

[about](#)[archive](#)[minis](#)[the shed](#)[dinner table](#)[store](#)[support wbw](#)

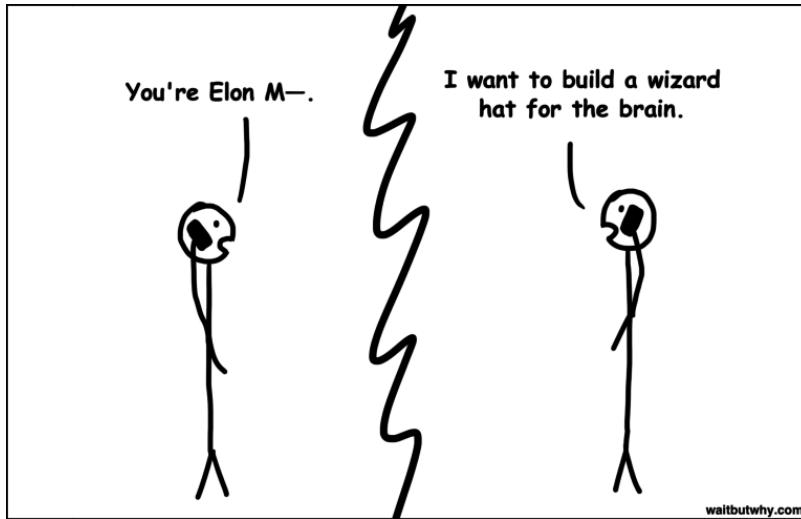
Neuralink and the Brain's Magical Future

April 20, 2017 By Tim Urban

Note: If you want to print this post or read it offline, the PDF is probably the way to go. You can [buy it here](#). And [here's a G-rated version of the post](#), appropriate for all ages.

48.3k
Shares

Month, I got a phone call.



Okay maybe that's not exactly how it happened, and maybe those weren't his exact words. But after learning about the new company Elon Musk was starting, I've come to realize that that's exactly what he's trying to do.

When I wrote about [Tesla](#) and [SpaceX](#), I learned that you can only fully wrap your head around certain companies by zooming both way, way in *and* way, way out. In, on the technical challenges facing the engineers, out on the existential challenges facing our species. In on a snapshot of the world right now, out on the big story of how we got to this moment and what our far future could look like.

Not only is Elon's new venture—Neuralink—the same type of deal, but six weeks after first learning about the company, I'm convinced that it somehow manages to eclipse Tesla and SpaceX in both the boldness of its engineering undertaking and the grandeur of its mission. The other two companies aim to redefine what future humans will do—Neuralink wants to redefine what future humans will *be*.

The mind-bending *bigness* of Neuralink's mission, combined with the labyrinth of impossible complexity that is the human brain, made this the hardest set of concepts yet to fully wrap my head around—but it also made it the most exhilarating when, with enough time spent zoomed on both ends, it all finally *clicked*. I feel like I took a time machine to the future, and I'm here to tell you that it's even weirder than we expect.

But before I can bring you in the time machine to show you what I found, we need to get in our zoom machine—because as I learned the hard way, Elon's wizard hat plans cannot be properly understood until your head's in the right place.

So wipe your brain clean of what it thinks it knows about itself and its future, put on soft clothes, and let's jump into the vortex.

Contents

[Part 1: The Human Colossus](#)

[Part 2: The Brain](#)

[Part 3: Brain-Machine Interfaces](#)

[Part 4: Neuralink's Challenge](#)

[Part 5: The Wizard Era](#)

[Part 6: The Great Merger](#)

NOTES KEY

 = **fun notes**: fun facts, extra thoughts, further explanation

 = **boring notes**: sources and citations

Part 1: The Human Colossus

600 million years ago, no one really did anything, ever.

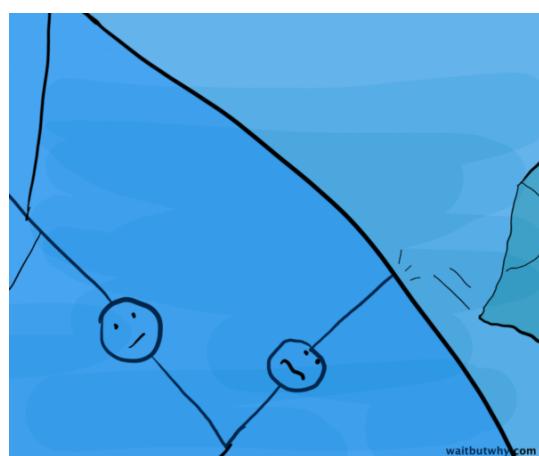


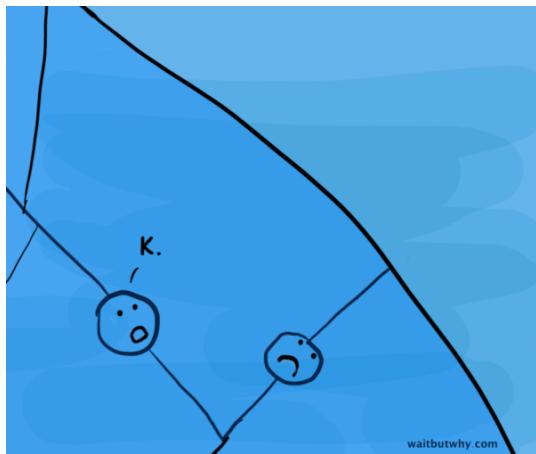
The problem is that no one had any nerves. Without nerves, you can't move, or think, or process information of any kind. So you just had to kind of exist and wait there until you died.

But then came the jellyfish.



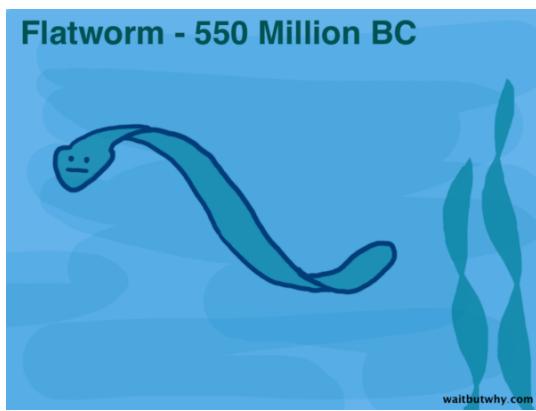
The jellyfish was the first animal to figure out that nerves were an obvious thing to make sure you had, and it had the world's first nervous system—a nerve net.



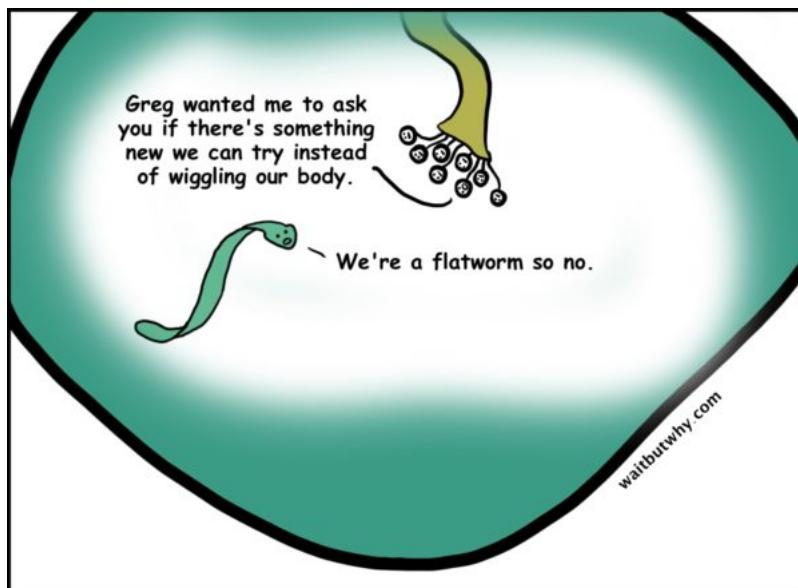
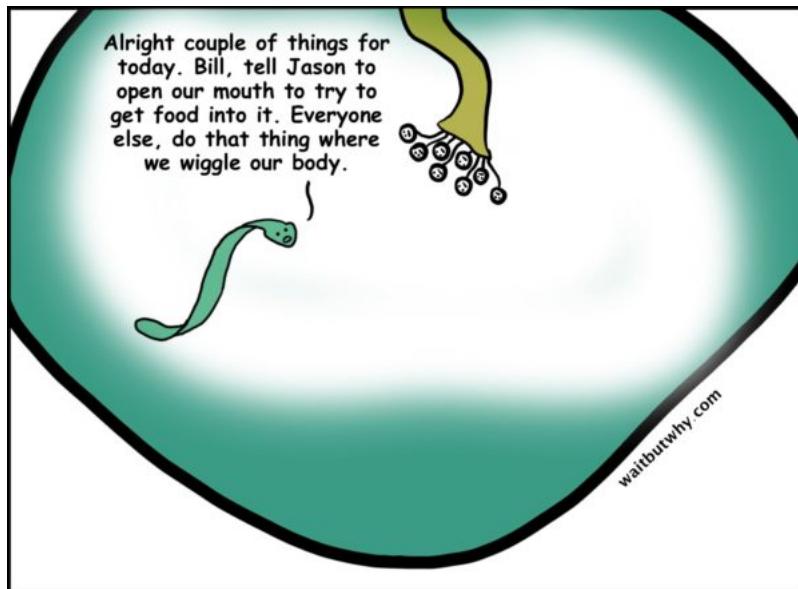


The jellyfish's nerve net allowed it to collect important information from the world around it—like where there were objects, predators, or food—and pass that information along, through a big game of telephone, to all parts of its body. Being able to receive and process information meant that the jellyfish could actually react to changes in its environment in order to increase the odds of life going well, rather than just floating aimlessly and hoping for the best.

A little later, a new animal came around who had an even cooler idea.



The flatworm figured out that you could get a lot more done if there was someone in the nervous system who was in charge of everything—a nervous system boss. The boss lived in the flatworm's head and had a rule that all nerves in the body had to report any new information directly to him. So instead of arranging themselves in a net shape, the flatworm's nervous system all revolved around a central highway of messenger nerves that would pass messages back and forth between the boss and everyone else:



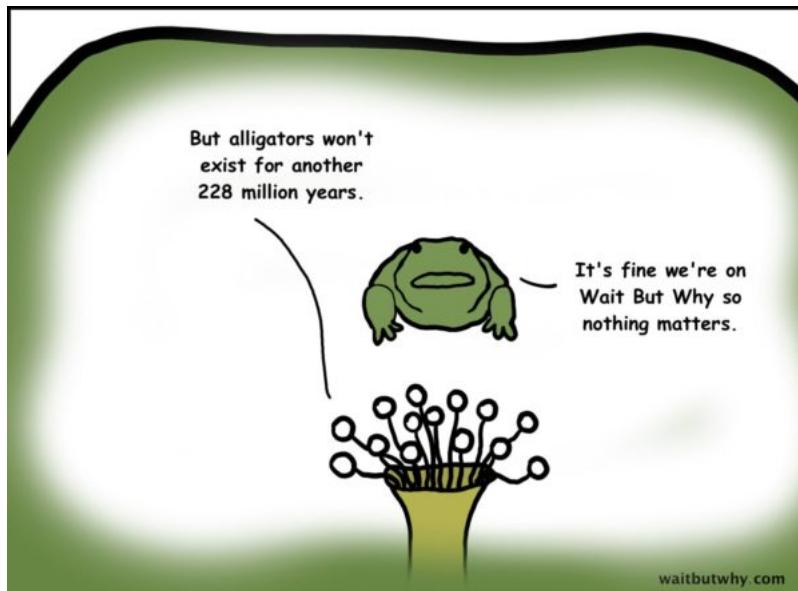
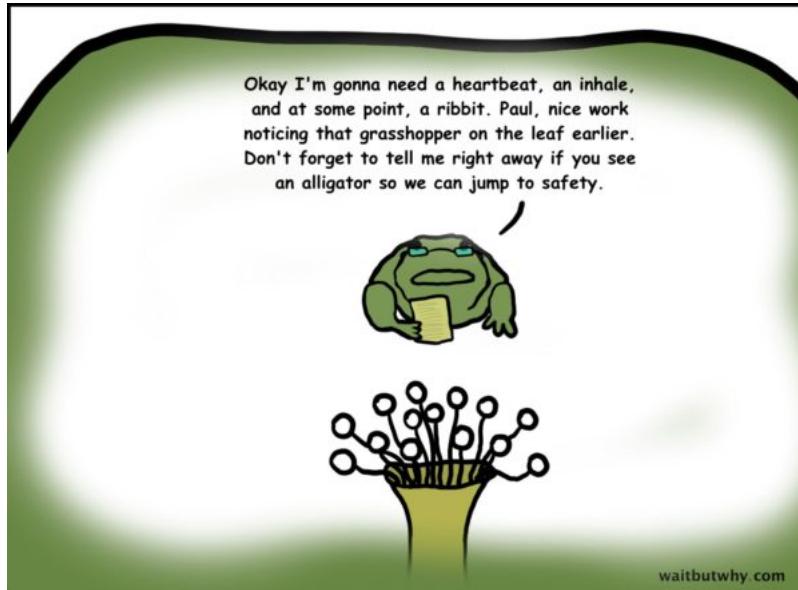
The flatworm's boss-highway system was the world's first central nervous system, and the boss in the flatworm's head was the world's first brain.

The idea of a nervous system boss quickly caught on with others, and soon, there were thousands of species on Earth with brains.

As time passed and Earth's animals started inventing intricate new body systems, the bosses got busier.

Frog - 265 Million BC

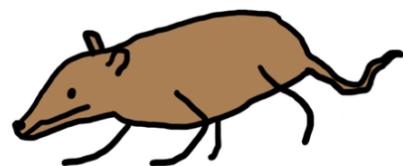


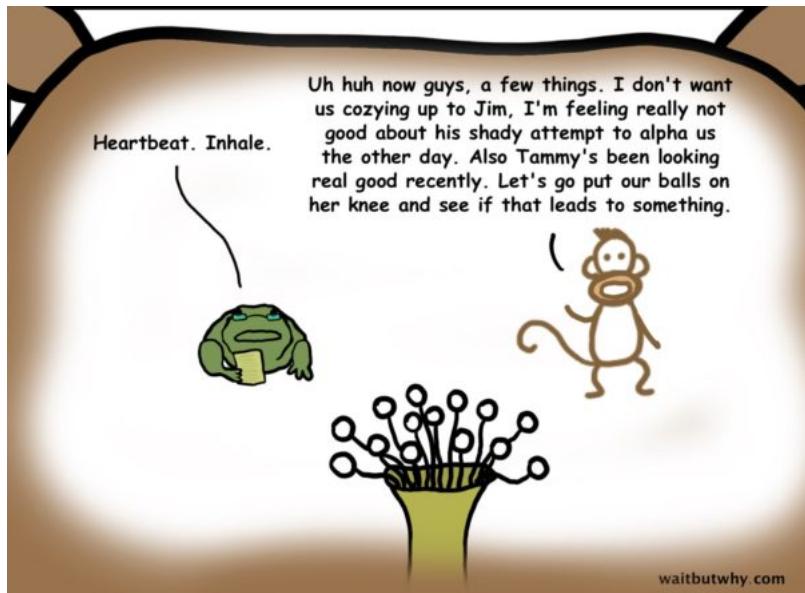


A little while later came the arrival of mammals. For the Millennials of the Animal Kingdom, life was complicated. Yes, their hearts needed to beat and their lungs needed to breathe, but mammals were about a lot more than survival functions—they were in touch with complex feelings like love, anger, and fear.

