Chapter 1:

Could You Become a Computer?

I propose to consider the question, "Can machines think?"

—Alan Turing
Computing Machinery and Intelligence
1950

What would it be like to be an intelligent computer? Would you have sensations and feelings? Would you ever get angry or fall in love? Here is a thought experiment which illustrates what it would be like to become an intelligent computer. The point is two-fold. First, to ask questions about the nature of intelligence. And second, for you to reach your own conclusions about the capabilities and limitations of future machines.

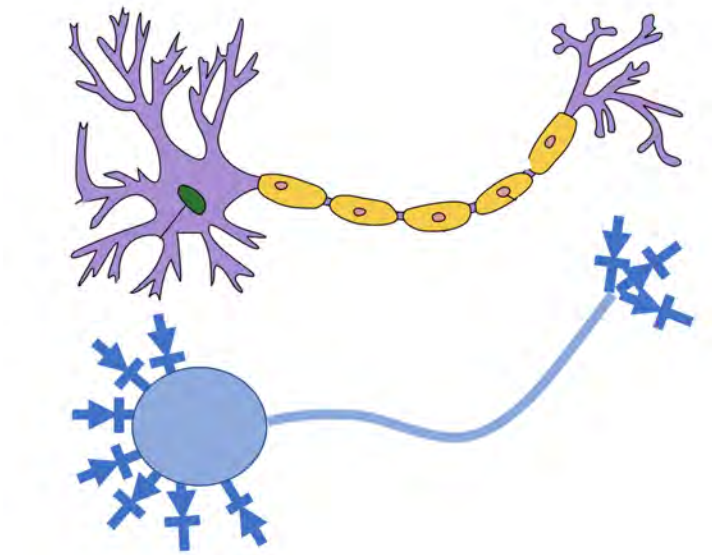
After all, who knows better how you think and what you feel than you do?

Automating your brain

Your brain is a collection of cells called neurons, so suppose we took neurons one at a time and replaced them with artificial neurons. What would you feel like? This is a question which cuts to the center of the definition of intelligence and the possibility of replicating it in non-biological hardware. Remember that this is purely a thought experiment, so we can ignore the technical difficulties.

Since the biological neuron is an electrochemical device, we could theoretically manufacture one with identical functionality from non-biological components such as transistors and capacitors. The practical problems of making neurons of identical size and shape could be insurmountable but imagine that you *could* build an artificial neuron with identical characteristics to an organic one. Consider that artificial joints already fully replace the action of natural ones.

Further, to the extent that intelligence may reside in cells other than neurons, you can consider replacing those with artificial cells as well. For ease of description, we will refer to all the cells involved in thinking



Consider replacing neurons with artificial equivalents built from electronic components such as transistors and diodes—perhaps with a microprocessor. [Diagram of neuron by Quasar Jarosz, license: CC BY-SA 3.0.]

collectively as "neurons" residing in the "brain", even though other types of cells may make a contribution. Indeed, cells outside the brain may contribute as well.

Imagine further that after significant research and development, we have an artificial neuron which can be implanted into your brain. It would take the place of any individual neuron and have an identical function.

Through our hypothetical, completely painless microsurgical techniques, we will remove a single neuron from your brain, measure its characteristics, and replace it with one of our artificial neurons which has been adjusted to fit perfectly. As we believe that individual neural synapses (which form the connections between neurons) harbor our memories, we would be very careful to adjust the simulated synapse transmitters and receptors of our artificial neuron to exactly match the neuron we replaced. Let's place the biological neuron we removed in a nutrient flask to keep it nourished for safekeeping.

Would you feel any different? Not in the slightest. We could simply have removed the neuron and not replaced it—in your brain, neurons die every day and you don't even notice. But by replacing the neuron with our artificial one, we can repeat the process as often as we like without having to worry about the possibility of depleting your brain.

So instead of replacing a single neuron, let's replace a cluster of a thousand neurons with an artificial set. Again, these neurons are perfectly adjusted to match those which were removed. For example, we could replace part of the visual cortex which processes the incoming image of the lower right-hand corner of your right eye. Again, we would save the neurons we removed in the flask. Your brain would still work the same way and you would not feel any difference.

Suppose we repeat the process and end up by replacing all the neurons in your brain with these precisely adjusted artificial neurons. You still would not notice the difference. The artificial brain in your head would be working in the same way as the one which had been removed. You would still be "you" and you would still feel like "you".

