to store more and more information and process it at faster and faster speeds, it doesn't change the fact that the processor of the computer is only able to handle relatively very small amounts of data at a time. In a processor with, say, only 8 registers of 4 bytes each, there are only up to 32 bytes of information being processed in a given cycle, and really less than that since some of those registers and operations are given over to administrative tasks like storing an instruction pointer and performing stack maintenance. Its computational power comes from swapping out values very quickly, but this is all so inefficient because the physical components responsible for the actual computations, the changing of stored data into new data, is so small in comparison to the physical architecture responsible merely for storage and maintenance. It would seem the brain, on the other hand, employs an overwhelming proportion of its real estate to actual processing, in a way not even bothering with storage per se but letting the storage come out of the processing.

What if we view the brain as only a processor? Not in the way a digital computer is, which is separated between memory and processor, the processor being arbitrarily applicable to whatever data it may pull from memory. Instead, the brain as a processor would be a physical realization of the data it is taken to represent. The architecture is by no means arbitrary, but carefully constructed to embody the complex subtleties of the environment in which the brain-controlled organism resides.

We should not even limit this to the brain itself but the nervous system as a whole, unique to an active body. A body gains its action through the use of muscle movements and coordination thereof, and those muscles are controlled ultimately via nerves that lie on a continuous path in the brain. Here then we don't think about the brain sending instructions to the muscles to move in this way and that, but instead: neurons fire in the brain, and those firings may or may not cause a certain muscle group to contract or relax. It is not an instruction being sent, but an outcome of a particular neural path firing that included the nerves responsible for causing movement in the muscle group of concern.

In this sense there is no information being sent to the muscles, though I suppose it depends on how you conceive of information. It is nothing that needs to be interpreted by any endpoint in or around the the muscles, it is merely an activation signal in that path, a reaction to the signal existing on the subpath containing the nerves responsible for that muscle movement. In a way, then, there is no processing per se, only activations of subpaths in a continuous network.

This seems counter to the way we approach robotics today. Robots have a computer as a brain and they are tasked to process information in order to send instructions to its mechanical parts, such as move this servo this many rotations and extend this pneumatic arm this many millimeters and so forth. At some point this will amount to sending a voltage to the appropriate part, so there is some question as to whether this can actually be seen as analogous to what happens in the human body and vice versa, only things are happening much more simply in the robot. In this case it would just be in how we are interpreting matters. It may be that the robot is a system operating by simply the system interacting as it does, only it is a much more simple system than the human system. Because at the system level, the robot is just physical components interacting as they do, initialized so as to produce a desired behavior. We are the ones who interpret things as, here are instructions, the processor processes those instructions and produces some output to ultimately guide the system in its interactions with the environment. Perhaps from a very fundamental point of view, even the digital robot and the human being are no different - they are each a physical system “designed” to exist in some environment. They differ very much in complexity though, and in a non-trivial way. It is not in the same sense that we would say a motorcycle is more complex than a bicycle; it is true in most cases but they operate similarly and on the same principles. It seems that the disparity between human and digital robot is of a different sort.

Part of my hesitation for outright saying that the brain is completely different from a digital computer, that it is not even close to being the same sort of system, comes from the fear that we think too simplistically of implemented digital computers. It is true that in operation an actual digital computer in behavior can be represented by your choice of abstract computation, e.g. Turing machines or register machines or what have you. Those abstract notions capture what we intend a physical computer to do, and it is obviously no accident that these two devices (physical and abstract) are so impressively coordinated. But the fact is that a computer is not really a computer until it is implemented. Even once implemented we tend to take the system as an abstract one and we talk and reason about it at an abstract level. We program in high-level languages and are provided the illusion that when we talk to the machine, so to speak, it blindly follows those instructions in all their semantic glory. But really to talk to a machine we need a process of transduction, just as the human body requires transduction to do any meaningful work. The human eye receives raw light signals from the environment and must transduce what it receives into a more suitable form for the brain to handle, and this happens via chemical activation in the cells of the retina. That chemical activation ultimately leads to electric activity in the optic nerve which reaches the brain in that form, and really all that electric activity is based in chemicals, ions and channels and such. The light patterns in and of themselves are useless to the human, but those light patterns transduced to structured chemical activity, that is useful. A similar thing must happen with computers, only we must alter how we think of the computer-environment relation. If you have a computer with no camera, no microphone, no appendages for reaching out into the environment and to engage with it, then either you must think the computer is very poorly constructed for the environment it exists in or, that is not its environment. Perhaps it is the case that the computer's environment is not the physical environment as we know it, the physical space that the actual computer is sitting in, but rather the abstract space consisting of the data it is to operate on.

This is kind of an interesting parallel, because the human-environment relation would be conceived of as a physical body coupled to a physical environment, each of the same fundamental character and if we look at some other human, the first human becomes a piece of the environment as well. (That a human from one perspective can, viewed from another be considered just another part of the environment seems to suggest that they are the same sort of thing.) In the case of a computer, we would have to conceive of an abstract operating system being some constellation of data, which itself then interacts with an environment consisting of abstract data.The data is “meaningful” to the computer because when it “interprets” the data, that data is well-defined, it causes action in the computer, action being some further processing step. You could give it data in a different language, non-binary, and it would be “meaningless” to the computer. Similarly you could give it something that doesn't define an address where it might find meaningful data, or a corrupt instruction that just doesn't specify an action for that particular computer, though the same data in a different computer might be very well-defined and produce an expected action. The digital computer is still dependent on an environment, there is only a limited range of contexts in which any given computer can operate. Keep in mind I am thinking the computer abstractly, loosely as being identical with the operating system of a computer. That seems to be a cogent distinction - operating system consisting more or less of a well-defined constellation of data, and a distinct constellation of data on which the OS may operate. (Of course they could overlap, but that doesn't seem to make a difference.)

Suppose you took every instruction of an OS and mapped it onto a graph and similarly mapped the data on which the OS operates onto a graph, potentially the same graph. This would be a complete representation of the system somehow containing all possible interactions and operations that can take place between OS and pre-defined data. And perhaps the data need not even be pre-defined, you could have new data spawn and emerge within some original context and the graph would be augmented accordingly. Suppose then that you ran the computer (physical computer implementing the OS and data) at the same time you had displayed the corresponding graph representing the complete OS-data coupling. You would see very rapid activation of various paths and nodes in the graph. Suppose you regulate it a bit to abstract away from the hard implementation of the physical computer and allowed multiple processes appearing to operate simultaneously to actually operate simultaneously (I am referring to how a single core processor will give the appearance of parallel execution by very rapidly cycling among processes, rapidly loading from and saving to memory various processes.

This is a limitation of the way we implement single core processors). You would then see activation in various parts of the graph, some of those paths potentially overlapping, some causing cascades of more and more activation (forking new processes, operating on the same data), some causing only very local activation and dieing out. There is a chance, it would look something like a brain.