For the reptilian brain, which had only had to deal with reptiles and other simpler creatures so far, mammals were just...a lot. So a second boss developed in mammals to pair up with the reptilian brain and take care of all of these new needs—the world's first limbic system.

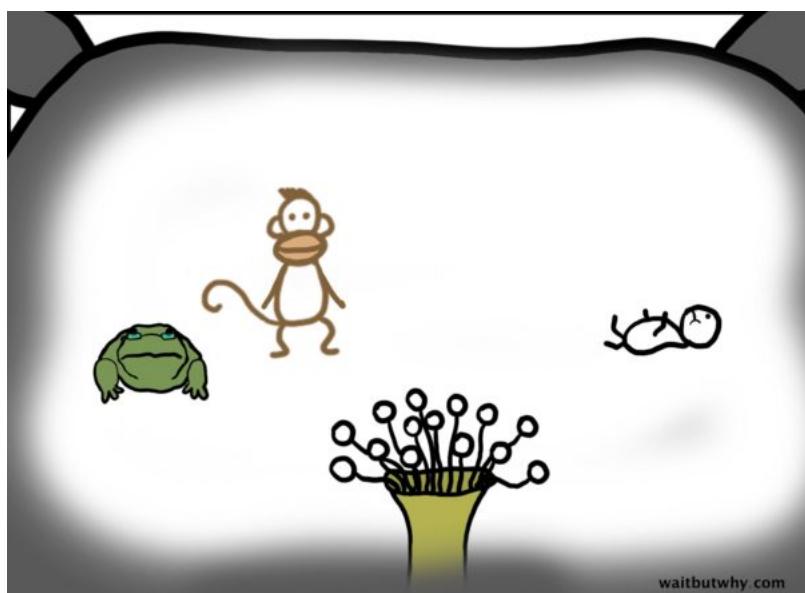
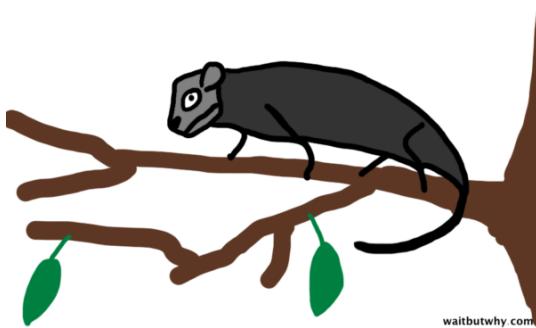
Rodent - 225 Million BC





Over the next 100 million years, the lives of mammals grew more and more complex, and one day, the two bosses noticed a new resident in the cockpit with them.

Tree mammal - 80 Million BC

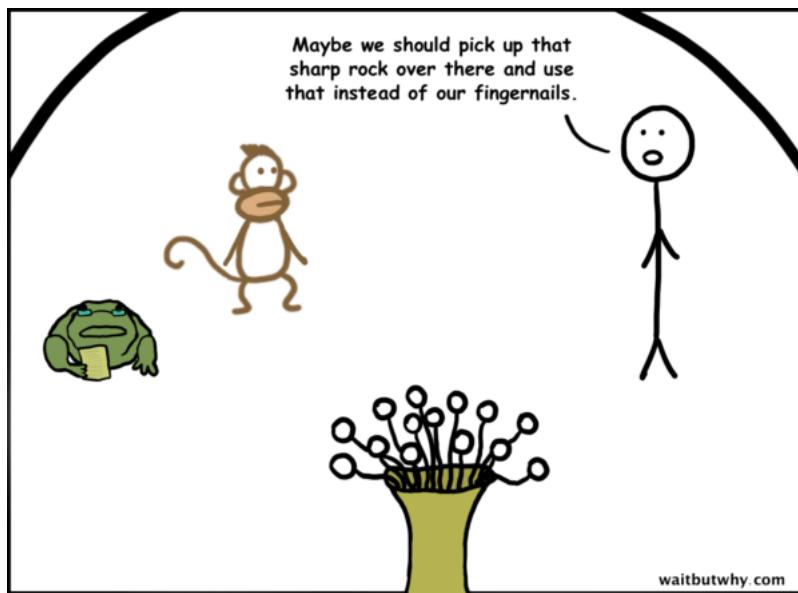
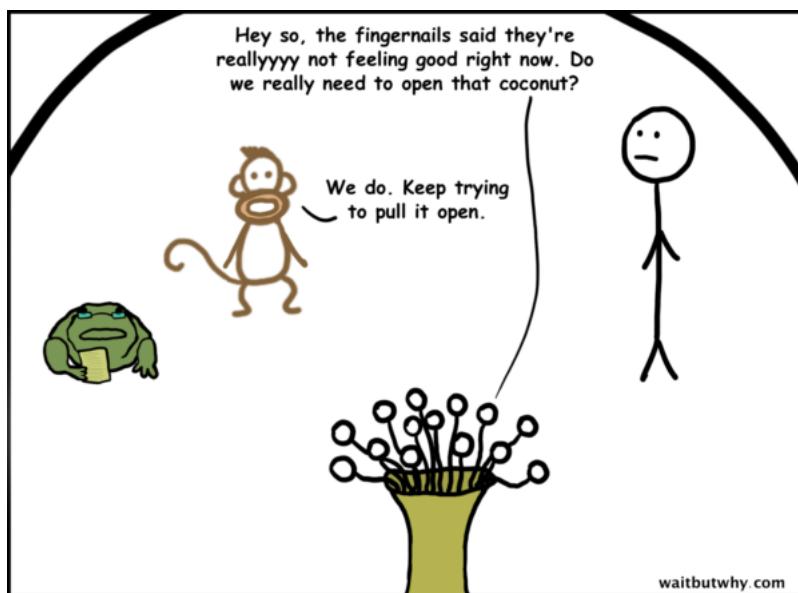


What appeared to be a random infant was actually the early version of the neocortex, and though he didn't say much at first, as evolution gave rise to primates and then great apes and then early hominids, this new boss grew from a baby into a child and eventually into a teenager with his own idea of how things should be run.

Hominid - 4 Million BC

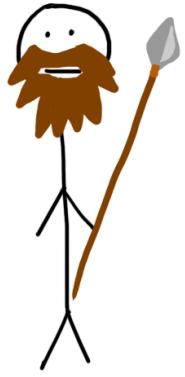


waitbutwhy.com



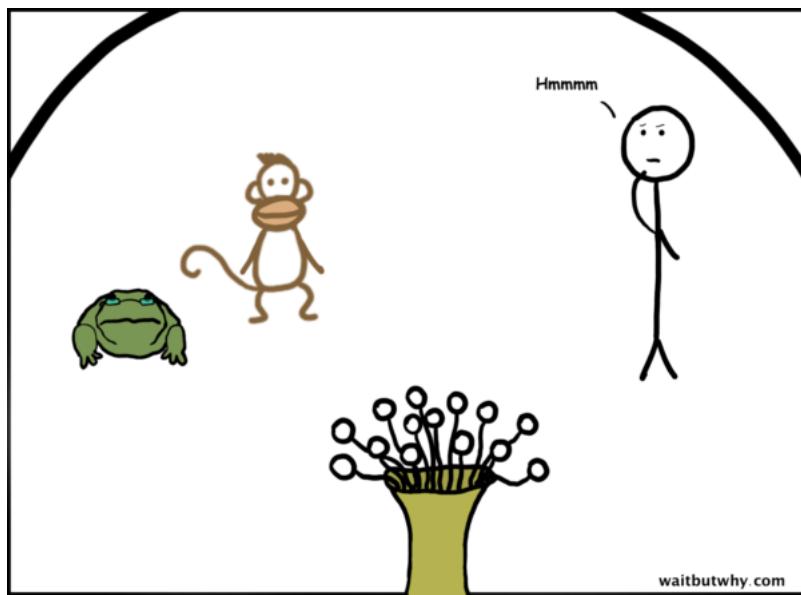
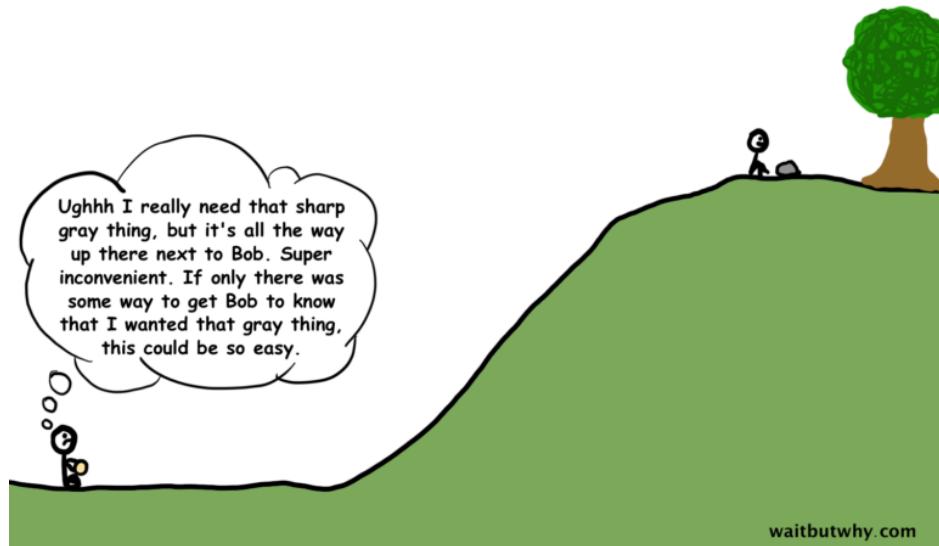
The new boss's ideas turned out to be really helpful, and he became the hominid's go-to boss for things like tool-making, hunting strategy, and cooperation with other hominids.

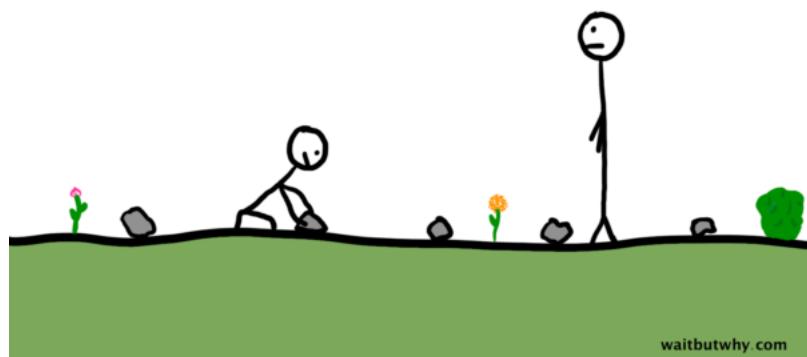
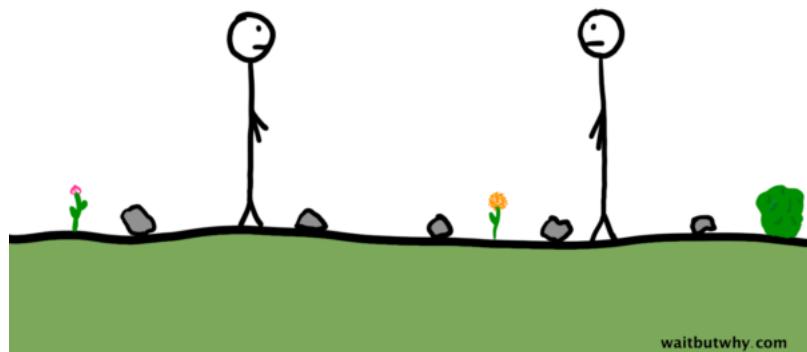
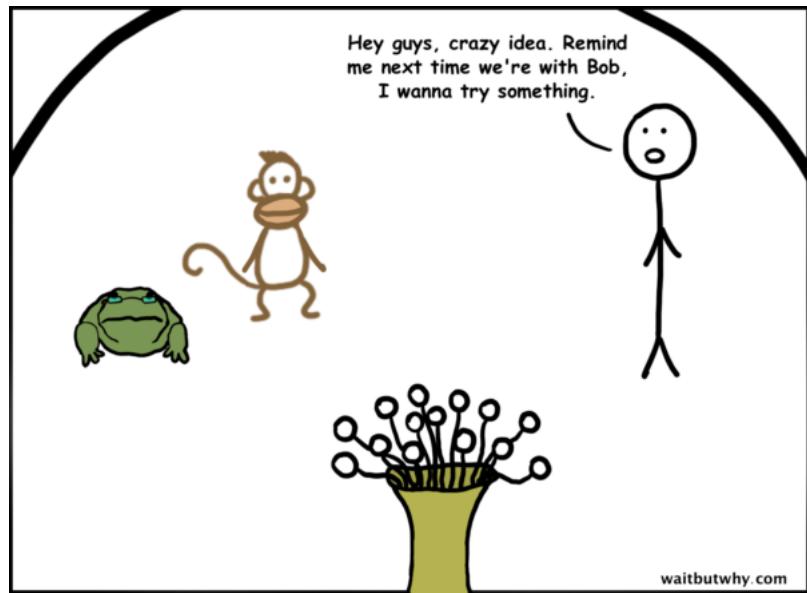
Over the next few million years, the new boss grew older and wiser, and his ideas kept getting better. He figured out how to not be naked. He figured out how to control fire. He learned how to make a spear.