Now a question: is *your* brain in your head or in the nutrient flask? Many would think that an *artificial* brain is in our head and our *real* brain is in the flask. And to the extent that you feel and believe that your sense of what is you is in your physical brain, where is the real you? In your head or in the flask?

This is exactly the point: what is the *real* you? If we had not bothered to maintain the structure of the original neurons, now, whatever is in the flask is *not* you—it's a more-or-less random collection of left-over biological neurons. Maybe it used to be you but it is no more. If anything is to be you, it must be contained in the artificial neurons we have installed in your head. It will be seeing what your eyes see, it will be hearing what your ears hear, it will be feeling what your body feels, it will be remembering what your brain used to remember. What makes you *you* is the structure and pattern of the neurons—their connections and the sizes and types of synapses.

As we were replacing neurons, at what point were you transferred from the biological to the artificial? This is a question which springs from the concept of "you" as a single specific entity. Rather, consider yourself to be the sum of processes and behaviors which go on in, and are controlled by, your brain. If we began to replace the fibers in a wooden beam one by one with carbon fibers, the question of when the beam stops being wooden and becomes carbon is a similar matter of definition. The fact is that the beam would be performing a function which it can continue to perform while its fibers are being replaced one by one. So if we chose to say that the beam became a carbon-fiber beam when 50% of its fibers had been replaced, we might just as well say that you possessed an artificial brain when 50% of the neurons were replaced. We might contend that some neurons or certain areas of your brain are more important than others. If we replaced the prefrontal lobes first, for example, the "you" would have been transferred to artificial neurons sooner. Fine, but the exact point of transfer is not a meaningful thing to look for. What gives you intelligence and your sense of yourself doesn't seem to be localized to a specific part of the brain.

If I am asked where in my body *I* reside, I respond that I feel that I am behind my eyes. I feel this way because my eyes' position defines my point of view. I feel this way even though I know that visual signals enter the brain at the very back of my head, so my face is really facing the opposite direction from that which my brain "sees". People whose principal spatial sensation is through touch or hearing might have a different perception of where they reside—but whatever the feeling is, replacing biological neurons with the new identical electronic ones would not make a difference to the feeling. The brain will work the same way and will process visual signals, for example, the same way. You will still have the same point of view that you did before the replacement.

Faster and bigger

Suppose all these artificial neurons could be simultaneously adjusted to make them faster. Whatever the function of the biological neuron, imagine our synthetic neurons could do the same function twice as fast as before. Would that make you smarter? Is having a faster brain equivalent to being more intelligent?

The biological brain is a delicate balance of neuron speeds so the initial effect of speeding up its neurons would have some unintended results. From your brain's perspective, the general effect of speeding up the neurons would be that the world would seem to run in slow motion. But although watching a movie in slow motion makes some details easier to see, it also makes speech virtually impossible to understand. The early years of your life which you spent learning to understand speech only taught you to understand it within a fairly narrow band of speeds. Either much faster or much slower speech becomes progressively more difficult to understand.

Further, while your perception could be somewhat quicker, your muscles would continue at the same speed, so they would seem to run in slow motion. Your muscular coordination would be reduced and your ability to speak would be impaired—similar perhaps to being intoxicated. All-in-all, the initial effects of having a faster brain would be to make you seem somewhat less intelligent.

But brains are adaptive, and you would be able to re-learn how to walk and talk with a brain which is faster than your muscles. Then you would be able to make use of your faster brain. You would have a faster reaction time and would be able to answer *Jeopardy!* questions more quickly. But you wouldn't be able to answer *Jeopardy!* questions that you could not have answered previously. Having a faster brain won't make you know more, it won't make you wiser, or give you a bigger vocabulary.

Things that you haven't been able to learn in the previous 20 years of your life won't suddenly become obvious to you. If it takes you 20 mental repetitions to memorize a phone number today, it will still take you 20 repetitions to learn it with a faster brain. You'll just be able to learn the phone number in less wall-clock time. That is, since all your perceptions will be faster, you won't even perceive that you can memorize faster—only that the clock on the wall will not have progressed as far while you accomplished the given mental task.

With a brain speeded up by a factor of 10, I could do simple arithmetic faster than an average human. I would be able to read books more quickly and I might be able to react more quickly to any given situation. But even with this greatly increased mental speed, I might not appear much smarter than I am today.