It would likely look even more like a brain if you incorporate activity to account for the physical implementation of the computer, namely activity to handle input from users (keyboard, mouse, etc.) and to handle a display screen, power management, general system maintenance so that the computer can stay online and operating. Perhaps there may be error-correcting circuits and redundancy in the data representations so that the system is very robust to corruption, some pathways being designed to maintain the integrity of the data and the OS so that neither one can veer too far into self-destruction. For all of this to exist in a self-regulating way and all being integrated on the same “architecture”, you would require a rather complex system to carry it all out. Whether I mean complex in the same way I mean that the human brain is complex, that remains to be seen and this perhaps is the question we should all like to answer in order to settle this business once and for all.

The problem is that you would never know of all this complexity, and really it is never actually realized because we filter it all through a single meager processor in our actual implementation of the digital computer. We forego an impressive and wildly difficult to achieve architecture in favor of the one that serves our immediate purpose and is easy to construct. It is simple, small, conceptually easy to grasp. But it defeats the complexity a computer is capable of. All of the complexity that might exist within the complete representation of an OS-data coupling is left unexplored because instead of implementing it in all its glory, we limit our implementation to exploring only infinitesimal regions of it at any given time. Even when we allow for massive parallel processing, it doesn’t change the fact that in a given instant, the processor can only handle a tiny fraction of the data available to it. And it is not even handling it continuously so long as it is cycling processes.

In this way we have build a machine that, in a sense, emulates the limitations of human consciousness and attention. Humans can only consciously attend to a very small fraction of their environment at a time, even when we conceive of the environment as being a very local thing to the human being, say just the small room they are sitting in. If you consider the wealth of complexity that room may offer, the conscious human may only grasp a very small portion of it at one time; and in order to try to grasp more of it at one time it must cycle its attention among the various tasks of attending to the various things requiring attention. We made a machine that acts in a similar way to the high level behavior of conscious activity. BUT, remember conscious activity is emergent from a wildly complicated architecture, and on that lower level you do not have the bottleneck you do at the higher level. On the lower level of the architecture itself you have an intense orchestration of activation, many many things going on simultaneously and communicating and coordinating all at once. This is because even a single isolated perception or a simple stimulus from the environment of the human organism requires an explosion of activity from the lower level brain architecture in order to handle it. A simple processor is not going to cut it. And yet, this is what we have given our machines that are optimistically supposed to emulate our conscious behavior in all of its most impressive forms. See the flaw?

There is much confusion between computations, systems that implement them, conscious behavior, systems that implement it, and the interrelations between all of these things and more. It is worth sorting out all of this confusion as it seems to be the serious roadblock to constructing complex systems that actually are like the human brain and that actually do emulate our conscious behavior. That is, we are currently prevented from constructing artificial brains that emulate the low-level architecture and behavior of the human brain, and in turn prevent ourselves from constructing artificial systems that emulate the emergent phenomena such architectures are capable of, all because we cannot sort out what each of these notions and systems require, and how exactly they differ.

======================

Self-organization in the brain is, I think, the most essential design feature contributing to the success of the brain as a control structure for human beings. If you consider a modern digital computer, if a hardware failure occurs in the processor or somewhere on the motherboard, you will typically have complete system failure and will potentially have to replace the entire board or simply buy a new machine. Even with peripheral storage, the only way for the system to behave as usual if the hard disk becomes corrupted is to have identical backups stored on redundant drives. Obviously it would be biologically expensive for the brain to employ this same design structure, and it instead employs a very different memory and processing model that is able to survive even a great trauma. From an engineering perspective this comes at the cost of having to utilize an extraordinarily complex design, but from a biological perspective this makes great sense since it must survive in and adapt to a dynamic and unpredictable environment.

It is a bit baffling to me that the enterprise of artificial intelligence did not begin with trying to replicate the design and operation of the brain. I suppose it was a matter of what kind of technology and knowledge was available at the outset; though the basic structure and operation of neurons was known, global brain structure as well as the subtleties of neuronal interactions were not known until much later, and we still struggle with these aspects today. I think the biggest leap forward we’ve made in AI is with machine learning and artificial neural networks. With ANNs we are at least beginning to understand the principles of truly distributed memory and distributed representations that do not rely on memory addresses in order to recall or recreate information. I think this can be taken a step further by implementing ANNs in hardware instead of just running them as simulations on a general purpose computer. There should be some way that we can take the design principles of the human brain and implement an artificial brain that combines processing and memory to be performed by the same architecture (as opposed to there being a separate processor with registers that will pull in data from a memory source). But in order to achieve this properly, self-organization must be a central design principle, allowing the system to care for itself and manage its organization without central control.

==============

There was a statement in the Doidge/Merzenich paper we read that stuck with me, where Merzenich was describing his experiments with severing nerves to the fingers in monkeys and saw brain maps changing nearly instantaneously after a procedure was done in some cases. He remarked something like the maps seemed as if they were always there, but just had to be uncovered or given the space to be used.

I would be willing to take that statement literally. When brain networks are discussed, it typically focuses on the neurons that fire as opposed to all the many neurons that are not firing during a particular activity or behavior. But inhibition in the form of inhibitory networks plays such a key role in the brain, and whatever neurons happen to be firing during a given task, they are activating at least as many inhibitory neurons to prevent other regions from firing simultaneously. When you have one map corresponding to the motor movements of, say, the middle finger of your left hand, the neurons firing in for that map are inhibiting other neighboring neurons to prevent other maps from encroaching on that space. But if you lop off your middle finger, the inhibitory networks that were associated with the map corresponding to the middle finger are no longer exercise, lose strength, and other firing can take over in that region (i.e. neighboring maps will no longer be inhibited and will start to spread into the open space).

That being said though, I don’t take the maps to be as definitively defined as much of modern neuroscience has led me to believe. You will often see articles reporting that scientists have found the neural correlates of BLANK, where “BLANK” takes on every sensation or emotion you could imagine. This is all based on a subtractive methodology: you scan a patient’s brain at rest to get a baseline reading, you scan the patient’s brain while they are performing some task or subjected to some stimulus to provoke some sensation, subtract the baseline activity from the activity seen in the experimental task, and voila! you have found the neural correlate of such and such. The problem is that it isn’t that simple. Those identified neurons are not unique to that task, and the same neurons may fire for a different task at a different time when the patient is it a different “baseline”. Also, the same neurons may fire differently for the same task at a different time.

What is lacking is a more holistic understanding of how those maps fit in with the rest of the firing. Just as peripheral vision provides the context for our central vision, the entire brain firing pattern at a given time sets the context for a particular region of neurons firing at that time. This is the case because the brain does not consist of isolated modules firing away, it consists of collections of neurons firing in concert, coordinated by oscillations persistently present in the brain and often taken to be noise. In my opinion, there is no such thing as noise in a healthy brain - every neuron firing is for some purpose and it may not be pure processing. It may be just to maintain the state of connections by periodically strengthening them appropriately. It may be oscillations traveling at varying frequencies to introduce little pockets of computation that can happen over time and share the architecture. In any event, I don’t trust brain maps too much, because I think they are taken out of context. It is based on the traditional reductionist mentality that has been so successful in science thus far. But the brain is a complex system, and analyzing complex systems is different since they are non-reductive systems - you can not break it into simpler parts because its function is a result of the entire system exercising its dynamics, not of any small piece of it.