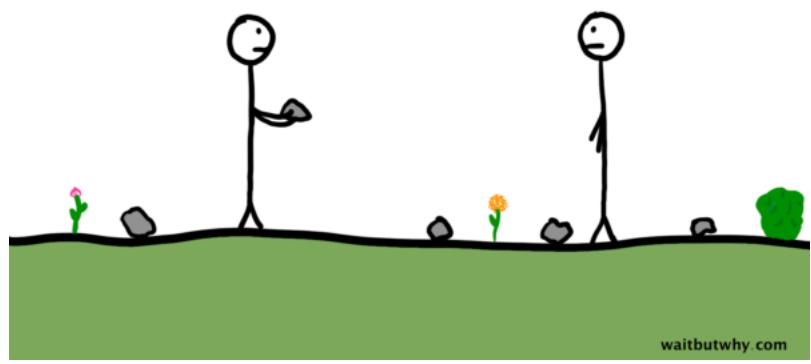
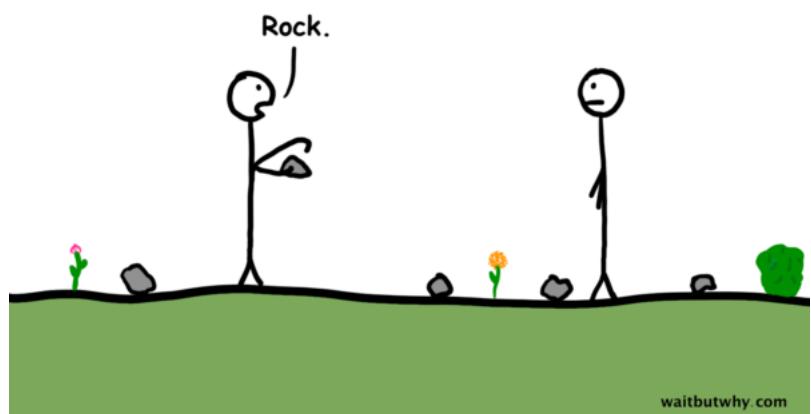
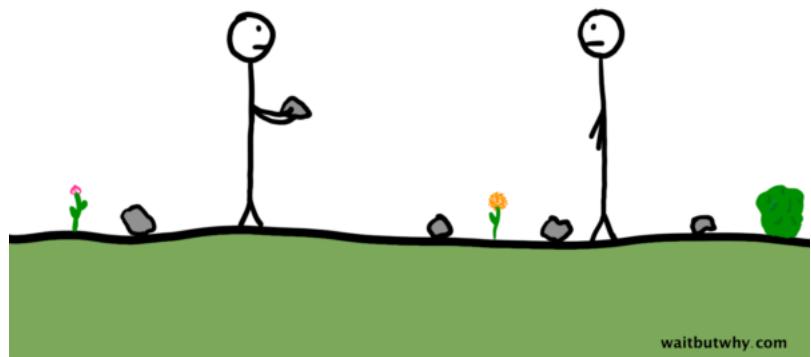


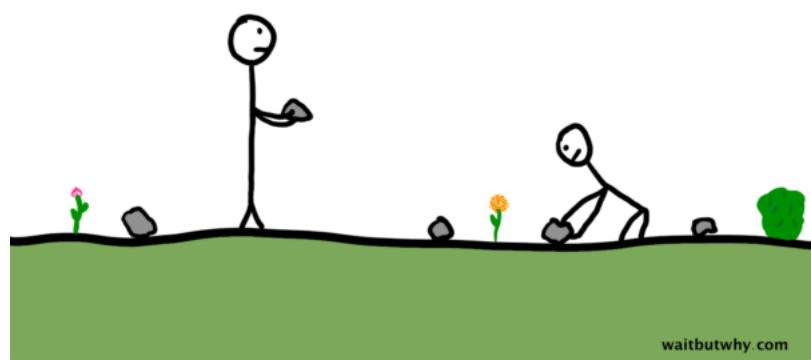
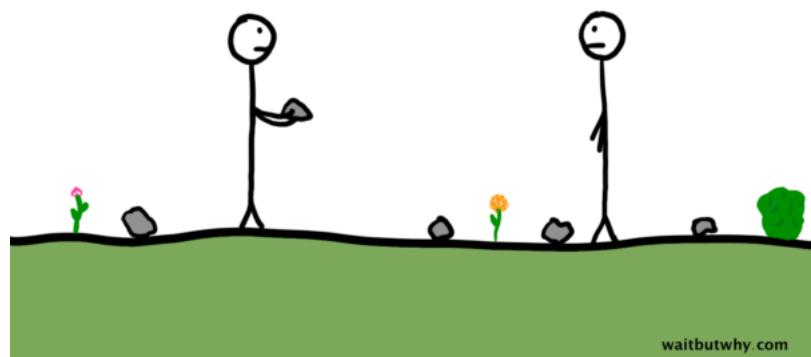
But his coolest trick was *thinking*. He turned each human's head into a little world of its own, making humans the first animal that could think complex thoughts, reason through decisions, and make long-term plans.

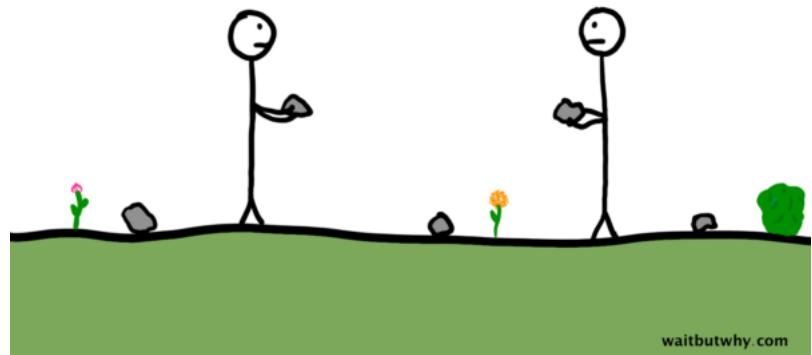
And then, maybe about 100,000 years ago, he came up with a breakthrough.



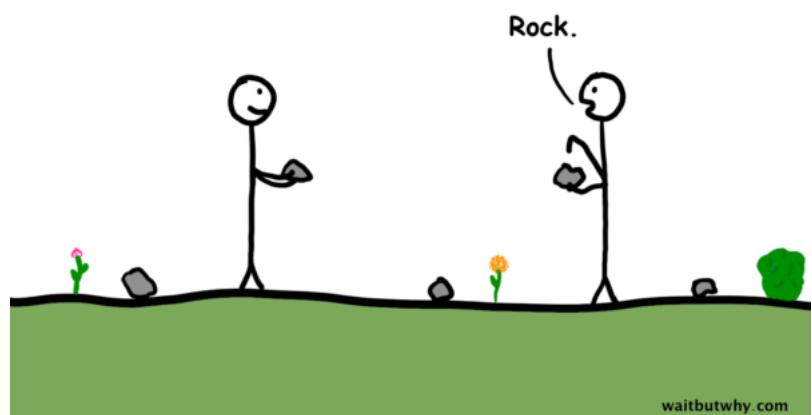




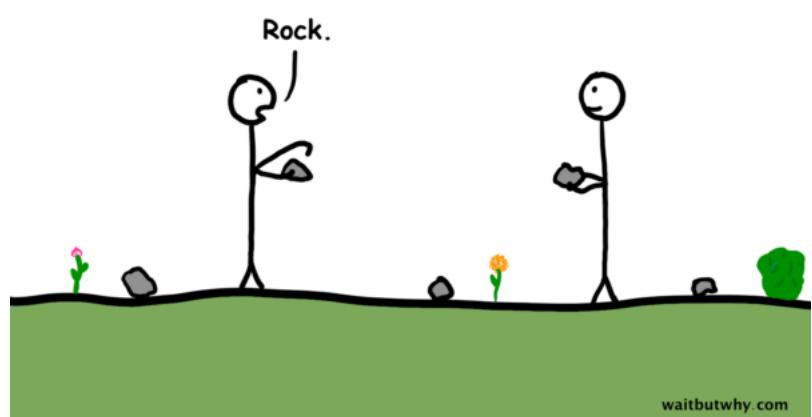




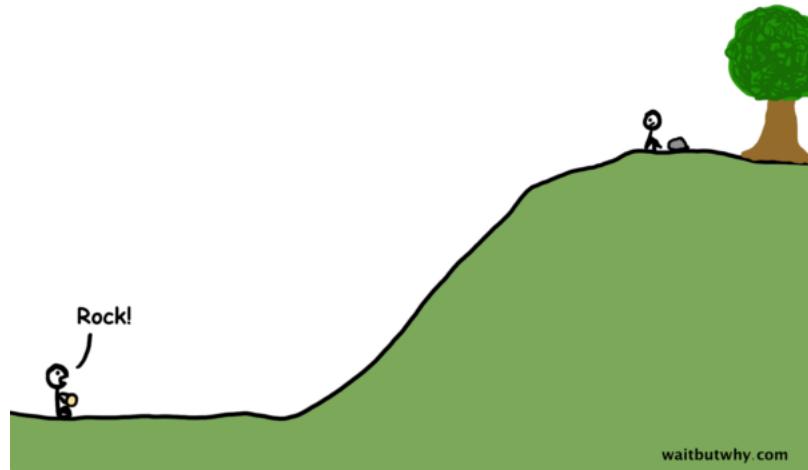
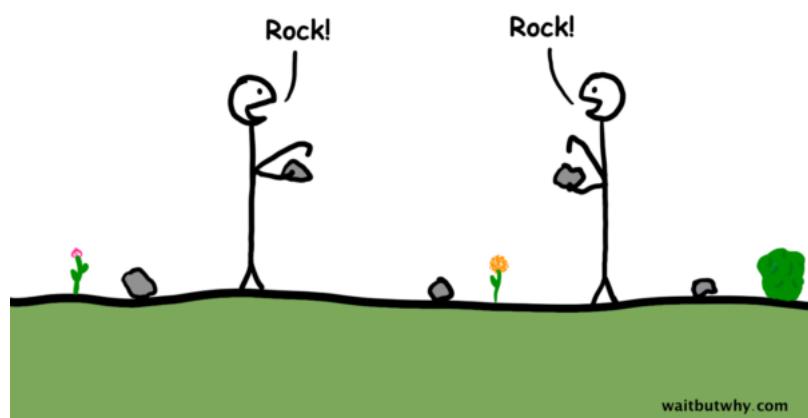
waitbutwhy.com

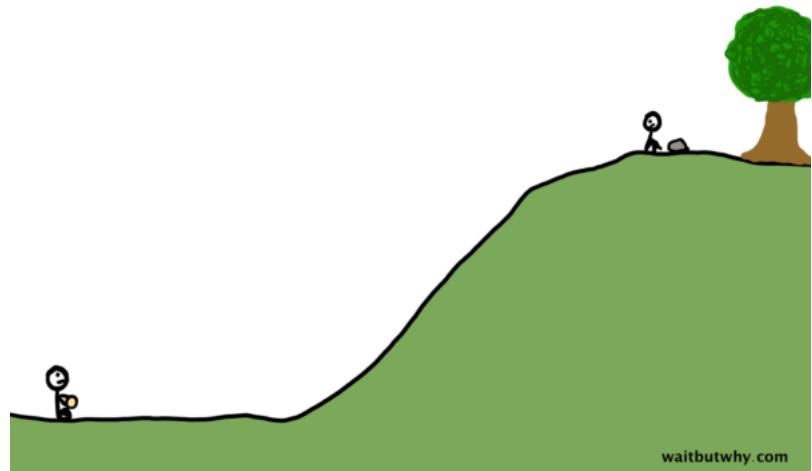


waitbutwhy.com

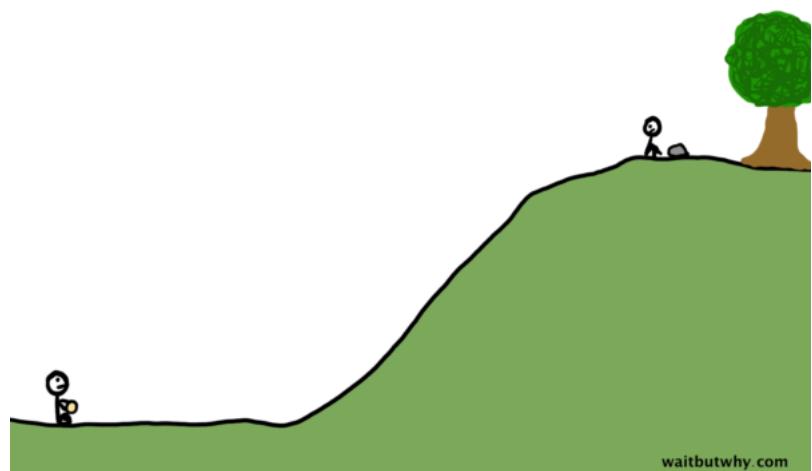


waitbutwhy.com

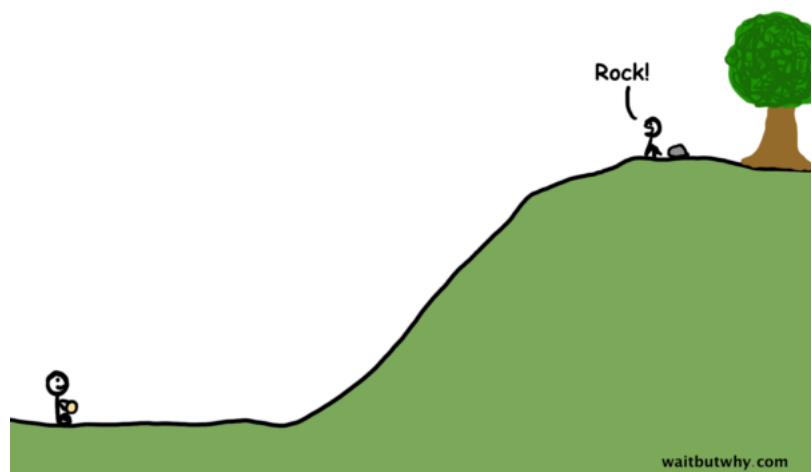




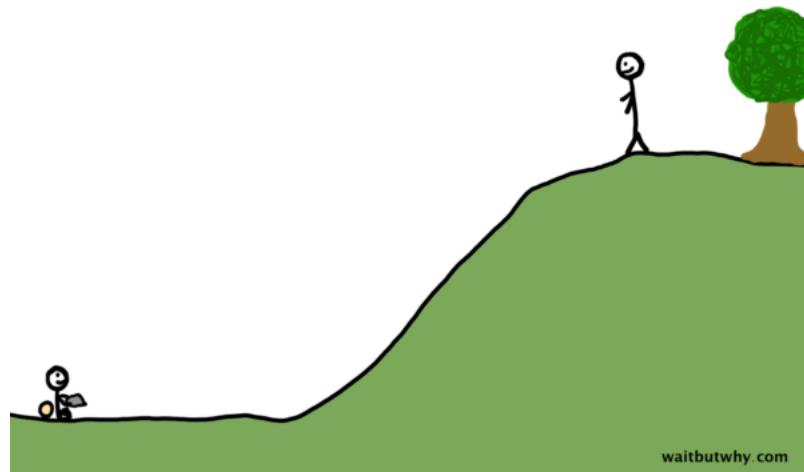
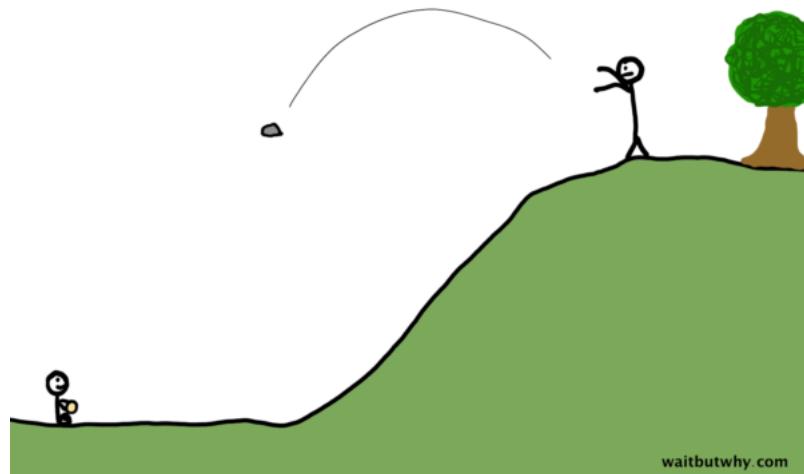
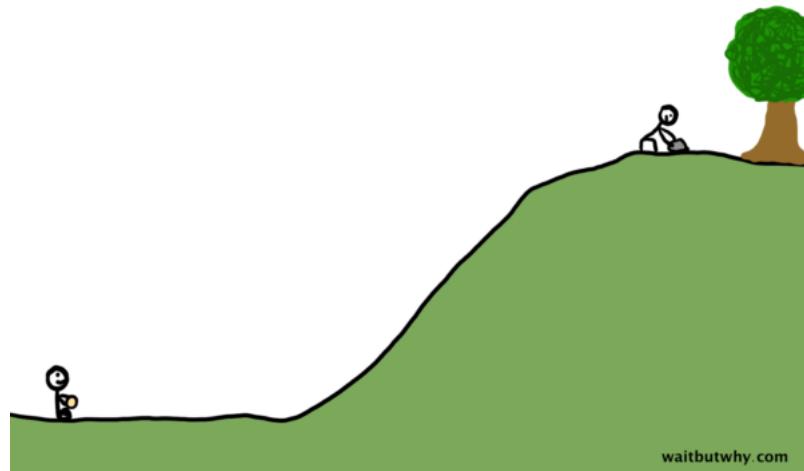
waitbutwhy.com



waitbutwhy.com



waitbutwhy.com



The human brain had advanced to the point where it could understand that even though the sound "rock" was not *itself* a rock, it could be used as a *symbol* of a rock—it was a sound that *referred* to a rock. The early human had invented language.