On the other hand, if you'd had those twice-as-fast neurons since birth, then at three years of age, you would have the behaviors of a six-year-old. You could have graduated from college at 11 instead of 22. You'd seem pretty smart. Perhaps with that accelerated mental ability, you'd have the time to accumulate more knowledge and experience and be able to make better decisions.

But perhaps your brain is already using its maximum abilities (we'll consider that later). If so, a faster brain gets you to adult-level thinking at an earlier age but doesn't make you an Einstein, because you won't have any thoughts which you wouldn't have had otherwise. You'll just have those thoughts at an earlier age.

Suppose we increased the size of your hypothetical brain with many additional synthetic neurons. Would that make you smarter? Again, yes and no. Today, it is not clear that any particular mental activity you attempt is limited by the number of neurons in your brain. I speculate that the things we learn are limited by our ability to apply ourselves to study rather than our mental limitations. This is similar to athletic abilities as well. I am confident that with the proper commitment and practice, I could make basketball free-throws pretty well. I am not an athlete, but if I were willing to practice eight hours a day, I would become better than most people (who don't have that level of commitment). There would still be natural athletes who would be much better but, with practice, I would certainly be much better than average. The problem is that I am not motivated to that level of commitment.

Similarly, I speculate that many people with the appropriate level of commitment could master college-level mathematics, for example. The reason that relatively few people actually master college-level mathematics has more to do with motivation, commitment or interest than limitations in mental capacity. Mathematics, specifically, also relies on a progression of learning. Individuals who do not learn the basics in elementary school have a difficult time progressing to higher levels. But again, I would contend that many individuals who did not learn the basics failed to do so for some reason other than mental limitation.

So back to the increased brain capacity. Again, the outward manifestations may not be obvious at all. With a greater brain capacity, I might be able to study both computer science and history with equal vigor. I could probably memorize more facts, for example, but it is not clear that more facts would lead to greater insight.

Instead, I imagine that a greater brain capacity would improve the detail of my memories. I would remember more vividly the scenes from last summer's sailing trip. I would remember the names of more schoolmates from my high school years. Long-term memories would not fade as much. Potentially, being able to remember more experiences might make me a wiser person. If I could make decisions based on a greater number of remembered experiences or recognize complex patterns in situations that other people could not, I would appear more intelligent. I would have a greater memory capacity.

But this is only useful to the extent that I commit myself to filling my memory with something valuable. And I contend that my mental capacity is already greater than my commitment and therefore increasing the capacity will make little difference. With greater mental capacity, I could *choose* to apply myself to learning more things, but I am not necessarily going to do so.

Here, one of the important distinctions between people and future computers becomes highlighted. While mental capacity may or may not be comparable between humans and computers, the computer can be directed to 100% commitment, learning everything it can in a specific field without any time out for eating or sleeping, or even just getting tired or bored. If elementary school students spend an hour a day studying arithmetic, then by the sixth grade they've spent about 1,000 hours at it.

A computer, with identical mental functionality but with complete dedication to that single activity, would cover the thousand hours in about six weeks. Six weeks vs. six years gives the computer the appearance of a substantial edge, even with an identical level of mental ability!

Further, the students' brains are not wholly focused on arithmetic during class. As the student thinks about recess, lunch, vacations, exams, friends and other distractions, how much of the brain is actually devoted to the material? In a computer, with full attention dedicated to the topic at hand, it would appear even smarter.

Brain in the basement

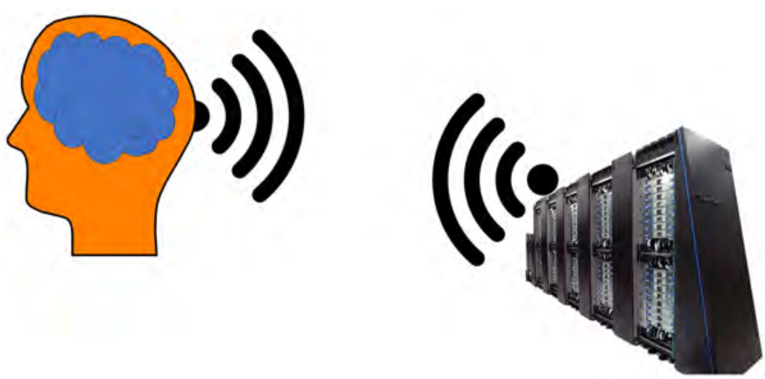
Now let's suppose that we can augment your synthetic neurons with remote transmission capabilities—Wi-Fi, if you will. We can take all the neurons out of your head and install them somewhere else. In your head, we'll install transmitters and receivers which send and receive signals to external neurons. In other words, we can move your brain outside your head. All signals from your senses are sent to your external brain and it sends signals back to your muscles. Now we can put your brain anywhere we like; let's put it in the basement of your home.