================

I am wondering about the maximum capacities of the brain in terms of processing ability and storage. I know I have come across some rough estimates for these notions in terms of number of computations per second and bytes of storage based on the basic structure of the brain, but I am not fully convinced by this approach. To my knowledge, we don’t fully understand how memories are encoded in the brain and how the brain is handling its distributed storage. Any calculation of memory capacity that is done strictly by counting nodes and connections would be in error, since such a calculation does not consider the temporal qualities of brain storage and processing (e.g. oscillations and their role in facilitating memory integration over disparate regions of the brain). Everything that occurs in the brain depends both on structure AND dynamics, and I think both need to be considered in depth to get an accurate estimate for brain capacities.

What is interesting I think is that the brain seems to have a governor on it that sort of throttles back its activities. This could clearly have some evolutionary advantage since the brain running at full capacity all day would burn an exorbitant amount of energy and we would have to consume much more food just to stay alive. It may also fatigue your system much more and would require longer rest periods (sleep) and could shorten the lifespan of the brain (and therefore the animal). But I think it is clear that the brain is, at any given time, capable of exercising any of its abilities to a greater extent than it typically does. This is suggested by studies that have been done with TMS (transcranial magnetic stimulation). I am recalling from an episode of “Through the Wormhole” where a researcher involved in developing technology for the military created an experimental task where subjects were to identify dangerous targets in an aerial satellite image (e.g. missile launchers or enemy communications towers), a task usually very error-prone and difficult for humans. They were trained so they knew what to look for, and then tried it out on a number of images. Then they did the same task but with a TMS unit attached to a strategic region of the brain. With the stimulation, the task was easy and their accuracy rates skyrocketed. They were able to focus more intently and things were just clearer and more discernible for them. Just by activating specific regions of the brain to a greater extent than they normally are (heightening activation), they were able to increase their abilities by orders of magnitude.

This may also be suggested by reports of people that lose a particular sense begin to experience heightened abilities in the remaining sensory modalities (e.g. a person going blind and beginning to hear much better than before). I think in this case it may be similar to the Merzenich monkey experiments in that the other sensory modalities are taking over the real estate that was previously used by the lost sense, so that each sense has a larger neural space to operate with, perhaps providing for deeper and more precise representations, and allowing attention to access those modalities better. It would seem that resources are limited in the brain and all regions are competing for space and processing capabilities, and when one is taken out of the picture, the remaining tasks/abilities going on in the brain take over the vacated region.

It is natural then to wonder about brain augmentation, and what the brain could be capable of it were only given more and/or better resources to do what it needs to do. Perhaps we could have a larger brain with more real-estate or a brain with even denser connections, or the connections could be made more efficient in transferring energy. We might figure out how to interface the brain with remote resources, e.g. wireless communication nanobots that could pull in more information from the environment or provide the brain with cloud storage capabilities to reduce the burden of storage the brain faces. With these sorts of nanobots feeding in information from sources that are not directly mediated by our current sensory interfaces, we may develop new senses that we can't as of yet imagine. When you think about it, everything we have learned and all of our memories and thoughts have their basis in external information that was only internalized through mediation of the senses. If we bypass our senses and begin to receive information through another means (nanobots), who is to say we wouldn't begin to develop new sensations and feelings of intuition (you know something to be true unequivocally, yet you cannot explain to anyone why it is so). It's tantamount to psychics and those claiming they have mystical knowledge, yet in the nanobot case there is at least some justification for it.

====================

I have had some issues with the Turing Test and what it is taken to resolve, and I think emotion is a good focal point for teasing out some of my concerns with it. The very notion of "intelligence" is difficult to get a solid hold on and I think people typically speak of it with two different intentions in mind:

1. Something behaves intelligently, though it is not taken to be conscious and having experiences in the way humans are (e.g. a chess computer or a self-driving car)
2. Something is intelligent in the way humans are, i.e. it has phenomenal experience and some sense of being and is able to conceptualize and reason about the world it inhabits.

It is too quickly forgotten that such a distinction is made and many confuse the two. I think this accounts for the over-optimism of many AI researchers/enthusiasts (believing they are dealing with (2) but have mistaken it for (1)) and for the over-pessimism of pretty much everyone else (believing that machines that achieve (1) are purported to be (2) as well).

The Turing Test addresses only (1), which is my primary issue with the test. It is based only on what is observable/extractable through conversation and inherently cannot with any reliability probe the matter of whether or not the machine in question is having a phenomenal experience. This point is addressed in Turing's 1950 paper in which he introduces the game, but he does it only cursorily. He quotes Geoffrey Jefferson's 1949 paper *The Mind of Mechanical Man* where Jefferson says essentially not until a machine is feeling emotions can we consider the machine to be intelligent in the way we conceive of humans as. The point can be made more broadly to say that a machine that says intelligent things and maybe even claims to be having experiences and feeling emotions is not necessarily doing so and that at best we have a machine that satisfies (1). I would want to say that the appearance of intelligence (sense (1)) is not sufficient for genuine intelligence (sense (2)), which I take to be a direct negation of the Turing Test.

The natural question then is, what is our metric for intelligence? If we can't base our determination on what we observe, how can we know anything but ourselves is intelligent, including another human being? I think the matter gets even worse before we can ask this question meaningfully. Jefferson sets a metric of sorts for a machine to be intelligent (sense (2)), and I believe this is the metric most people utilize. We tend to believe that if something is alive and aware, it should be experiencing something like we are. I think this is horribly misguided and far from warranted, considering we have animate creatures literally walking around about us that are intelligent, are having phenomenal experiences, though perhaps to a lesser extent than humans are. So in truth there should be at least one other sense than the two above, namely

3. Something is intelligent and having phenomenal experience with a sense of being, but is not having the same experience as a human being.

I would think that all non-human animals and insects fall under this sort of intelligence. Still, in sense (2), there is something strange that emotions will bring out. Sense (2) assumes some kind of archetypal human that sets the metric for human experience, that all humans feel the same emotions. To a large extent it can be said that humans do feel the same emotions, perhaps to different degrees and in unique ways, but for the most part our experience is fairly uniform throughout the population. There are certainly exceptions though. Consider psychopaths. They are certainly human, but must be having such a different experience than your typical human being, as is evidenced by the fact that your typical human being simply cannot understand why a psychopath behaves as they do. Psychopaths may feel some emotions, but for the most part they get by just by mimicking the emotional responses of others. This is how they fit in, but they are not actually FEELING these emotions, they are simply exhibiting the tell-tale signs of them so that observers will believe they are normal and feeling the appropriate emotions at the appropriate times. The world of a psychopath is mentally inaccessible to a healthy human, and vice versa; each cannot understand the experience of the other because they see and understand the world in drastically different ways.