Soon there were words for all kinds of things, and by 50,000 BC, humans were speaking in full, complex language with each other.

The neocortex had turned humans into magicians. Not only had he made the human head a wondrous internal ocean of complex thoughts, his latest breakthrough had found a way to translate those thoughts into a symbolic set of sounds and send them vibrating through the air into the heads of other humans, who could then decode the sounds and absorb the embedded idea into their own internal thought oceans. The human neocortex had been thinking about things for a long time—and he finally had someone to talk about it all with.

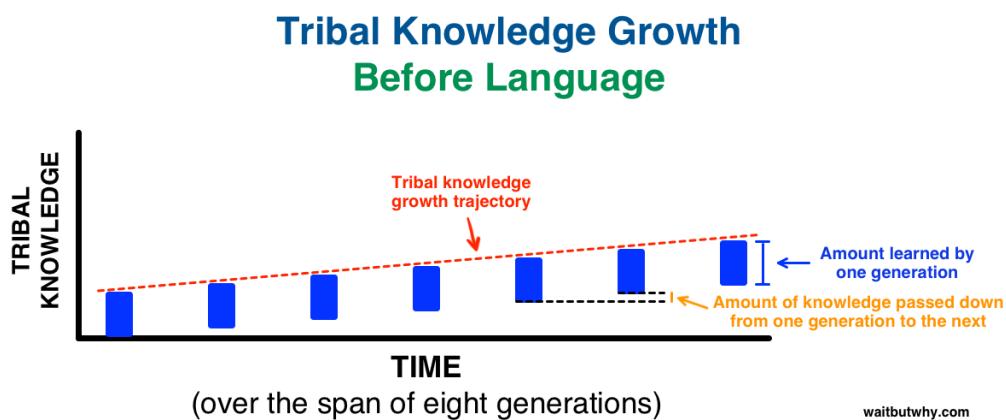
A neocortex party ensued. Neocortexes—fine—neocortices shared everything with each other—stories from their past, funny jokes they had thought of, opinions they had formed, plans for the future.

But most useful was sharing what they had *learned*. If one human learned through trial and error that a certain type of berry led to 48 hours of your life being run by diarrhea, they could use language to share the hard-earned lesson with the rest of their tribe, like photocopying the lesson and handing it to everyone else. Tribe members would then use language to pass along that lesson to their children, and their children would pass it to their own children. Rather than the same mistake being made again and again by many different people, one person's "stay away from that berry" wisdom could travel through space and time to protect everyone else from having their bad experience.

The same thing would happen when one human figured out a new clever trick. One unusually-intelligent hunter particularly attuned to both star constellations and the annual migration patterns of wildebeest^① herds could share a system he devised that used the night sky to determine exactly how many days remained until the herd would return. Even though very few hunters would have been able to come up with that system on their own, through word-of-mouth, all future hunters in the tribe would now benefit from the ingenuity of one ancestor, with that one hunter's crowning discovery serving as every future hunter's *starting point* of knowledge.

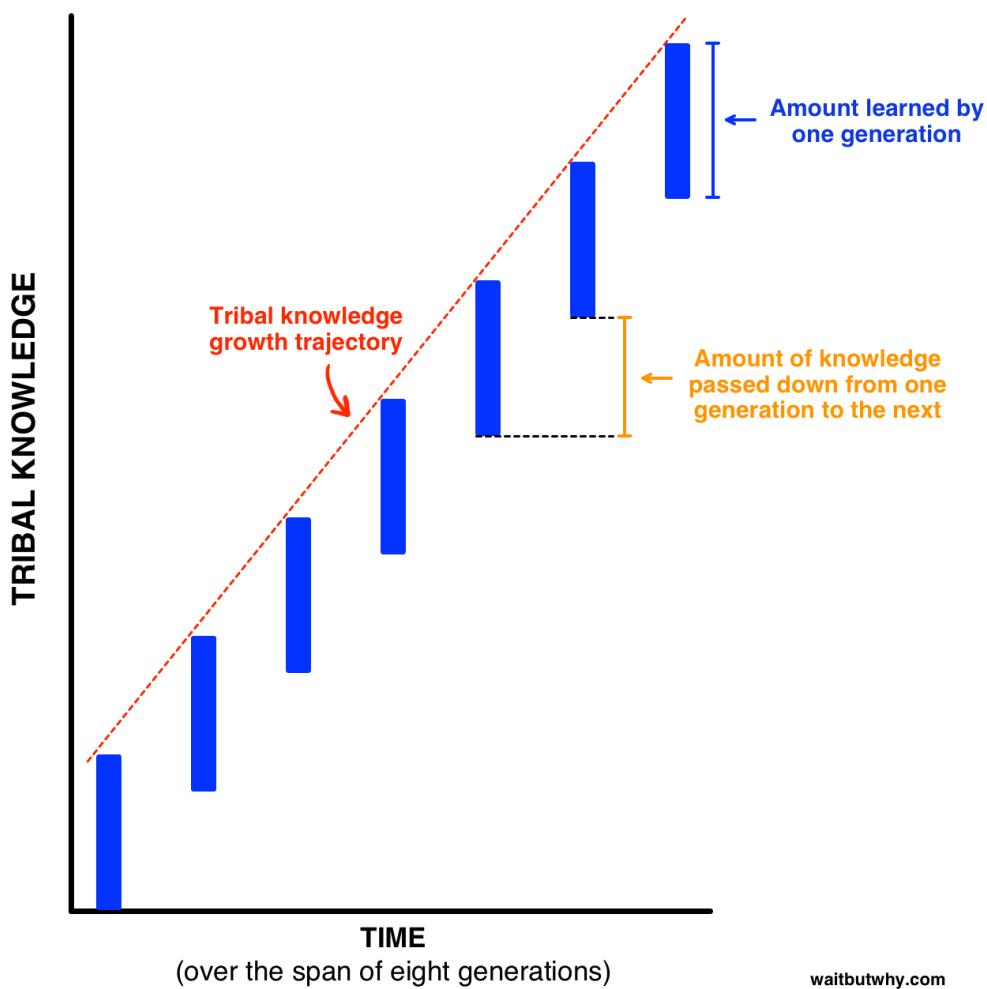
And let's say this knowledge advancement makes the hunting season more efficient, which gives tribe members more time to work on their weapons—which allows one extra-clever hunter a few generations later to discover a method for making lighter, denser spears that can be thrown more accurately. And just like that, every present and future hunter in the tribe hunts with a more effective spear.

Language allows the best epiphanies of the very smartest people, through the generations, to accumulate into a little collective tower of tribal knowledge—a "greatest hits" of their ancestors' best "aha!" moments. Every new generation has this knowledge tower installed in their heads as their starting point in life, leading them to new, even better discoveries that build on what their ancestors learned, as the tribe's knowledge continues to grow bigger and wiser. Language is the difference between this:



And this:

Tribal Knowledge Growth After Language



waitbutwhy.com

The major trajectory upgrade happens for two reasons. Each generation can learn a lot more new things when they can talk to each other, compare notes, and combine their individual learnings (that's why the blue bars are so much higher in the second graph). And each generation can successfully pass a higher percentage of their learnings on to the next generation, so knowledge *sticks* better through time.

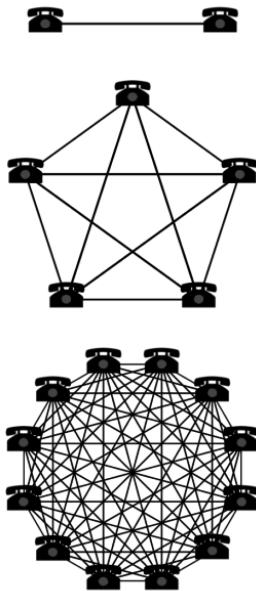
Knowledge, when shared, becomes like a grand, collective, inter-generational collaboration. Hundreds of generations later, what started as a pro tip about a certain berry to avoid has become an intricate system of planting long rows of the stomach-friendly berry bushes and harvesting them annually. The initial stroke of genius about wildebeest migrations has turned into a system of goat domestication. The spear innovation, through hundreds of incremental tweaks over tens of thousands of years, has become the bow and arrow.

Language gives a group of humans a *collective intelligence* far greater than individual human intelligence and allows each human to benefit from the collective intelligence as if he came up with it all himself. We think of the bow and arrow as a primitive technology, but raise Einstein in the woods with no existing knowledge and tell him to come up with the best hunting device he can, and he won't be *nearly* intelligent or skilled or knowledgeable enough to invent the bow and arrow. Only a *collective* human effort can pull that off.

Being able to speak to each other also allowed humans to form complex social structures which, along with advanced technologies like farming and animal domestication, led tribes over time to begin to settle into permanent locations and merge into organized super-tribes. When this happened, each tribe's tower of accumulated knowledge could be shared with the larger super-tribe, forming a super-tower. Mass cooperation raised the quality of life for everyone, and by 10,000 BC, the first cities had formed.

According to Wikipedia, there's something called Metcalfe's law, which states that "the value of a telecommunications network is proportional to the square of the number of connected users of the

system." And they include this little chart of old telephones:¹



But the same idea applies to people. Two people can have one conversation. Three people have four unique conversation groups (three different two-person conversations and a fourth conversation between all three as a group). Five people have 26. Twenty people have 1,048,555.

So not only did the members of a city benefit from a *huge* knowledge tower as a foundation, but Metcalfe's law means that the number of conversation possibilities now skyrocketed to an unprecedented amount of variety. More conversations meant more ideas bumping up against each other, which led to many more discoveries clicking together, and the pace of innovation soared.

Humans soon mastered agriculture, which freed many people up to think about all kinds of other ideas, and it wasn't long before they stumbled upon a new, giant breakthrough: writing.

Historians think humans first started writing things down about 5 – 6,000 years ago. Up until that point, the collective knowledge tower was stored only in a network of people's memories and accessed only through livestream word-of-mouth communication. This system worked in small tribes, but with a vastly larger body of knowledge shared among a vastly larger group of people, memories alone would have had a hard time supporting it all, and most of it would have ended up lost.

If language let humans send a thought from one brain to another, writing let them stick a thought onto a physical *object*, like a stone, where it could live forever. When people began writing on thin sheets of parchment or paper, huge fields of knowledge that would take weeks to be conveyed by word of mouth could be compressed into a book or a scroll you could hold in your hand. The human collective knowledge tower now lived in physical form, neatly organized on the shelves of city libraries and universities.

These shelves became humanity's grand instruction manual on everything. They guided humanity toward new inventions and discoveries, and those would in turn become new books on the shelves, as the grand instruction manual built upon itself. The manual taught us the intricacies of trade and currency, of shipbuilding and architecture, of medicine and astronomy. Each generation began life with a higher floor of knowledge and technology than the last, and progress continued to accelerate.

But painstakingly handwritten books were treated like treasures,² and likely only accessible to the extreme elite (in the mid 15th century, there were only 30,000 books in all of Europe). And then came another breakthrough: the printing press.

In the 15th century, the [beardy](#) Johannes Gutenberg came up with a way to create multiple identical copies of the same book, much more quickly and cheaply than ever before. (Or, more accurately, when Gutenberg was born, humanity had already figured out the first 95% of how to invent the printing press, and Gutenberg, with that knowledge as his starting point, invented the last 5%.) (Oh, also, Gutenberg didn't invent the printing press, [the Chinese did](#) a bunch of centuries earlier. Pretty reliable rule is that everything you think was invented somewhere other than China was probably actually invented in China.) Here's how it worked:

It Turns Out Gutenberg Isn't Actually Impressive Blue Box

To prepare to write this blue box, I found [this video](#) explaining how Gutenberg's press worked and was surprised to find myself unimpressed. I always assumed Gutenberg had made some genius machine, but it turns out he just created a bunch of stamps of letters and punctuation and manually arranged them as the page of a book and then put ink on them and pressed a piece of paper onto the letters, and that was one book page. While he had the letters all set up for that page, he'd make a bunch of copies. Then he'd spend forever manually rearranging the stamps (this is the "movable type" part) into the next page, and then do a bunch of copies of that. His first project was 180 copies of the Bible,
³ which took him and his employees two years.

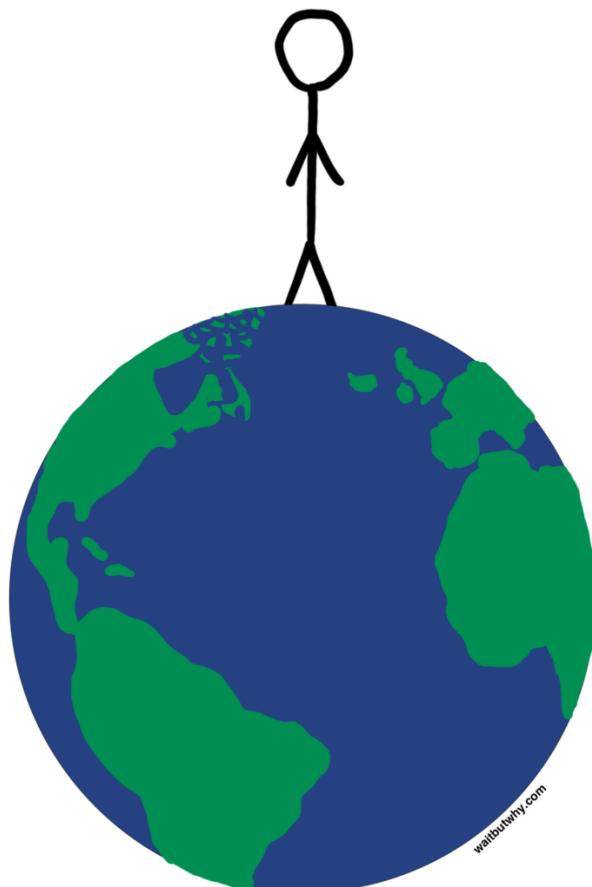
That's Gutenberg's thing? A bunch of stamps? I feel like I could have come up with that pretty easily. Not really clear why it took humanity 5,000 years to go from figuring out how to write to creating a bunch of manual stamps. I guess it's not that I'm unimpressed with Gutenberg—I'm neutral on Gutenberg, he's fine—it's that I'm unimpressed with everyone else.

Anyway, despite how disappointing Gutenberg's press turned out to be, it was a *huge* leap forward for humanity's ability to spread information. Over the coming centuries, printing technology rapidly improved, bringing the number of pages a machine could print in an hour from about 25 in Gutenberg's time⁴ up 100-fold to 2,400 by the early 19th century.²

Mass-produced books allowed information to spread like wildfire, and with books being made increasingly affordable, no longer was education an elite privilege—millions now had access to books, and literacy rates shot upwards. One person's thoughts could now reach millions of people. The era of mass communication had begun.

The avalanche of books allowed knowledge to transcend borders, as the world's regional knowledge towers finally merged into one species-wide knowledge tower that stretched into the stratosphere.

The better we could communicate on a mass scale, the more our species began to function like a single organism, with humanity's collective knowledge tower as its brain and each individual human brain like a nerve or a muscle fiber in its body. With the era of mass communication upon us, the collective human organism—the Human Colossus—rose into existence.