After this upgrade, you will have no neurons at all in your brain, only transmitters which are taking signals from your optic nerves and other sensory nerves and sending the signals to your basement. Receivers get transmissions from your basement and control your muscles through your motor nerves. Will you still feel just as you did before? Your brain is in the basement but because even today's Wi-Fi speeds are so fast relative to the speed of your neurons, there is no perceptible delay. Your brain still has the same content. You still have your memories, your knowledge, your biases.

Now, how do you answer the question, "Where do you feel you are?". Your brain is in the basement but you'll still feel as though you are behind your eyes. Suppose you go into the basement and look at your electronic brain. It's probably no more interesting than your furnace; it's just another piece of equipment in the basement. As long as it's working properly, you don't think about it at all.

Now we have the brain-in-the-basement concept, imagine that your friend's brain is in *his* basement as well. Your friend is asleep so you decide to sneak down to his basement and look at his brain and check on his status.

You know he's asleep and you think about what would happen if you temporarily disconnected his brain. Of course, you'll be sure to allow the



Consider connecting the entire thinking part of your brain to electronics at a remote location with a wireless connection to your body. [Image of IBM Blue supercomputer by Argonne National Laboratory's Flickr page, license: CC BY-SA 2.0.]

neurons which control his circulation, respiration, and other bodily functions to continue to operate. So you can go ahead and shut down all the "thinking" parts of his brain without him being aware of it. If he was dreaming before, he will be no longer if his brain is shut down. But will he be aware of the difference? If you were asleep and were anesthetized without being awakened, would you be aware of the difference?

Backups and the passage of time

Suppose there is a power failure. Even if the autonomic nervous system continues and keeps your friend's body alive, the current state of the brain is lost if it was being stored in volatile electronic memory. Being prepared for this possibility, let's invent a system for daily backups. The backup includes the current configuration of every neuron, and details of every synapse. Although this represents a vast amount of data (and we have only estimates of how much) we can use some yet-to-be-invented molecular storage and store it all in a few grams of material.

After the power failure, you need to reload the brain from yesterday's backup. You do this and restart the brain. When your friend wakes up, does he feel any different? He may be confused at the loss of a day. It's Wednesday today and he has no recollection of Tuesday at all, as though he has slept through it entirely. It's as though he's been transported into the future (by one day) by a time machine. For him, the intervening day simply didn't exist. His brain has no recollection of the intervening time except that it can look at clocks and calendars and everyone else seems to agree that the time has passed. It could have been a month or a year if we'd loaded an older backup.

But about that backup… to the extent that it represents the complete state of your brain, it contains the complete "essence" of you. If we can reload your brain from the backup and you are restored, we're forced to conclude that virtually all of what makes you *you* is stored in the backup. In computer science we sometimes think of a collection of data bits as representing a number. We can think of a text character as also being a number. A string of characters is just a number with more digits. In fact, the entire content of Wikipedia can be considered as just a single really big number.

Similarly, then, the backup of you is just a really, really, really big number. The backup of your friend is a different, similarly sized number. Does this mean that, contrary to all appearances and all your fervent desires, you *are* "just a number"?

Distributed intelligence

Getting back to the image of your brain in a box in your basement… Imagine that instead of a single cabinet, the neurons of your synthetic brain were housed in two computer cabinets, with a network interconnection between them. Again, as long as the necessary information can get from neuron to neuron quickly enough, the operation of your brain should be no different. It doesn't really matter how we choose to divide your brain either. You could imagine that the left hemisphere is housed in one cabinet and the right in the other, for example, but any random division would yield the same result.

As long as we can keep the network connection fast enough, it doesn't matter if the two cabinets remain close to each other in your basement. If it were more convenient, we could move one of them next door and your brain would continue to operate and feel the same. Since the speed of electronic signals is so much quicker than the speed of neural signals, it would not become significant until the distance between the cabinets exceeded perhaps 200 miles. At that point, the speed-of-light transmission between the cabinets would take about 1 millisecond and might begin to become significant vs. the 1-millisecond speed of neural pulses.