Now consider that we have this kind of variation within our own species. How can we set the metric of intelligence and conscious experience as human experience when we cannot even consistently say of each human what it means for them to be intelligent and conscious, especially just through observation? My thought is that the test/metric for intelligence/consciousness/phenomenal-experience/intentionality/cognition/sentience/awareness/so on and so forth is not a simple matter at all. In fact, it is something we have little to say about until we understand precisely what it is about the human brain that allows humans to be as they are. I believe there are necessarily principles and features that underlie the human nervous system that could allow us to formulate a precise notion of what it means to be any of the things just listed. It is still easy to say with confidence that another human is any of those things, because not only do we observe them to be as such, but we know that each of us is having an experience and that fundamentally every person is no different than any other. We are all running the same machinery here, so to speak. We all have brains and nervous systems and we know that they are not so different, and for what differences there are, we can observe some difference or another in higher levels of observation from the low level hardware (behavior, or appearance).

I can say of any human that it is highly likely they are having some kind of phenomenal experience, as inaccessible as it is to me, because I observe them acting as if they were and I know that their body is effectively the same as my body, the engine of my own experience. But if I observe, say, an android that looks and behaves like a human, but I further know it is being run by a digital computer, I cannot confidently say it is having a phenomenal experience in the way that I am, because I know its machinery is different and I have no prima facie reason to believe that sort of machinery can give rise to phenomenal experience. This makes the question of intelligence/experience a more involved question, especially when entering the arena of ARTIFICIAL intelligence/experience. We cannot rely on simple observations, we must know something about the operation and causal structure of the machine/being we are observing.

This gives us a concrete problem to work on. We need to understand the human brain, and specifically what about it gives rise to conscious experience. If we understood that much, we could build an artificial brain that mimics the human brain's operation. If a machine being powered by such an artificial brain behaved like a human, I would feel confident in saying it is intelligent and genuinely having an experience. If we could uncover some necessary and sufficient conditions for intelligence, as infeasible as it sounds, we could then say of anything whether or not it is intelligent and having an experience. It starts with the brain though. Our best shot at genuine artificial intelligence is re-creating the brain in all its glorious complexity.

====================

Some of aspects of how people categorize things are relevant to machine learning and it's interesting to compare how these aspects aspects affect human categorization and how they affect machine learning algorithms:

Correlation Effect - people's classifications in the disease classification task were skewed by features that were highly correlated. In machine learning, it is common to completely delete a feature if two features are highly correlated, the reasoning being that additional features should be used in computing a classification only if it will lead to more accurate results. If you have two features that are perfectly correlated, the classification will be identical using both features as using just one, and if there is a high correlation, the computed classification will not be much different than using just one of the variables. In general it is desirable to get by with as few features as possible since adding more increases computation time, and training sets are often very large.

Attribute Weighting - I was interested by this idea as it seems like it could greatly help a supervised learning algorithm by telling it what features to put the most importance on. In an artificial neural network, bias weights could be attached to specific input features to make their effect on the network greater than the features. There may be a good amount of classification tasks where the experimenter knows which features should have the most influence on the generated classification, and incorporating that knowledge explicitly into the algorithm could improve accuracy and reduce training time.

=============================

Consider the role of autoencoders in the brain. For an ANN, an autoencoder has a target output layer that is the same as the input layer. The idea is to get the network to essentially come as close as possible to computing the identity function, and what is gained is a re-representation of the data in the hidden layer. If a fewer number of hidden nodes are used than the input representation, then you can get a compressed representation in the hidden layer. This may explain in part the use of feedback and recurrent networks in the brain. A neuronal network in the brain could tune itself to a certain input (perhaps a visual representation from the retina) and thus learn a compressed and distributed representation of the input to be efficiently stored and integrated into memory.

=========================

Searle, at the beginning of his metaphor paper, identifies cases where a speaker means something different from what they are explicitly saying. Metaphor is a subset of this class, and he also identifies irony and indirect speech acts. I am not fully sure I know what he intends by indirect speech acts, but I take it to be saying something that leaves out a lot of essential information, but the listener understands it all the same as they are able to fill in the gaps or infer the rest of what you were saying. I use this often in my own conversations. I enjoy saying the least amount I need to in order to get someone to understand me; I see a certain elegance in it, and it's satisfying when you can come up with a statement that requires a real mental leap of your listener, and they get it anyways.

I would also identify sarcasm. Sarcasm must be very difficult for a computer to model, as the uttered sentence can make sense if interpreted literally, and the only thing that forces us to interpret it otherwise is that it makes little sense to be taken literally when embedded in the context it is uttered. You have to know a lot about the situation and environment in which a sarcastic sentence is said in order to know that is being sarcastic, and it may require a complex history between the speakers or detailed knowledge of the subject of the sentence.

Also important to point out is how we sometimes communicate our feelings. Often when you ask a significant other or family member how they are doing, they might respond “fine” or if you ask them if something is wrong, they will insist everything is great. But often times one is compelled to ask these questions because they know with certainty that something is a little off, and they just want to find out what it is. Especially when you spend a lot of time with someone, you learn their moods and know when they are mentally pre-occupied with something or, worse, when they are mad at you or frustrated with you and they won't tell you why. But it's interesting that if you ask them how they are and they say "fine", if you were taking it literally you would think "oh that's great" and be on with your day, but knowing them and knowing that people are rarely just "fine", you infer the complexity of how they are feeling and know that they are likely troubled by something that is maybe difficult for them to speak to you about, which may either point to the difficulty of their situation or maybe some tension between the two of you. We seem to always be subtle when it comes to expressing our feelings and emotions, rarely coming straight and showing it as it is.

==================================

It was mentioned in class that some scientific results can be important even if they are not useful (the example was knowing that humans can see even 1/10th of a second into the future; obviously that is not enough time to act on that knowledge of the future so it really won't do anyone any good. But, it would require a whole new theory of time and require us to reconceptualize our entire world, so still important). This is a notion that strikes a deep chord with me, as I believe knowledge in itself about our world and existence is essential to our existence having any sort of purpose and meaning. It has been difficult for me to articulate this and I still cannot fully justify it, but I can roughly say that everything that we as humans enjoy about life or find meaning/purpose in has as its basis that things exist as they do, and that humans exist as they do. Our world has come about in a very particular way and operates under certain principles, and without these contingencies humans would never be. Further, humans owe their existence to the complexities of the human brain and the fact that a tiny lump of matter can somehow produce consciousness and intelligence to the degree that we have it.