With the entire body of collective human knowledge in its brain, the Human Colossus began inventing things no human could have dreamed of inventing on their own—things that would have seemed like absurd science fiction to people only a few generations before.

It turned our ox-drawn carts into speedy locomotives and our horse-and-buggies into shiny metal cars. It turned our lanterns into lightbulbs and written letters into telephone calls and factory workers into industrial machines. It sent us soaring through the skies and out into space. It redefined the meaning of “mass communication” by giving us radio and TV, opening up a world where a thought in someone’s head could be beamed instantly into the brains of a billion people.

If an individual human’s core motivation is to pass its genes on, which keeps the species going, the forces of macroeconomics make the Human Colossus’s core motivation to create value, which means it tends to want to invent newer and better technology. Every time it does that, it becomes an even better inventor, which means it can invent new stuff even faster.

And around the middle of the 20th century, the Human Colossus began working on its most ambitious invention yet.

The Colossus had figured out a long time ago that the best way to create value was to invent value-creating *machines*. Machines were better than humans at doing many kinds of work, which generated a flood of new resources that could be put towards value creation. Perhaps even more importantly, machine labor freed up huge portions of human time and energy—i.e. huge portions of the Colossus itself—to focus on innovation. It had already outsourced the work of our arms to factory machines and the work of our legs to driving machines, and it had done so through the power of its brain—now what if, somehow, it could outsource the work of the brain *itself* to a machine?

The first digital computers sprung up in the 1940s.

One kind of brain labor computers could do was the work of information storage—they were *remembering* machines. But we already knew how to outsource our memories using books, just like we had been outsourcing our leg labor to *horses* long before cars provided a far better solution. Computers were simply a memory-outsourcing upgrade.

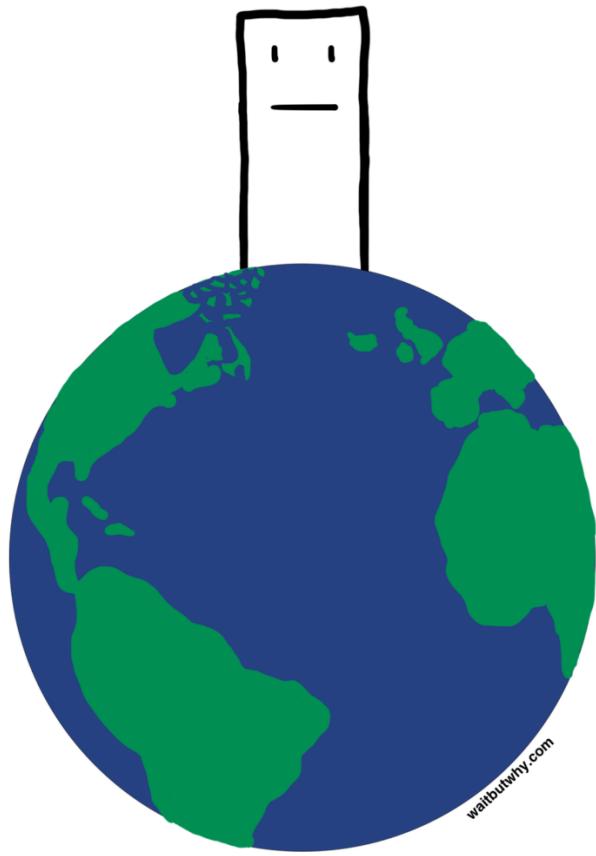
Information-processing was a different story—a type of brain labor we had never figured out how to outsource. The Human Colossus had always had to do all of its own computing. Computers changed that.

Factory machines allowed us to outsource a physical process—we put a material in, the machines physically processed it and spit out the results. Computers could do the same thing for information processing. A software program was like a factory machine for information processes.

These new information-storage/organizing/processing machines proved to be useful. Computers began to play a central role in the day-to-day operation of companies and governments. By the late 1980s, it was common for individual people to own their own personal brain assistant.

Then came another leap.

In the early 90s, we taught millions of isolated machine-brains how to communicate with one another. They formed a worldwide computer network, and a new giant was born—the Computer Colossus.

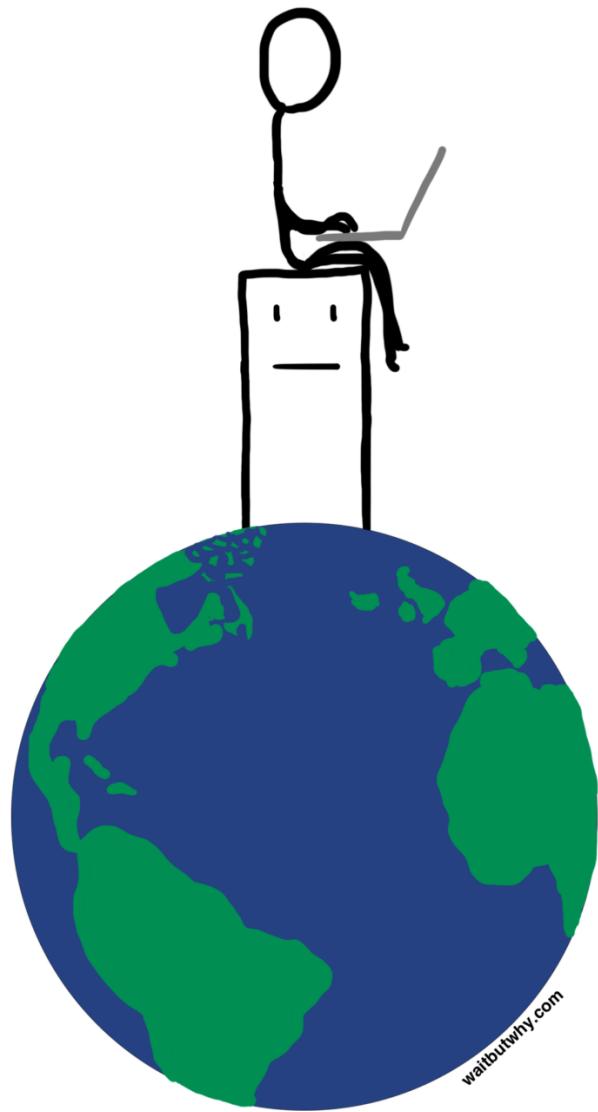


The Computer Colossus and the great network it formed were like popeye spinach for the Human Colossus.

If individual human brains are the nerves and muscle fibers of the Human Colossus, the internet gave the giant its first legit nervous system. Each of its nodes was now interconnected to all of its other nodes, and information could travel through the system with light speed. This made the Human Colossus a faster, more fluid thinker.

The internet gave billions of humans instant, free, easily-searchable access to the entire human knowledge tower (which by now stretched past the moon). This made the Human Colossus a smarter, faster learner.

And if individual computers had served as brain extensions for individual people, companies, or governments, the Computer Colossus was a brain extension for the entire Human Colossus itself.



With its first real nervous system, an upgraded brain, and a powerful new tool, the Human Colossus took inventing to a whole new level—and noticing how useful its new computer friend was, it focused a large portion of its efforts on advancing computer technology.

It figured out how to make computers faster and cheaper. It made internet faster and wireless. It made computing chips smaller and smaller until there was a powerful computer in everyone's pocket.

Each innovation was like a new truckload of spinach for the Human Colossus.

But today, the Human Colossus has its eyes set on an even bigger idea than more spinach. Computers have been a game-changer, allowing humanity to outsource many of its brain-related tasks and better function as a single organism. But there's one kind of brain labor computers still can't quite do.

Thinking.

Computers can compute and organize and run complex software—software that can even learn on its own. But they can't *think* in the way humans can. The Human Colossus knows that everything it's built has originated with its ability to reason creatively and independently—and it knows that the *ultimate* brain extension tool would be one that can really, actually, legitimately *think*. It has no idea what it will be like when the Computer Colossus can think for itself—when it one day opens its eyes and becomes a *real* colossus—but with its core goal to create value and push technology to its limits, the Human Colossus is determined to find out.

We'll come back here in a bit. First, we have some learning to do.

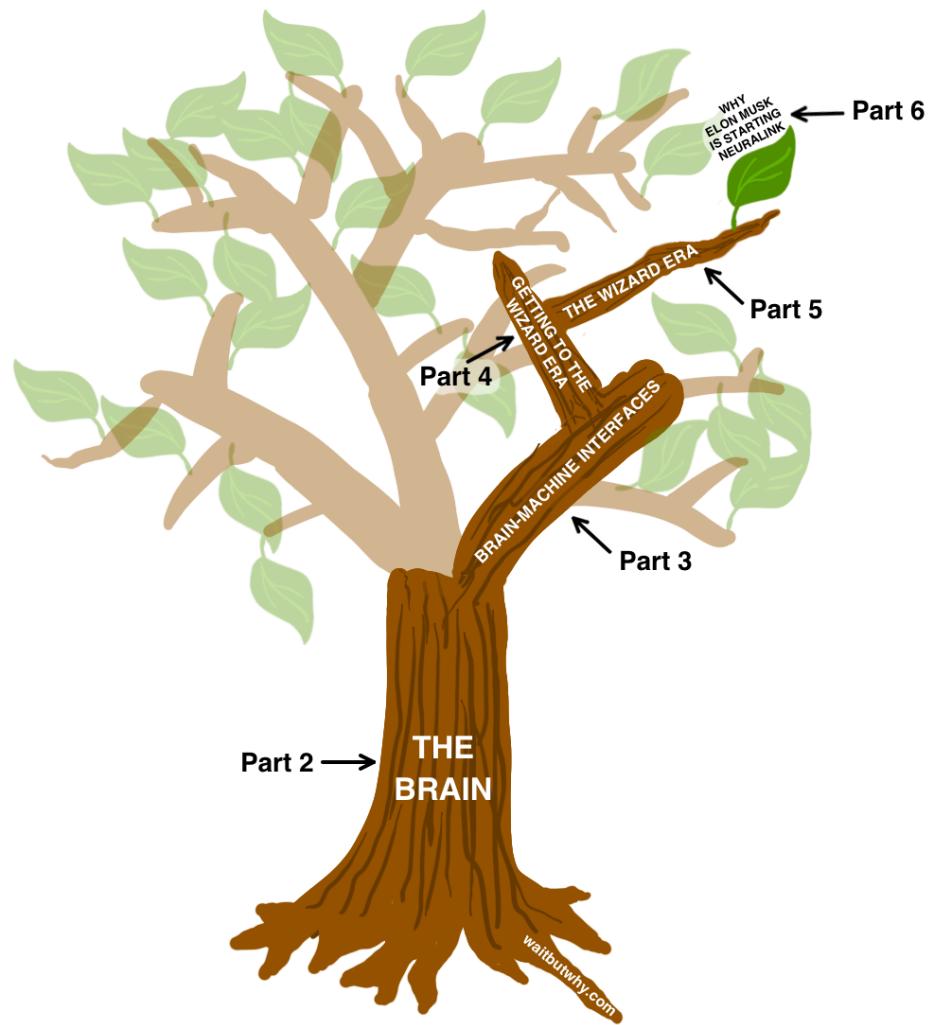
As we've discussed [before](#), knowledge works like a tree. If you try to learn a branch or a leaf of a topic before you have a solid tree trunk foundation of understanding in your head, it won't work. The branches and leaves will have nothing to stick to, so they'll fall right out of your head.

We've established that Elon Musk wants to build a wizard hat for the brain, and understanding why he wants to do that is the key to understanding Neuralink—and to understanding what our future might actually be like.

But none of that will make much sense until we really get into the truly mind-blowing concept of what a wizard hat is, what it might be like to wear one, and how we get there from where we are today.

The foundation for *that* discussion is an understanding of what brain-machine interfaces are, how they work, and where the technology is today.

Finally, BMIs themselves are just a larger branch—not the tree's trunk. In order to really understand BMIs and how they work, we need to understand the brain. Getting how the brain works is our tree trunk.

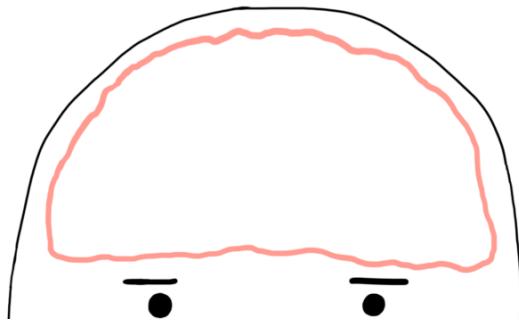


So we'll start with the brain, which will prepare us to learn about BMIs, which will teach us about what it'll take to build a wizard hat, and that'll set things up for an insane discussion about the future—which will get our heads right where they need to be to wrap themselves around why Elon thinks a wizard hat is such a critical piece of our future. And by the time we reach the end, this whole thing should click into place.

Part 2: The Brain



This post was a nice reminder of why I like working with a brain that looks nice and cute like this:

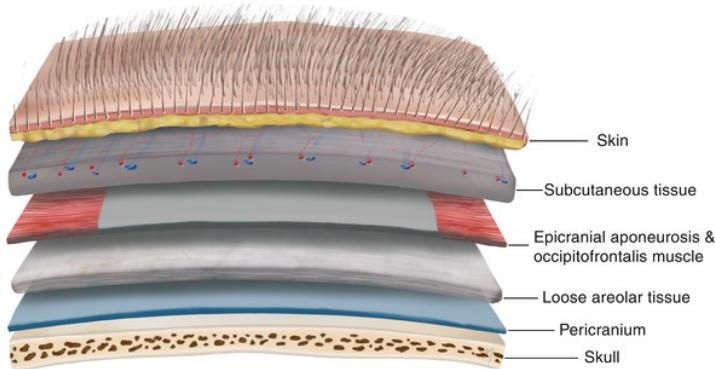


Because the real brain is extremely uncute and upsetting-looking. People are gross.

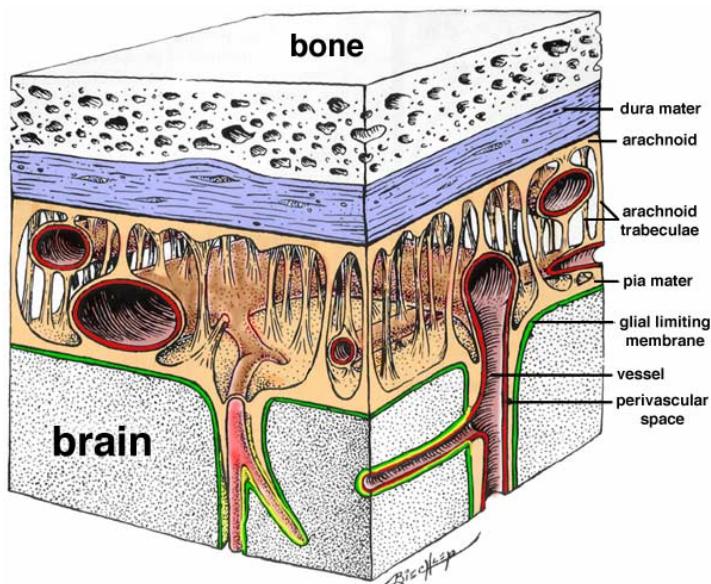
But I've been living in a shimmery, oozy, blood-vessel-lined Google Images hell for the past month, and now you have to deal with it too. So just settle in.

We'll start outside the head. One thing I will give to biology is that it's sometimes very satisfying,⁵ and the brain has some satisfying things going on. The first of which is that there's a real Russian doll situation going on with your head.