To move the cabinets further apart, we would need to be selective as to how we divided the neurons between the cabinets. If we generally kept the neurons which were close to each other in the brain in the same cabinet, the transmission delays would probably remain unnoticeable as we moved the cabinets thousands of miles apart. As an example, in watching a movie, if the audio and video aren't perfectly synchronized, you don't notice until the asynchrony approaches a tenth of a second. That's equivalent to putting your visual processing in a cabinet in your basement and your auditory processing in a cabinet halfway around the world.

So if we stay within a 10-mile radius, we can put the cabinets anywhere we want and distribute neurons between the cabinets without any consideration about which neuron goes in which cabinet. Further, if we used smaller cabinets and had more of them, we could distribute the neurons any way we liked. If we had 1,000 cabinets instead of two and distributed them throughout your city, the operation of your brain would remain the same.

If you were to consider what an artificial intelligence would be like if it were running on a network of many different computers, the answer is that it would be, potentially, just like you. You have no awareness of the physical architecture or placement of your brain. The fact that all of your sensations are being received through a single "body" is the factor which gives you a sense of place, not the positions of your various neurons. So if you were to write a program which emulated the function of 100,000 neurons and ran it simultaneously on each of 1,000,000 computers, it could conceivably have a sense of self similar in scope to that of a person. That is, if it had a single location from which it received its inputs, such as a camera signal. If, instead of a single camera, the network was connected to 1,000 different cameras in 1,000 different locations, it would reasonably be expected to have an entirely different sense of self, if it had one at all.

Body swapping

As the development of brain replacements becomes more commonplace, the remote connection between the brain and the body will become standardized. Each body and brain are paired, so individual brains are always operating with the correct body in the same way that today's systems always send the correct phone and Wi-Fi signals to the correct devices using address numbers. But what if the addresses were shifted? You and your friend agree to swap addresses and you go down to the basement and make the adjustments.

Suddenly, your brain is sensing and controlling his body, and vice versa. You are receiving signals from his eyes and ears. You will notice the immediate shift in point of view, but it need not be any more disorienting than a scene change in a movie.

You will notice when you get up that your friend is six inches taller than you, because there is a significant shift in your point of view from having the eyes six inches higher. And maybe your friend was in better physical condition than you and you immediately notice that the body responds more quickly and powerfully. You feel clumsy, though, because your brain has spent years learning the coordinated movements of your smaller body and the timing is off when you reach for things or try to walk. But you are still you, you have just shifted bodies. The part of you which makes you *you* is still safely in your basement.

Now, your mind and your body are completely disconnected. After many people have undergone our hypothetical brain-remoting procedure, you can try out a new body if you want to. Further, since it's the brain not the body that is believed to need sleep, there are usually available bodies for rent whenever their "owners" are sleeping and not in need of them. Swapping bodies is a simple matter of changing addresses.

But suppose one day, all the good bodies are already in use and you don't have one. You are awake but you have no input from the outside world. You are blind and deaf and cannot feel or smell or taste. You have your memories and can think about what you have done already and what you plan to do when you eventually get a body again. You might contemplate the fact that you are just a brain in a basement but you have complete sensory deprivation. How long would you want to remain in this situation? *Could* you remain for long in this situation?

If my desktop computer today *could* think, how likely is it to be able to appear thoughtful? It's blind, deaf, and immobile. It has no physical experience to reminisce about. It has no future for which it can plan. Any expectation that it should manifest human-like qualities is not realistic—regardless of how fast its CPU becomes or how much RAM it has. Today's desktop computer is inadequate for thought. We need interaction with an environment; i.e. a robot. Yes, there is often a camera and microphone on the desktop, but today's computer doesn't have the capacity to understand what it sees and hears in a general way. It's more akin to owning a camera vs. having eyes.

Immortality

You essentially became immortal when we moved your brain to the basement. As your body gets old and wears out, you might simply acquire one of the new fully artificial robot bodies. They're faster and see better but they do have to be recharged periodically.

Are you still *you*? With a robotic body and a brain in the basement, you're completely artificial. If there is to be an argument over this, it is one of semantics, not substance. You might feel somewhat different because you are in an artificial body, but you could not point to a specific time when you stopped being you and started being a machine. Do you feel different than you did? Perhaps the robotic body has no sense of smell or a limited sense of touch. It is certainly a unique feeling when, for example, your foot is "asleep" or the dentist anesthetizes part of your face.