To me, it seems unjustified to enjoy all the pleasures of life or to be taking issue with any undesirable facet of our existence if we do not have a detailed understanding of how our world exists, what features of that world are responsible for our conscious lives, and the principles that allow such complex behavior to emerge and persist. I liken it to cell phones. Everyone has a cell phone these days, and hardly anyone knows how they work or acknowledges the collective effort that goes into getting that little device into their hands and doing what it does. They just use it. And then you will hear complaints that they are getting no reception or the battery doesn't last long enough or one of the text messages that was sent never arrived. Ungrateful! The modern cell phone is a masterpiece of engineering and the infrastructure required to keep it working in such diverse locations is astonishing. How can anyone complain if at some point in time they made at least one phone call successfully - they heard the voice of another person that was hundreds of miles away and they did it with no wires and no tangible mode of connection - that's amazing! What more do you want? And who has the right to complain if one did nothing to contribute to the endeavor of making it happen, and has no understanding of the difficulty in getting it to work?

I see it as a similar situation with the brain. Obviously we had nothing to do with making it happen, but it somehow seems wrong to simply use it without a second thought. We should seek to understand the principles by which it operates and what is ultimately responsible for our level of consciousness and intelligence. Since basic consciousness/cognition is a precondition for everything that we could possible enjoy or find appealing in life, should we not at least try to understand that first step? Something tells me that if I knew exactly how consciousness worked, if I had some set of principles that I could understand that could prescribe how consciousness is generated and how I, specifically, am here, that would go a long way in providing a meaningful basis for everything else that I do. I think the cell phone analogy comes close to making that idea clearer, but I still feel that I cannot communicate properly the compulsion to understand as a means of justifying my existence. In any event, there is a large subset of the human population involved in this endeavor of understanding not just consciousness and the brain but physical existence as a whole, and I think that is terrific.

=========================

In an article called “Distributed Information Processing in Biological and Computational Systems” by the biological systems group in the ML department at CMU, they discussed how biological systems can give us really good guidance on how to construct our computational and distributed systems, and also that as we come to understand how to build better computational/distributed systems, we will come to understand the mechanisms in biology better as well. This is possible because the constraints in each domain are similar (e.g. constrained energy, resources, unpredictable environments, susceptibility to failure and attack, etc.).

This is a beautiful relationship to have in an area of research because you are really killing two birds with one stone - we learn more about these complex biological systems that we still struggle to make perfect sense of, and we get more complex and resilient computational systems that are more suited for the trials of our natural world. I think this also points to where we are headed with our technology: our computational systems are going to become more and more like biological systems, striving more for robustness and adaptability than rigid efficiency. This is a sensible goal because the technological network we have created (consisting of networked machines via the internet and networked devices via our mobile communication infrastructure) is becoming increasingly complex, to the point that the complexity is on the order of the complexity we have already in our natural environment. Not only this, but as noted in the article, we have introduced new constraints and challenges by relying more and more on mobile technology (cell phones, laptops, wireless LANs, sensor networks, etc.) instead of the traditional computers and servers sitting in warehouses and offices and connected by fixed wired connections. Now our networks more than ever have to deal with volatility coming from weak connections and network hub outages, signal interference, lost connections, changing network topology, and so forth.

This research focus will go along nicely with a similar project in engineering: creating materials that act more like biological systems. I’m thinking of some things I’ve seen in the works like self-healing bridges that can dynamically repair cracks in the structure instead of requiring continual inspection and maintenance, or nanotechnology that can introduce monitoring systems into effectively any physical product, e.g. a paint that has nanosensors you can roll onto any surface and have it monitor things like internal and external temperature and surface tension. The idea here is to take inspiration from systems that have some mechanism for self-protection and self-healing so that less resources and manpower are spent maintaining these systems. Now imagine if we built our network infrastructure using these new materials, i.e. self-healing and self-monitoring wires and network hubs. This would greatly reduce the need for human involvement in maintaining the physical infrastructure, and would interface nicely with the computational systems it supports.

The nanotechnology seems particularly appealing in this area. If we do in fact have something like a bio-friendly paint that has nanocomputers embedded in it, then all of our man-made structures could be coated in the material and could support and augment our communication networks. And perhaps it could go beyond just a surface coating, and we could mix nano-computers into all of our building materials (bricks and cinder blocks we use for buildings, steel we use in skyscrapers and bridges, even the asphalt we lay down for our roads). Every physical object engineered and constructed by humans could then be made to support computational tasks, and we could offload the burden of our existing infrastructure specifically designed for supporting this kind of activity onto the infrastructure that exists independently of a communication network. That is, we have/need buildings and roads and bridges anyways, and we need pervasive communication networks - why not make the existing things (buildings roads and bridges) BE the communication network. You need nothing extra then. And if the technology is in fact self-healing and self-supporting to some extent (e.g. receives energy from the sun for power) then all we do is take a small extra step by mixing in these nano-computers into our materials and we get a complex, robust, and adaptable communication infrastructure for free.

Also in the article they note how biological systems will often sacrifice speed for robustness and adaptability. This seems counter to the way we go about developing computers today. A lot of tasks requiring some sense of intelligence are being solved today by throwing a tremendous amount of computational power at them, and this approach is feasible only because our computers have matured to have lightning fast processing speeds and highly optimized procedures. The issue is that suppose you build a robot that exhibits a high degree of intelligence and is able to function in the natural world to a high degree; it is able find resources to keep itself powered and can navigate effectively and perform any number of tasks that keeps it functioning and somehow contributing to the environment or society it is a part of. But it happens that it trips and falls and cracks one small tiny portion of its processing chip. System failure. This is a catastrophic event for the system that it just won’t recover from unless it come equipped with redundant processors. These systems are just not robust and cannot recover from such an event. On the other hand, you can remove an entire half of the brain and the organism will go on living and functioning to a high degree because the brain can reorganize itself to keep its essential systems and functions operating.

Optimizing our systems for speed makes some sense now because a lot of our technology applications are about just getting some task done as soon as possible. But if we are in the business of building intelligent machines, and further, sentient machines, well it seems a little unethical to be equipping such machines with very fickle hardware. Human machinery can go through a lot before cognition and sentience is lost, and that is certainly good for us. But even if it weren't, we would have no one to blame it would just be the nature of things. But if we are engineering sentience, it seems like there is some obligation for us to make sure it is not going to be overly transient if what is produced is a sufficiently advanced intelligence. That is, if we produce something that is conscious and aware in the way we are and is capable of recognizing its existing and of fearing non-existence, wouldn't it be cruel to have given it an underlying sentience-generating machinery that is highly susceptible to failure? If it is in our control, shouldn't we seek to only produce advanced sentience that has a fighting chance in the natural world?

This brings up some issues regarding the morality of engineering sentience. For instance, it seems wrong to engineer a sentient system with a kill switch. We shouldn't be allowed to just turn a sentient being off with the push of a button or flick of a switch. You certainly cannot do this with other humans, and if you did something analogous to a human, the human would not come back. They would likely just be dead or brain dead at least (and who knows, depending on what technology is sufficient for generating sentience, this may be the case with an artificially sentient being as well). My intuition is that we shouldn't have that kind of power. We should only be able to create them and let them be as they are, we should not have some God-like control over their existence, even though we are responsible for them existing, it doesn't give us ownership over it. At best we have ownership of the engineering process, we may take credit for the complexity of the engineering task, but the outcome is something we have give its own free-standing existence. This seems unique to engineering sentient systems. Most systems we not only should get credit for engineering them, but we also have ownership over the system itself and can do with it as we please. But a sentient system must be given an existence of its own, just our biological offspring should be given a life of their own. We would be parents to our engineered artificially sentient offspring just as we are parents to our biological offspring - the responsibilities would be the same I think.