You have your hair, and under that is your scalp, and then you think your skull comes next—but it's actually like 19 things and then your skull:⁶



Then below your skull,⁶ another whole bunch of things are going on before you get to the brain⁴:



Your brain has three membranes around it underneath the skull:

On the outside, there's the **dura mater** (which means "hard mother" in Latin), a firm, rugged, waterproof layer. The dura is flush with the skull. I've heard it said that the brain has no pain sensory area, but the dura actually does—it's about as sensitive as the skin on your face—and pressure on or contusions in the dura often account for people's bad headaches.

Then below that there's the **arachnoid mater** ("spider mother"), which is a layer of skin and then an open space with these stretchy-looking fibers. I always thought my brain was just floating aimlessly in my head in some kind of fluid, but actually, the only real space gap between the outside of the brain and the inner wall of the skull is this arachnoid business. Those fibers stabilize the brain in position so it can't move too much, and they act as a shock absorber when your head bumps into something. This area is filled with spinal fluid, which keeps the brain mostly buoyant, since its density is similar to that of water.

Finally you have the **pia mater** ("soft mother"), a fine, delicate layer of skin that's fused with the outside of the brain. You know how when you see a brain, it's always covered with icky blood vessels? Those aren't actually on the brain's surface, they're embedded in the pia. (For the non-squeamish, [here's a video](#) of a professor peeling the pia off of a human brain.)

Here's the full overview, using the head of what looks like probably a pig:



From the left you have the skin (the pink), then two scalp layers, then the skull, then the dura, arachnoid, and on the far right, just the brain covered by the pia.

Once we've stripped everything down, we're left with this silly boy:⁵



This ridiculous-looking thing is the most complex known object in the universe—three pounds of what neuroengineer Tim Hanson calls “one of the most information-dense, structured, and self-structuring matter known.”⁵ All while operating on only **20 watts** of power (an **equivalently powerful computer** runs on 24,000,000 watts).

It’s also what MIT professor Polina Anikeeva **calls** “soft pudding you could scoop with a spoon.” Brain surgeon Ben Rapoport described it to me more scientifically, as “somewhere between pudding and jello.” He explained that if you placed a brain on a table, gravity would make it lose its shape and flatten out a bit, kind of like a jellyfish. We often don’t think of the brain as so smooshy, because it’s normally suspended in water.

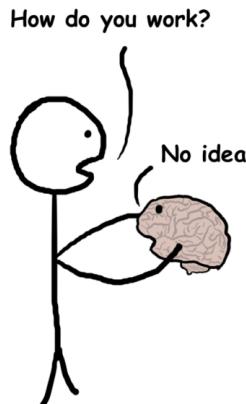
But this is what we all are. You look in the mirror and see your body and your face and you think that’s you—but that’s really just the machine you’re riding in. What you actually are is a zany-looking ball of jello. I hope that’s okay.

And given how weird that is, you can’t really blame Aristotle, or the ancient Egyptians, or many others, for assuming that the brain was somewhat-meaningless “cranial stuffing” (Aristotle believed the heart was the center of intelligence).⁷

Eventually, humans figured out the deal. But only kind of.

Professor Krishna Shenoy **likened** our understanding of the brain to humanity’s grasp on the world map **in the early 1500s**.

Another professor, **Jeff Lichtman**, is even harsher. He starts off his courses by asking his students the question, “If everything you need to know about the brain is a mile, how far have we walked in this mile?” He says students give answers like three-quarters of a mile, half a mile, a quarter of a mile, etc.—but that he believes the real answer is “about three inches.”⁸

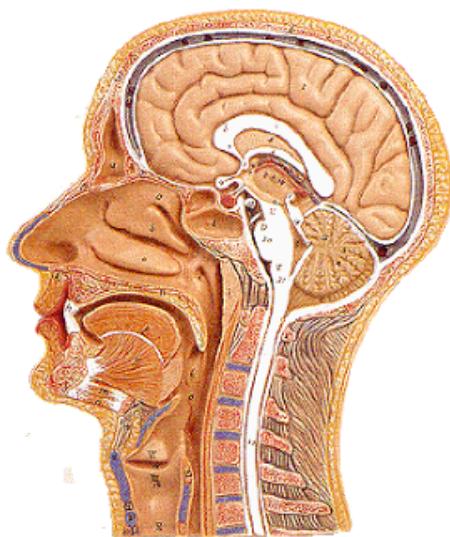


A third professor, neuroscientist **Moran Cerf**, shared with me an old neuroscience saying that points out why trying to master the brain is a bit of a catch-22: “If the human brain were so simple that we could understand it, we would be so simple that we couldn’t.”

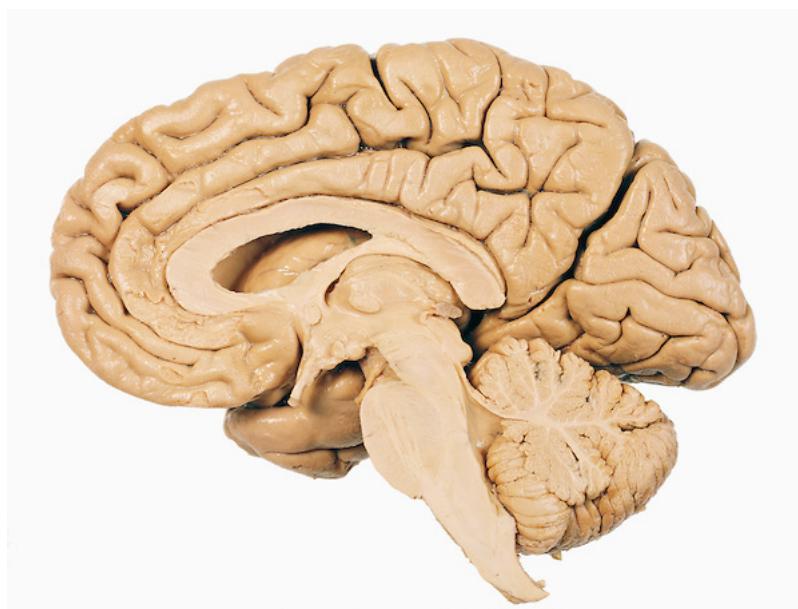
Maybe with the help of the great knowledge tower our species is building, we can get there at some point. For now, let’s go through what we do currently know about the jellyfish in our heads—starting with the big picture.

The brain, zoomed out

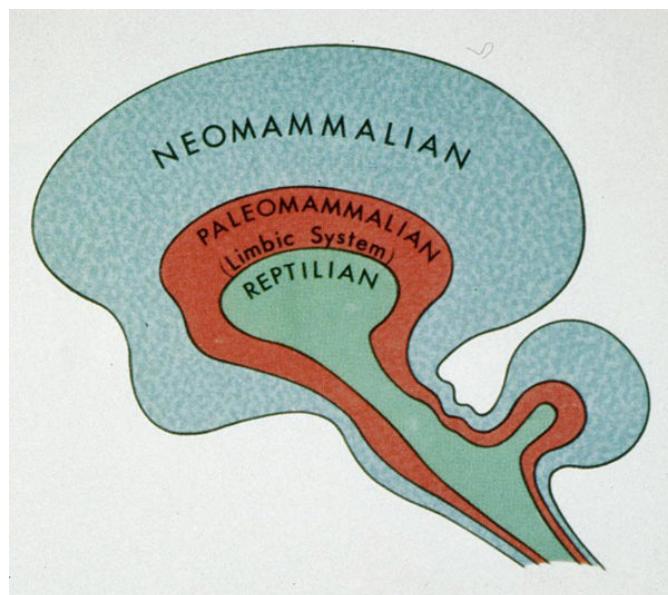
Let’s look at the major sections of the brain using a hemisphere cross section. So this is what the brain looks like in your head:



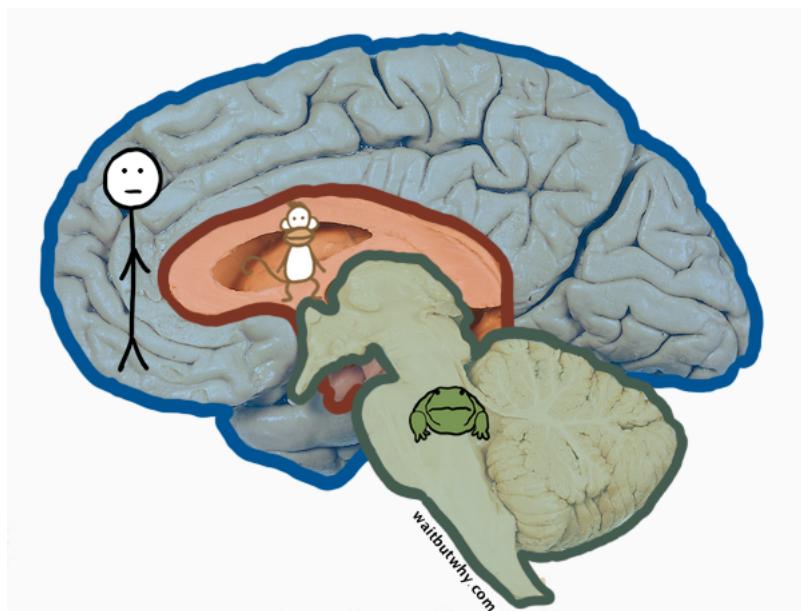
Now let's take the brain out of the head and remove the left hemisphere, which gives us a good view of the inside.⁹



Neurologist [Paul MacLean](#) made a simple diagram that illustrates the basic idea we talked about earlier of the reptile brain coming first in evolution, then being built upon by mammals, and finally being built upon again to give us our brain trifecta.



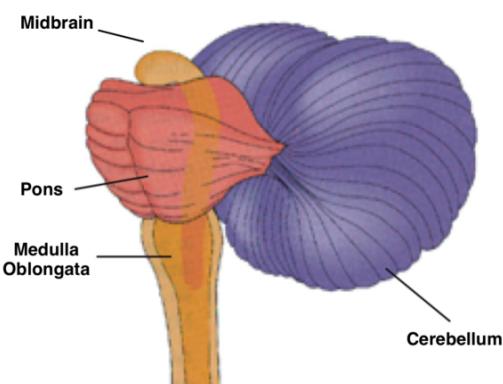
Here's how this essentially maps out on our real brain:



Let's take a look at each section:

The Reptilian Brain: The Brain Stem (and Cerebellum)

This is the most ancient part of our brain:^[10]



That's the section of our brain cross section above that the frog boss resides over. In fact, a frog's entire brain is similar to this lower part of our brain. Here's a real frog brain:^[11]



When you understand the function of these parts, the fact that they're ancient makes sense—everything these parts do, frogs and lizards can do. These are the major sections (click any of these spinning images to see a high-res version):

The medulla oblongata



The medulla oblongata really just wants you to not die. It does the thankless tasks of controlling involuntary things like your heart rate, breathing, and blood pressure, along with making you vomit when it thinks you've been poisoned.

The pons

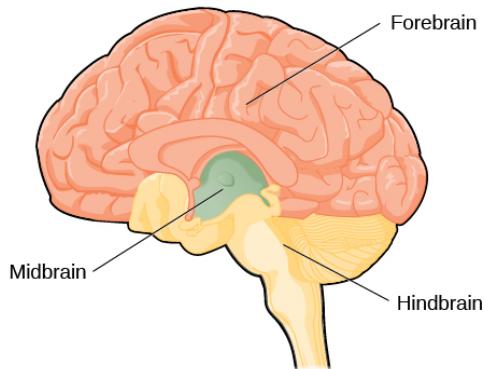


The pons's thing is that it does a little bit of this and a little bit of that. It deals with swallowing, bladder control, facial expressions, chewing, saliva, tears, and posture—really just whatever it's in the mood for.

The midbrain



The midbrain is dealing with an even bigger identity crisis than the pons. You know a brain part is going through some shit when almost all its functions are already another brain part's thing. In the case of the midbrain, it deals with vision, hearing, motor control, alertness, temperature control, and a bunch of other things that other people in the brain already do. The rest of the brain doesn't seem very into the midbrain either, given that they created a ridiculously uneven "forebrain, midbrain, hindbrain" divide that intentionally isolates the midbrain all by itself while everyone else hangs out.¹²



One thing I'll grant the pons and midbrain is that it's the two of them that control your voluntary eye movement, which is a pretty legit job. So if right now you move your eyes around, that's you doing something specifically with your pons and midbrain.⁷

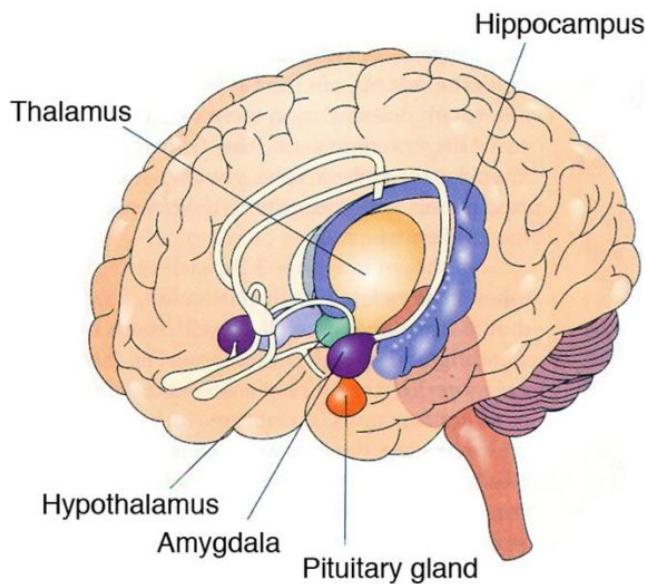
The cerebellum



The odd-looking thing that looks like your brain's scrotum is your cerebellum (Latin for "little brain"), which makes sure you stay a balanced, coordinated, and normal-moving person. [Here's that rad professor again](#) showing you what a real cerebellum looks like. ¹²

The Paleo-Mammalian Brain: The Limbic System

Above the brain stem is the limbic system—the part of the brain that makes humans so insane. ¹³



The limbic system is a survival system. A decent rule of thumb is that whenever you're doing something that your dog might also do—eating, drinking, having sex, fighting, hiding or running away from something scary—your limbic system is probably behind the wheel. Whether it feels like it or not, when you're doing any of those things, you're in primitive survival mode.

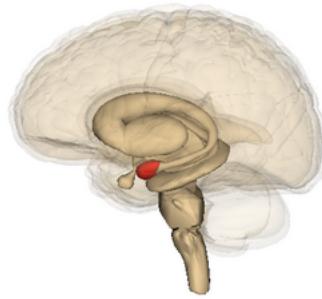
The limbic system is also where your emotions live, and in the end, emotions are also all about survival—they're the more advanced mechanisms of survival, necessary for animals living in a complex social structure.

In other posts, when I refer to your [Instant Gratification Monkey](#), your [Social Survival Mammoth](#), and [all your other animals](#)—I'm usually referring to your limbic system. Anytime there's an internal battle going on in your head, it's likely that the limbic system's role is urging you to do the thing you'll later regret doing.

I'm pretty sure that gaining control over your limbic system is both the definition of maturity and the core human struggle. It's not that we would be better off without our limbic systems—limbic systems are half of what makes us distinctly human, and most of the fun of life is related to emotions and/or fulfilling your animal needs—it's just that your limbic system doesn't get that you live in a civilization, and if you let it run your life too much, it'll quickly ruin your life.

Anyway, let's take a closer look at it. There are a lot of little parts of the limbic system, but we'll keep it to the biggest celebrities:

The amygdala



The amygdala is kind of an emotional wreck of a brain structure. It deals with anxiety, sadness, and our responses to fear. There are two amygdalae, and oddly, the left one has been shown to be more balanced, sometimes producing happy feelings in addition to the usual angsty ones, while the right one is always in a bad mood.