But what about your sense of yourself? It's about the same. You are still aware of the passage of time and have the memories of your childhood and the experiences of your lifetime. When you make a decision to do something, it's based on your previous experience, your preconceptions, and your understanding of the way things will work out. Everything which makes you *you* is still there except for the body and the brain.

My point is: what makes you *you* is all on that backup. The physical structure of synapses in your brain is changed when you learn something. Your memories are believed to be stored in the sizes and types of the synapses themselves. When you acquire a physical skill, learn a song, learn a multiplication fact, or have the experience of walking in the park on a spring day—all these experiences modify the physical structure of your brain.

So what is you is defined by the current state of your brain. Thus, what was you yesterday is not the same as what is you today and cannot be the same as what will be you tomorrow. What makes you *you* incorporates the sum total of the experiences you've had. Suppose all of the basement brains have universal hardware so that other people's backup disks could be loaded on your hardware brain. That would make you someone else!

During the time your backups are not loaded anywhere, what will you be? You simply won't exist—similar to being simultaneously comatose and invisible! But at any time, someone could load your backup disks onto some hardware and you would spring back to life—as good as ever, with no damage done. In the intervening time, though, are you alive? It is the information on the backup which is the "essence of you", so to speak. Whatever hardware that information is loaded on will become you. When the information is loaded, you will "awaken" with your memories and personality intact.

You could arrange for your backup to be loaded simultaneously on more than one brain and truly talk to yourself. You would be completely cloned. It might not be a very interesting conversation because you and your clone would see things the same way, know the same things, carry the same secrets, have the same reactions, and always be in agreement. You would have nothing to learn from your clone and nothing to teach it that it didn't already know. Now if your clone took a three-week vacation while you read a few good books, *then* you'd have something to talk about. But you'd no longer be the same person—you and your clone would be similar but no longer identical. This is entirely different from biological cloning—identical twins—which can create identical bodies but allows the minds to subsequently be filled with entirely different content.

A future artificial intelligence system could have the initial state of a universal hardware brain and a robotic body without any backup data loaded. It has no experience, no memories, no abilities. Whatever it is to do, it must learn. Starting out as an artificial entity would make it entirely different from you. No memories of childhood, no nursery school, no camping trips, no baseball games, only what it gained from its *own* experience. So if it were "brought up" in a human environment where it was treated like a child and expected to play and learn and progress, it would eventually act vastly more human than an identical system which is immediately put to work combing the internet for tidbits of information.

This brings into focus the distinction between man and machine. From the outside, the backup which represents "essence of you" and another which is "essence of robot" appear identical. If *your* backup is loaded into a brain/body, it will be you while the other will be a robot.

One is filled with information about "human" experiences while the other is filled with "robotic" experiences. Imagine that the experiences of the robot were loaded into an artificial brain with a biological body. Would that be more or less human than an entirely artificial body and brain?

Could it really be that the only difference between you and an artificial intelligence will be the content of your mind? It is reasonable to conclude that since our biological "hardware" (both brain and body) are essentially similar to those of humans who lived 50,000 years ago, the only thing which makes us "civilized humans" is that we are brought up to be civilized humans. If the empty artificial system is brought up to be a civilized human as well, couldn't it participate in society with the same abilities and shortcomings as any other person?

Ideas to consider

Here are some thoughts I draw from these thought experiments—perhaps you have had others:

- A computer intelligence unit is not necessarily different from you and me. Given the same sensory apparatus and experiences, an electronic brain could be quite similar to a biological one.
- Experiences, in a biological brain, are stored as the configuration of neurons and synapses. If we could capture this configuration, we could store and preserve the "essence" of a person.
- Having a brain which is faster or has a greater capacity does not necessarily make a person or computer appear more intelligent. The key is in how that speed and capacity are used.
- The electronic brain is not intrinsically senseless, or emotionless, or malevolent, or any of the stereotypes which are found in science fiction. There may be senseless, emotionless, and/or malevolent electronic brains, just as there can be senseless, emotionless, and/or malevolent people.
- The real distinction between people and machines will be in our ability to control how the machines are initially made and with which of life's experiences they are "programmed".

Having explored a fantasy future of thinking computers, we're ready to move on to the reality of intelligent machines.