=================

You will often see in neuroscience papers that they are accounting for noise in the brain when interpreting brain scans, e.g. associating a region of the brain with some ability by taking a baseline reading and then another reading when they are performing a task and subtracting the difference, with some filtering for noise. It occurs to me that there is likely to be very little noise in the brain itself, and that all activation is purposeful. It is only viewed as noise in relation to some subset of neuronal activity and when one is trying to isolate a neural correlate to some behavioral quality, but the brain doesn't work in such a modular way. Activation that is seen as noise to one neuronal subset is the main activity of another neuronal subset, and even if it is activation not directly associated with some activity, it may be activation that is serving as system maintenance (e.g. keeping some pathways strong so they do not lose their connectivity or associated firings) or it could be preliminary activation for coordinating some future activation. It could also be activation that is initiating inhibition of some other region from the regions that are being activated for the purpose of performing some bodily task. It’s also likely that there is a great deal of activation that is going on just to maintain the various body systems and keeping them in sync as they are controlled by the nervous system.

I think it is very flawed to view the brain as consisting of interconnecting modules where each module serves some task. Yes, there do seem to be well-defined regions that can be associated with various behaviors or memories (penfield maps) but that doesn’t mean they will be as such out of the context of all the other neural firings surrounding those maps. They are embedded into this larger network and connected to other activating and inhibiting paths, and so for a map to function in the context of an organism’s nervous system, it has to be highly integrated and cannot be meaningfully viewed as just a module connected to other modules. Its processing is highly dependent of the processing of those other “modules” or mapped regions, not just in an input/output sense but entirely. This comes out of the fact that there is no central coordination between sub-regions in the brain. It has to happen on its own and this is why inhibitory networks are crucial - it allows sub-regions to coordinate their behavior in a self-organizing way. So even if you do map a region that seems more or less responsible for some behavior, it is only so in virtue of the context it exists in, and it requires all the various activating and inhibitory paths it connects to and is generally dependent on, perhaps not even through explicit connections but through a temporal coordination through brain oscillations.

===================

I am certainly in favor of pushing forward and making an honest attempt at producing artificial sentience. I’m sufficiently convinced that it’s possible and I think we will at least uncover within my lifetime some basic foundational features that need to be present in a system in order to allow sentience to emerge. However, I do not think we as a society are ready for artificial sentience, that is, I think we are collectively too immature to handle it with grace. If we built, say, a humanoid that was somehow theoretically provable to be sentient (by some isomorphism argument to the human brain let’s say), it behaved like a sentient being, and we really have no legitimate reason to doubt its sentience any more than we can doubt the sentience of another human. Most would marvel at this I think, particularly because it was built from our own sweat and tears. But, there will no doubt exist some portion of humankind that will be in great opposition to this endeavor. It may be that they simply don’t believe it is sentient but still see it as a threat, or that they do believe it to be sentient but still see it as a threat, or they may see the entire endeavor as misguided for whatever reason, perhaps religious, and believe that it is not our place to be playing God. Some may also believe that only humans are allowed to be as sentient and intelligent, some sort of priority of humankind. Whatever the reason, there will undoubtedly be opposition, especially in the nascent days of artificial sentience.

I say this because we, as a human race, cannot even get along with each other, and it is some difference or another between people that brings that about. You have countries that are enemies, races that generally dislike one another, religious groups that are prepared to go to war, literally and figuratively, against anyone of a different religion. There are also just differences in policies or collective desires that clash and, unfortunately, the most popular mode by which we resolve our differences is to seek to eliminate the opposition. And in the context of the examples I've identified, that amounts to war, and destruction, and killing, and imprisonment, and perhaps even enslavement. And this is just between our own people! And then artificial sentience walks onto the scene, something that is not just a different person but a different being, different from every person no matter their race or beliefs or origin. If we cannot seem to handle differences within our kind, how are we to handle differences between our human race another intelligent race entirely?

I do believe we have collectively improved in this area, especially within the past century. People in the United States are today way more accepting of other races and religions and sexual orientations. Just in the past generation, racism was rampant, same sex marriages were unheard of, and gender roles were upheld vehemently. Not long before that, women and African Americans could not vote, and not long before that, it was common practice to OWN other people, and have them as personal slaves. In the grand scheme of things, those events are not so distant. But I feel we are becoming more accepting of our differences, especially as we acquire more differences and greater variation, and I think this trend will only continue. This is not to say we will one day be perfect and there will be no disagreements and no more war; I think this is impossible given the scale of the human population. But I think one day we will progress to the point that we are truly ready for another sentient race to join us and to coexist with us.

Right now we are still very wary of artificial intelligence/sentience and I think most are willing to entertain the idea now only because they doubt it will ever really happen. That is, I think most have the underlying suspicion that no machine will ever be sentient in the way we are, and that these "intelligent" robots we seek to build are not to be trusted. This is why most films and science fiction stories about robots and artificial intelligence amounts to some kind of battle of human vs. robots. We mistrust the idea of them and are already mentally preparing for going to war with machines! We haven't even built them yet and already we want to destroy them, because we will believe their first action as sentient beings will be to destroy us! I of course believe this is all misguided, but the point is that I think we need a little more time to get used to the idea that one day, we may not be the only highly intelligent and sentient beings around, and that we might one day have to share this landscape with other beings that are similar to but sufficiently different from us. Hopefully, the time it takes to build an artificially sentient being will be enough time for us to collectively get truly comfortable with idea and we can develop a game plan for how we will handle the day when it arrives.

=================

Watching a video tutorial about developing deep neural networks using the Theano framework, the speaker mentioned that it is naive to suppose that you can get a meaningful classification of an object, say a visual image of an object, just by processing it using one layer of computation. This is because the structural relations in a natural image are too complex to be represented in a single layer of representation. There are nonlinearities within a natural image and so we need a way to capture that nonlinearity in the image. Not only that, but images usually have a hierarchy of structure, with individual pixels forming into low-level basic features, those features being composed into more complex shapes until eventually you have the full complex natural image. There can exist nonlinear relationships at any level in that hierarchy, and thus a multilayer nonlinear processing system is required to get an adequate architecture for representing the input space.