The hippocampus



Your hippocampus (Greek for “seahorse” because it [looks like one](#)) is like a scratch board for memory. When rats start to memorize directions in a maze, the memory gets encoded in their hippocampus—quite literally. Different parts of the rat’s two hippocampi will fire during different parts of the maze, since each section of the maze is stored in its own section of the hippocampus. But if after learning one maze, the rat is given other tasks and is brought back to the original maze a year later, it will have a hard time remembering it, because the hippocampus scratch board has been mostly wiped of the memory so as to free itself up for new memories.

The condition in the movie *Memento* is a real thing—anteriorgrade amnesia—and it’s caused by damage to the hippocampus. Alzheimer’s also starts in the hippocampus before working its way through many parts of the brain, which is why, of the slew of devastating effects of the disease, diminished memory happens first.

The thalamus



In its central position in the brain, the thalamus also serves as a sensory middleman that receives information from your sensory organs and sends them to your cortex for processing. When you’re sleeping, the thalamus goes to sleep with you, which means the sensory middleman is off duty. That’s why in a deep sleep, some sound or light or touch often will not wake you up. If you want to wake someone up who’s in a deep sleep, you have to be aggressive enough to wake their thalamus up.

The exception is your sense of smell, which is the one sense that bypasses the thalamus. That’s why smelling salts are used to wake up a passed-out person. While we’re here, cool fact: smell is the function of the [olfactory bulb](#) and is the most ancient of the senses. Unlike the other senses, smell is located deep in the limbic system, where it works closely with the hippocampus and amygdala—which is why smell is so closely tied to memory and emotion.

The Neo-Mammalian Brain: The Cortex

Finally, we arrive at the cortex. The cerebral cortex. The neocortex. The cerebrum. The pallium.

The most important part of the whole brain can't figure out what its name is. Here's what's happening:

The What the Hell is it Actually Called Blue Box

The **cerebrum** is the whole big top/outside part of the brain but it also technically includes some of the internal parts too.

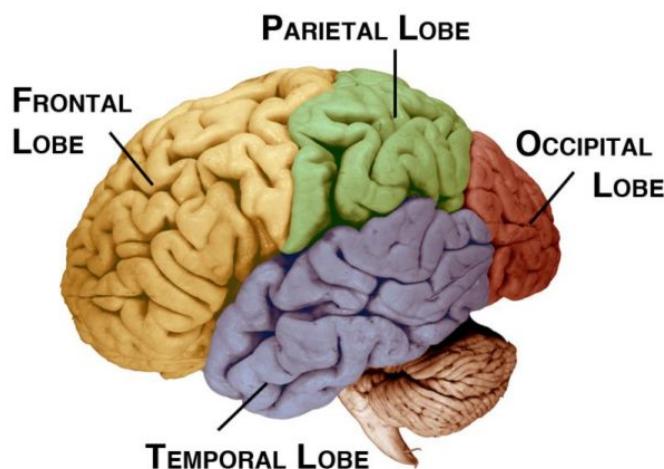
Cortex means "bark" in Latin and is the word used for the outer layer of many organs, not just the brain. The outside of the cerebellum is the cerebellar cortex. And the outside of the cerebrum is the **cerebral cortex**. Only mammals have cerebral cortices. The equivalent part of the brain in reptiles is called the **pallium**.

The **neocortex** is often used interchangeably with "cerebral cortex," but it's technically the outer layers of the cerebral cortex that are especially developed in more advanced mammals. The other parts are called the allocortex.

In the rest of this post, we'll be mostly referring to the neocortex **but we'll just call it the cortex**, since that's the least annoying way to do it for everyone.

The cortex is in charge of basically everything—processing what you see, hear, and feel, along with language, movement, thinking, planning, and personality.

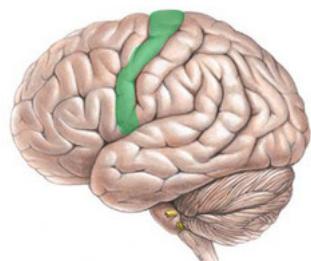
It's divided into four lobes:¹⁴



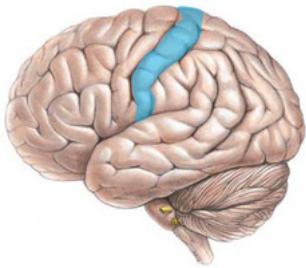
It's pretty unsatisfying to describe what they each do, because they each do so many things and there's a lot of overlap, but to oversimplify:

The **frontal lobe** (click the words to see a gif) handles your personality, along with a lot of what we think of as "thinking"—reasoning, planning, and executive function. In particular, a lot of your thinking takes place in the front part of the frontal lobe, called the **prefrontal cortex**—the adult in your head. The prefrontal cortex is the other character in those internal battles that go on in your life. The rational decision-maker trying to get you to do your work. The authentic voice trying to get you to stop worrying so much what others think and just be yourself. The higher being who wishes you'd stop sweating the small stuff.

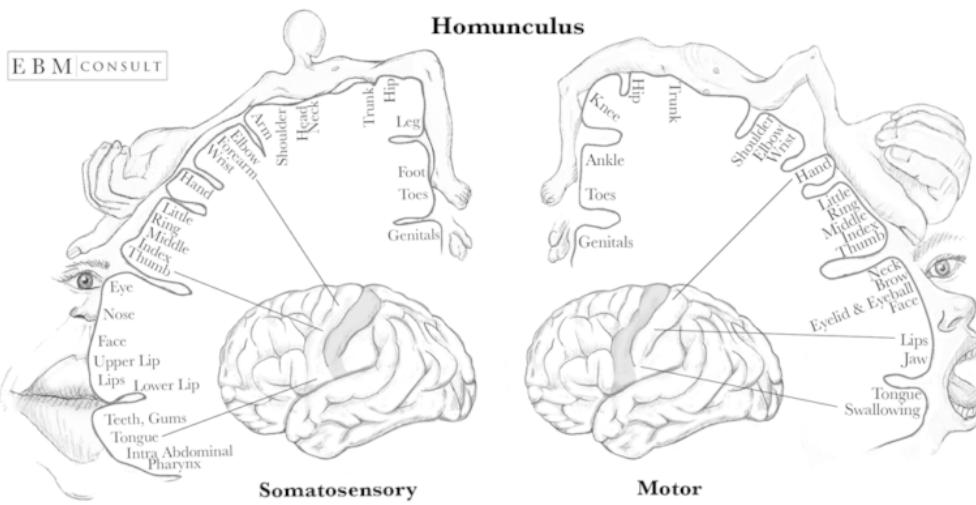
As if that's not enough to worry about, the frontal lobe is also in charge of your body's movement. The top strip of the frontal lobe is your **primary motor cortex**.¹⁵



Then there's the **parietal lobe** which, among other things, controls your sense of touch, particularly in the **primary somatosensory cortex**, the strip right next to the primary motor cortex.¹⁶



The motor and somatosensory cortices are fun because they're well-mapped. Neuroscientists know exactly which part of each strip connects to each part of your body. Which leads us to the creepiest diagram of this post: the *homunculus*.

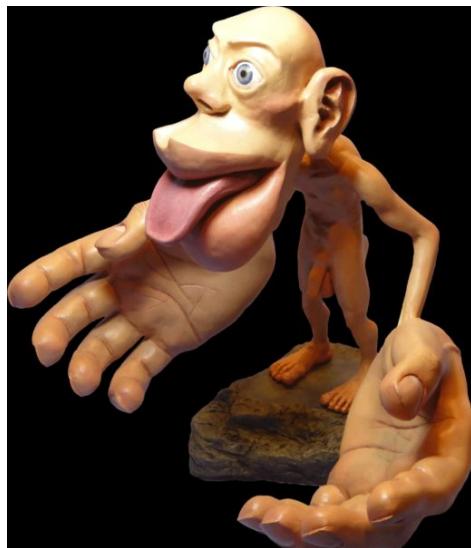


The homunculus, created by pioneer neurosurgeon Wilder Penfield, visually displays how the motor and somatosensory cortices are mapped. The larger the body part in the diagram, the more of the cortex is dedicated to its movement or sense of touch. A couple interesting things about this:

First, it's amazing that more of your brain is dedicated to the movement and feeling of your face and hands than to the rest of your body combined. This makes sense though—you need to make incredibly nuanced facial expressions and your hands need to be unbelievably dexterous, while the rest of your body—your shoulder, your knee, your back—can move and feel things much more crudely. This is why people can play the piano with their fingers but not with their toes.

Second, it's interesting how the two cortices are basically dedicated to the same body parts, in the same proportions. I never really thought about the fact that the same parts of your body you need to have a lot of movement control over tend to also be the most sensitive to touch.

Finally, I came across this shit and I've been living with it ever since—so now you have to too. A 3-dimensional homunculus man.¹⁷



Moving on—

The **temporal lobe** is where a lot of your memory lives, and being right next to your ears, it's also the home of your **auditory cortex**.

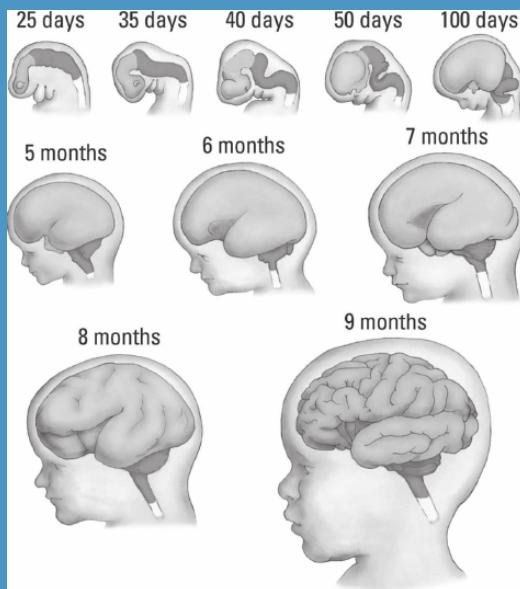
Last, at the back of your head is the **occipital lobe**, which houses your **visual cortex** and is almost entirely dedicated to vision.

Now for a long time, I thought these major lobes were *chunks* of the brain—like, segments of the whole 3D structure. But actually, the cortex is just the outer *two millimeters* of the brain—the thickness of a nickel—and the meat of the space underneath is mostly just wiring.

The Why Brains Are So Wrinkly Blue Box

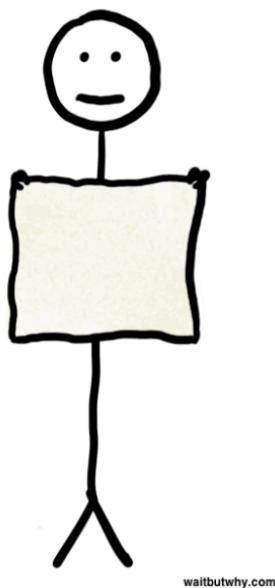
As we've discussed, the evolution of our brain happened by building outwards, adding newer, fancier features on top of the existing model. But building outwards has its limits, because the need for humans to emerge into the world through someone's vagina puts a cap on how big our heads could be.⁹

So evolution got innovative. Because the cortex is so thin, it scales by increasing its surface area. That means that by creating lots of folds (including both sides **folding down into the gap between the two hemispheres**), you can more than triple the area of the brain's surface without increasing the volume too much. When the brain first develops in the womb, the cortex is smooth—the folds form mostly in the last two months of pregnancy:¹⁸



Cool explainer of how the folds form here.

If you could take the cortex off the brain, you'd end up with a 2mm-thick sheet with an area of 2,000-2,400cm²—about the size of a 48cm x 48cm (19in x 19in) square. ¹⁰ A dinner napkin.

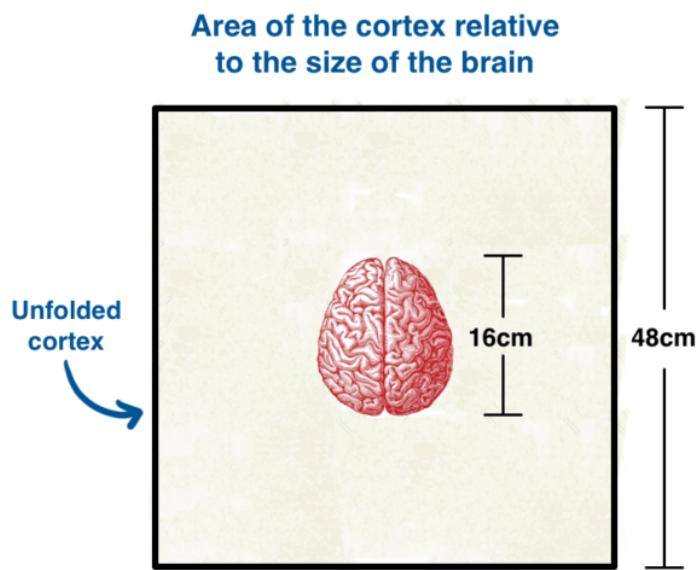


waitbutwhy.com

This napkin is where most of the action in your brain happens—it's why you can think, move, feel, see, hear, remember, and speak and understand language. Best napkin ever.

And remember before when I said that you were a jello ball? Well the *you* you think of when you think of yourself—it's really mainly your cortex. Which means you're actually a napkin.

The magic of the folds in increasing the napkin's size is clear when we put another brain on top of our stripped-off cortex:



So while it's not perfect, modern science has a decent understanding of the big picture when it comes to the brain. We also have a decent understanding of the little picture. Let's check it out:

The brain, zoomed in

Even though we figured out that the brain was the seat of our intelligence a long time ago, it wasn't until pretty recently that science understood what the brain was made of. Scientists knew that the body was made of cells, but in the late 19th century, Italian physician Camillo Golgi figured out how to use a staining method to see what brain cells actually looked like. The result was surprising:



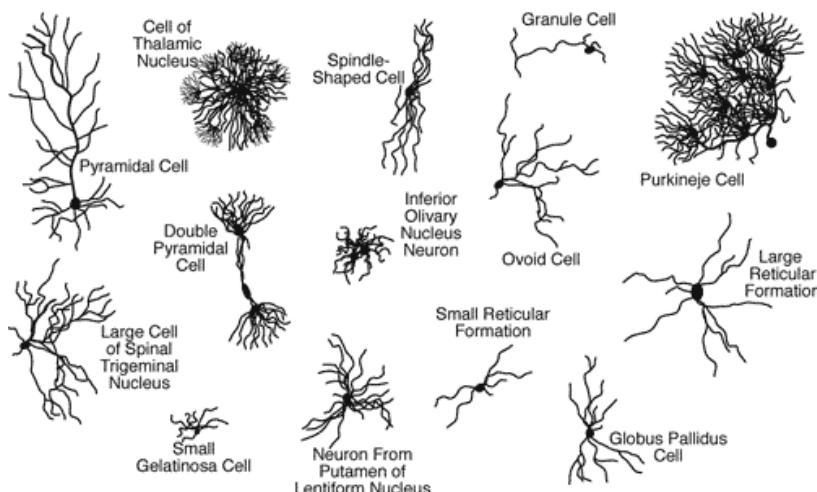
That wasn't what a cell was supposed to look like. Without quite realizing it yet,¹¹ Golgi had discovered the neuron.

Scientists realized that the neuron was the core unit in the vast communication network that makes up the brains and nervous systems of nearly all animals.