It can be proved that just adding more layers is insufficient when each layer is only a linear mapping of its input space, you can reduce such layers to a single layer without any loss. But when a layer is a nonlinear representation of its input space, you achieve more than could be accomplished with a single layer. Because our natural world is so complex and nonlinearities are pervasive in the natural world, it would make good sense that the brain developed into an architecture that could represent all of the various complexities of that environment. This is good support for why we see a layered structure in the cortical regions of the brain - those layers likely are producing nonlinear representations of their input space and thus progressively processing in layers the complexities of the input from the natural world. If this is what is needed in our artificial neural nets, its likely then that the brain is no exception, and the fact that we do in fact see layering in the brain, it is very likely that that is no accident and is essential to our brains being able to process the complex information coming in from our natural environment.

==========================

With the symbolic approach to AI, everything about the agent has to be explicitly engineered, except perhaps where the programmer employs things like genetic algorithms or allows the system to modify itself over time. But even if that is so, it seems like there is nothing stopping the programmer for explicitly coding in some features that are desired and explicitly leaving out some features that are not. On the other hand, the AI practitioner working from a bottom-up approach will not have as much control on what the agent actually turns out to be like. The symbolic engineer seems to have much more control over what they would like the agent to turn out as, while the systems engineer can only constrain the final product by constraining the possible dynamics of the underlying architecture. Perhaps sometime in the far future we will have enough control over the architecture that the systems engineer could bring about any characteristic desired by designing the architecture appropriately, but it seems for the present and for a long time to come, the engineer will have a method of producing a system that exhibits sentience but can do little in the way of customizing what kind of sentience ultimately comes about.

This is quite analogous to human development. Most parents have some expectations for how their children will grow and what sort of people they will turn out to be, and despite their efforts children sometimes just grow into a completely different person than the parents had hoped for. For example, the parents might exercise great control and send their child to a great school and have them play musical instruments and sports and encourage healthy friendships, only to have the child rebel against such control as a teenager and they become drug addicts. The thing is that children are influenced in very subtle ways as they grow up that we cannot really account for. Some seemingly small event in a child’s life can explode into a huge ordeal that they have to cope with much later in their life, and seemingly minor environmental features can slowly mold them into having a certain disposition or behaving in a certain way. There are also genetic factors to consider, the kind of people that are in the child's life, whether they have a reliable mentor, the personality they come to embody. There are a lot of subtle factors influencing things and I think the same would have to be said of truly artificially sentient beings. We would end up having very little explicit control on how they turn out, and would have to care and nurture them through their infancy in the way we do actual humans.

I suppose I'm assuming that sentience and intelligence has to develop and come to be and mature over time, as human sentience and intelligence does. I think our initial designs will be of this nature, because our initial designs will likely mimic the designs and methods we know work, namely the ones that allow humans to come into existence. That is just a natural place to begin. We will build learning agents and we will have to teach them the way we want them to behave and the qualities they should embody, constrained by their underlying systems (and hopefully we have equipped them with an appropriate architecture and dynamics that can support the kind of behavior we wish them to learn). It may come about in that process though that the agent, despite our efforts, learns to behave differently than desired or anticipated and exhibits rather unexpected behavior. In a way, I think you could say that humans exhibit some rather unexpected behavior that should have been rather unpredictable if someone were monitoring our development over the ages. I'm thinking of some of the really brutal cruelty we can exhibit (ISIS beheadings and torture) and also extreme selflessness (someone giving up their own life to save a complete stranger). If we see these sorts of features in humans, and we end up creating artificial sentience my mimicking the process by which humans become sentience (which I maintain is the most sensible way to start this project, because we KNOW, with certainty, it works), we would expect these artificially sentient beings to exhibit some of the same features, one of which being unpredictability of long-term behavior despite explicit efforts to impose desired behavior that is different.

============================

I just read a paper titled “Computer systems are dynamical systems” by Mytkowicz et al. which made the shocking conclusion (to me) that digital computers are in fact nonlinear dynamical systems. They ran a simple program on two different Intel x86 processors (same ISA), and found that for a run on one processor it exhibited periodic dynamics while on the other the low-level dynamics were chaotic. I have been under the impression that a digital computer and the brain were operating under different physics in a sense, but really they both are exhibiting nonlinear behavior in their lower level systems.

I think the difference then is that computers seek to hide those nonlinearities, or mask them, since it is desired that every time a computer program is run, you get the same results (sans execution time). So even though there are nonlinearities in the low level dynamics, the computer imposes linear behavior at the higher level so we can run our programs in a predictable and reliable fashion. The brain, on the other hand, seems to exploit its nonlinearities and use it as a mechanism for control and self-organization. The linearities actually drive the system instead of just being low level side effects of the nature of the system. This would mean that we would model the computer and the brain the same, using the same physical framework and principles of nonlinear dynamics, but they differ then in how those nonlinearities are utilized and the behavior they then exhibit at a higher level, as determined by those differences in utilization.

=======================

It was noted in class that “we all have the same hardware up there”. That is certainly true in that we all have brains, but I believe everything about human cognition is emergent from the physical brain, and this has the implication that if you change the hardware sufficiently, you can have different cognition emerge (or in the extreme, destroy cognition entirely which I think few would argue against the possibility of, since we have a lot of good evidence that people die). I expect then that in people that exhibit truly exceptional abilities, orders of magnitude beyond your Average Joe, you will see notable differences in their physical brain and likely to predictable areas of the brain that correlate somehow with their abilities. It was also noted that when you have people with a truly exceptional ability, they also have a truly exceptional deficiency (poor social skills or behavioral disorders). It seems very likely to me that it is a matter of real estate being commandeered, or at least forcing the brain to reorganize itself sufficiently to optimize some ability. Perhaps something in the development of the brain went awry and failed to prioritize a certain phase of development so that some region developed too aggressively and forced the other area out.

=======================

The hardest thing to get around conceptually regarding the brain is how in the world something like phenomenal experience can come out it. One thing that I think contributes to this problem is that we view neurons as passing information from one to other, that there are messages that are flying around in the brain at rapid speed from one cell to another. But there is nothing over and above the firing of components, there is nothing in the brain to try to interpret a message from its cells. There just are, cells, and cells firing. Nothing is there to interpret any of it. I think that is a very difficult thing for us to wrap our heads around, because we really want to view each neuron as receiving information and processing it and passing another message onto its neighbors, but more realistically that does not seem to be what is happening. It is just cells in the system responding to pure charges, but together they are doing it in a sufficiently complex way that unexpected things emerge, namely, consciousness and cognition. But it is still profoundly unclear that happens. I suspect that is one step in the right direction though, trying to account for the system itself and not implicitly supposing there are subsystems within the brain responsible for interpretation of the system activity. It just is the system behaving as it does, and somehow it embodies very complex activity and behavior.

======================

We are only able to have abilities and knowledge that our brains are able to represent, i.e. have neuronal maps that are sensitive to input or internal representations and that have the complex structure sufficient for representing the target phenomenon. I'm thinking of the example in class where Clark could not give the directive "turn right" because there was no notion of "right" and he also could know that such a concept was missing because, as noted, the mechanism needed to recognize the absence of something is the same mechanism needed to implement it in the first place. I was just reading also in a book by Judith Dayhoff called Neural Network Architectures that bats are highly sensitive to a frequency range of around 60hz, and they use that sensitivity to detect minor deviations around that range for echo-location, so they have sonar abilities and are able to get around quite proficiently relying on it. Cats, on the other hand, have a more typical broad spectrum sensitivity and no targeted sensitivity to a very small range, so they here acutely in that broad range but have nothing close to the echolocation abilities of a bat.

In general we will have abilities that our brains have been adapted for, and that adaption process will take the form of building the appropriate representations in the wetware of the brain and appropriately integrating that representation into the greater context of everything else that has to be represented to support general organism functions. It seems like a kind of trivial thought but it strikes me as very fundamental, that we will only be capable of having abilities and knowledge that our nervous systems are capable of embodying, and that requires a very rich representational space given the complexity of the environment we are expected to survive in and the complexity of the abilities the human body and mind are capable of. We can indeed say that the brain offers general support for a very rich representational space, given the number of components and connections and dynamics it is capable of, but it has been optimized to embody certain abilities that can be seen as necessary for human survival individually and then collectively as a race. It seems if we took the basic framework but got rid of some of the peculiarities of our evolution and survival, we could impose on this basic structure any specialized representations we wanted, and that would be a way in which we could mold artificial agents into a creature we would like to have around. We could augment and/or constrain their representational space appropriately to be able to support the abilities we wish them to have, and optimize the structures and dynamics and organization to that end.

========================

I think it is true that human life is very different now than it was the millennia ago when the brain was developing into its current form. Of course the development is continuous, but we see a lot of machinery for dealing with an environment that is really nothing like our modern environment. We still have these biological tendencies for mate selection and for defense and hunting, but at least in the developed world these things are not present in the form they were when these mechanisms were of use. This is especially true when you're looking at the world with computers and internet and the infrastructure we have in place for e-commerce. If one has money, one does not even need to leave their home. Everything can be ordered online and delivered right to their door. And the means by which people get money are not always from effort, i.e. people inherit fortunes or win the lottery. A lot of our biological tendencies just aren't relevant any more, though they are perhaps good to keep around in the event our technology and infrastructure collapse.

This same line of reasoning could be applied to consciousness. It would certainly benefit an animal to be conscious and aware and to reason critically about their environment and well being. They would be able to plan for the future and outsmart predators, gauge weather conditions and navigate through dangerous regions with a mind for safety. But it may be that the mind continued to advance to the point that it is now being repurposed for other, more abstract ends. We may still have the basic biological machinery for mate selection and reproduction, but perhaps romantic love is something on top of that in the conscious domain. We may have biological systems for reward and for seeking high states of well-being, but humor and having fun may be what exists in the conscious layer that lays on top of the basic reward system. It may be that consciousness now takes the form of something that rides over top the basic biological machinery that in the past guided our survival and proliferation but now underlies the more complex and abstract facets of human life.

====================

In discussing the capabilities of the brain itself (and not the higher level human abilities that come out of it), it seems like most of the discussion centers on computational power and memory capacity. I think the more relevant thing to be concerned about is representational capacity, i.e. the complexity of representation it is capable of, and how much it can then store when its representations are of a certain complexity. Just knowing byte capacities is not so useful, because it’s unclear that information in the form of bitstreams is what the brain is actually representing. And I think this is a key difference between current digital computers and the brain - digital computers use binary information at their heart, while the brain uses something else. It simply doesn’t matter that you could approximate (or even perfectly represent) the information the brain utilizes with a binary representation. What matters is that the brain is not using that binary representation, and so it needs the more complex architecture it has to represent the full complexity of the environment it is expected to survive in.

One thing I would like to entertain is the possibility that the three dimensional architecture is necessary for the brain to achieve what it does. It seems like current digital computers utilize a 2d architecture, as they use transistors and silicon printed circuits on a flat surface. I suppose it’s possible they could be using stacked boards, but unless there are a lot of cross connections (which I doubt), it is likely a 3d board emulating a larger 2d board. The reason I think this could be relevant, is that there are graph structures that cannot be drawn in 2d space and require at least a third dimension. For example, the complete graph on 5 (K5) vertices is a non-planar graph, as is a complete 3-3 bipartite graph (K3,3; utility problem). And there is a proof showing that any graph that contains either K5 or K3,3 as a subgraph is also non-planar. And these are very simple graphs, so it is expected that a lot of graphs will contain them as subgraphs and have the feature of being non-planar. The only way to construct them then is to make them three dimensional, and then there are no issues (though I wonder if there is a similar notion of non-planarity for higher dimensions, something like non-hyperplanar?).

In a way you could get around this in 2d by just adding another vertex and simulating K5 by not considering the added vertex to be a real vertex, just there for aiding in representation. But then that changes the structure! I want to conceive here of what the actual architecture connections are capable of, and if you go this route it is cheating. Just this simple example seems to indicate that 3d architectures are more capable than 2d ones in terms of representation when representation is taken to happen on the connections themselves or in the dynamics that proceed on the connections and nodes. Does it matter? I'm not sure, but it seems like even just that basic feature of the brain gives it a richer representational space than today's digital computers.

Another source of richer representation may come out of the temporal abilities of neuron firings. They are able to coordinate through oscillations and it seems feasible that representations could be modulated also by oscillations. Neuron firing is highly dependent on the context, i.e. what neurons have just fired in a short history of firings in nearby and distant regions. The brain can then maintain representations over time by appropriate coordination of neuronal groups. Computers use oscillations too, but I think in a different way, i.e. they use them to coordinate forcing the right data through the processor at the right time (such a bottleneck!).

A key area of research is how exactly the brain is representing things in its distributed architecture, and how those representations differ from the representations of a digital computer. I maintain that the representations are indeed of a different nature, and the question then is in what way?

=====================

I think the brain can certainly be viewed as a computational device, but the question is then is that the best way to be looking at it? I take this point from Searle in his book *The Mystery of Consciousness*, he notes that the very notion of "computation" is a relative thing. A window can be viewed as performing a simple computation: window open == true and window closed == false. Sure someone can do computations on the window over time by opening and closing it appropriately and they might be able to do basic arithmetic in this way, or if not with one certainly with a building of windows. Is it then appropriate to call the window or the building a computational device? Sure you can view it as one in this rather contrived way, but of course it is not, not in the way a laptop is a computational device. I think we have a notion of a computational device as a device that is INTENDED to perform computations to some end, and computation itself is not always precise. Which is why Turing's work was so important, he gave a precise mathematical definition to the notions of 'computation' and 'algorithm' and he was then able to prove meaningful theorems about computational devices because it was unambiguous what he meant by it. But our usage in common parlance is not unambiguous and I think it is then misleading to treat the brain as a computational device, as it imposes the idea that the brain is just like a digital computer, and I think the brain is undoubtedly not like a digital computer. They are doing different things, and the very least they are doing things in a fundamentally different way, so why should we treat them as the same?

=====================