But it wasn't until the 1950s that scientists worked out *how* neurons communicate with each other.

An axon, the long strand of a neuron that carries information, is normally microscopic in diameter—too small for scientists to test on until recently. But in the 1930s, English zoologist J. Z. Young discovered that the squid, randomly, could change everything for our understanding, because squids have an [unusually huge axon](#) in their bodies that could be experimented on. A couple decades later, using the squid's giant axon, scientists Alan Hodgkin and Andrew Huxley definitively figured out how neurons send information: the action potential. Here's how it works.

So there are a lot of different kinds of neurons—¹²



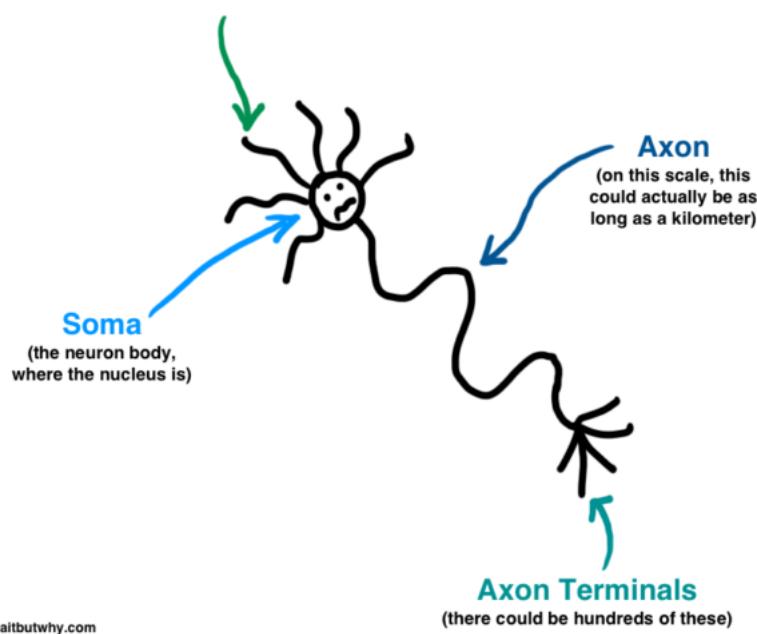
—but for simplicity, we'll discuss the cliché textbook neuron—a pyramidal cell, like one you might find in your motor cortex. To make a neuron diagram, we can start with a guy:



And then if we just give him a few extra legs, some hair, take his arms off, and stretch him out—we have a neuron.

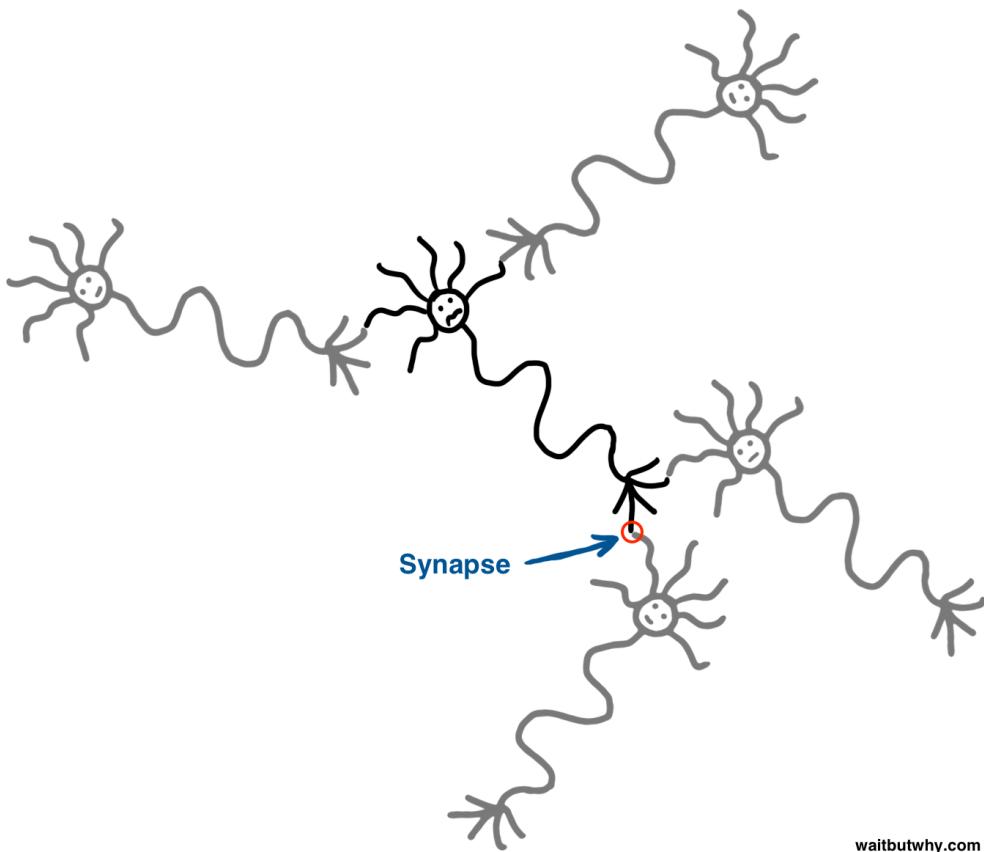
Dendrites

(there could be hundreds of these, and on this scale, they may reach a few meters long)



waitbutwhy.com

And let's add in a few more neurons.

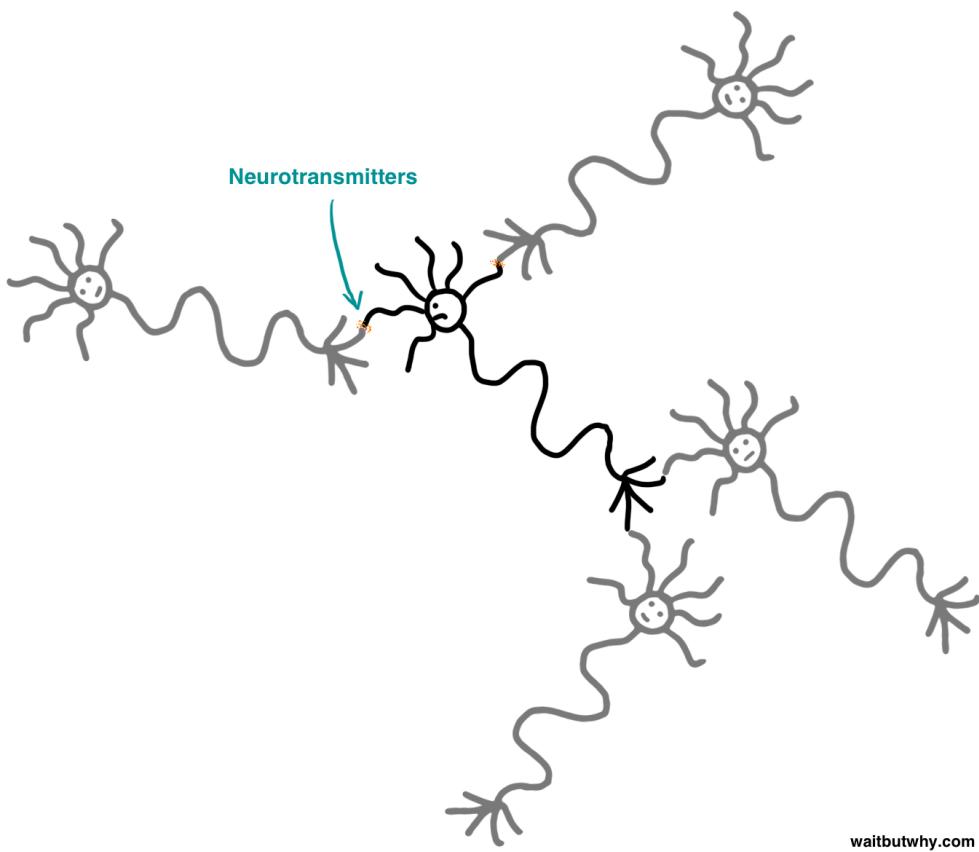


waitbutwhy.com

Rather than launch into the full, detailed explanation for how action potentials work—which involves a lot of unnecessary and uninteresting technical information you already dealt with in 9th-grade biology—I'll link to this great Khan Academy [explainer article](#) for those who want the full story. We'll go through the very basic ideas that are relevant for our purposes.

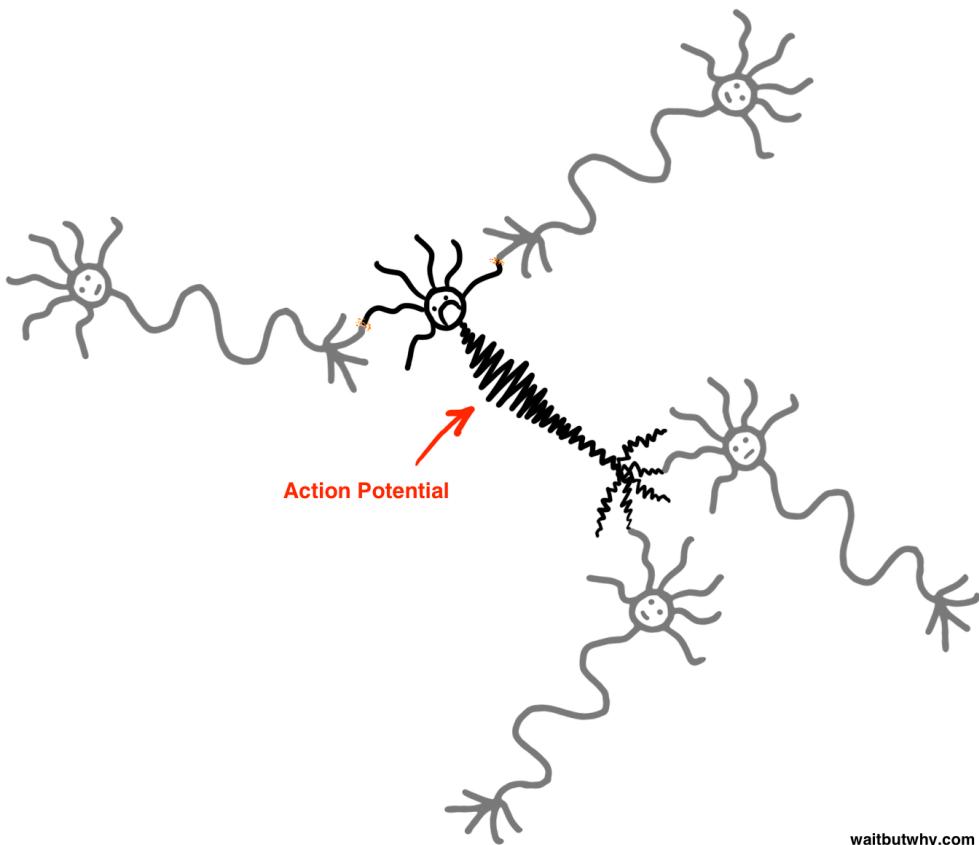
So our guy's body stem—the neuron's **axon**—has a negative “resting potential,” which means that when it's at rest, its electrical charge is slightly negative. At all times, a bunch of people's feet keep touching ¹² our guy's hair—the neuron's **dendrites**—whether he likes it or not. Their feet drop chemicals called **neurotransmitters** ¹³ onto his hair—which pass through his head (the cell body, or

soma) and, depending on the chemical, raise or lower the charge in his body a little bit. It's a little unpleasant for our neuron guy, but not a huge deal—and nothing else happens.



waitbutwhy.com

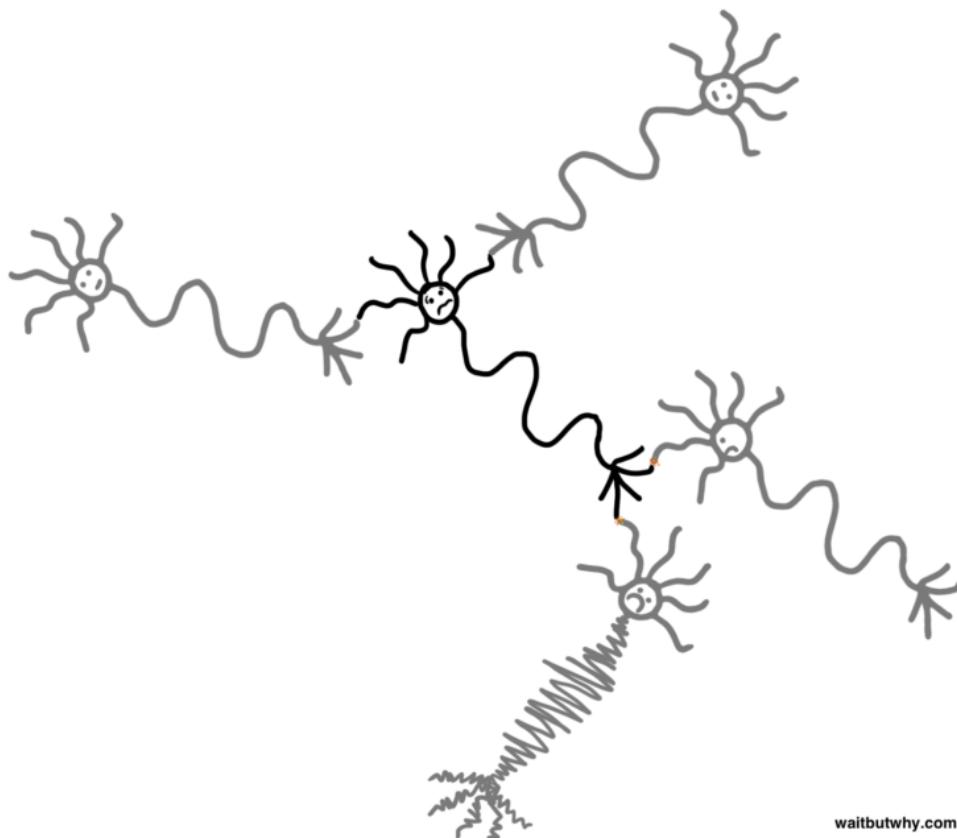
But if enough chemicals touch his hair to raise his charge over a certain point—the neuron's "threshold potential"—then it triggers an **action potential**, and our guy is electrocuted.



waitbutwhy.com

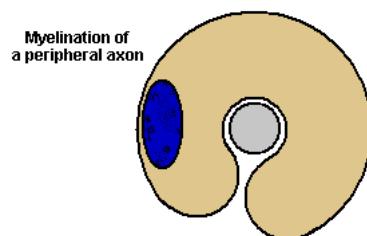
This is a binary situation—either *nothing* happens to our guy, or he's fully electrocuted. He can't be kind of electrocuted, or extra electrocuted—he's either not electrocuted at all, or he's fully electrocuted to the exact same degree every time.

When this happens, a pulse of electricity (in the form of a brief reversal of his body's normal charge from negative to positive and then rapidly back down to his normal negative) zips down his body (the axon) and into his feet—the neuron's **axon terminals**—which themselves touch a bunch of other people's hair (the points of contact are called **synapses**). When the action potential reaches his feet, it causes them to release chemicals onto the people's hair they're touching, which may or may not cause those people to be electrocuted, just like he was.



This is usually how info moves through the nervous system—chemical information sent in the tiny gap between neurons triggers electrical information to pass *through* the neuron—but sometimes, in situations when the body needs to move a signal extra quickly, neuron-to-neuron connections can themselves be electric.

Action potentials move at between 1 and 100 meters/second. Part of the reason for this large range is that another type of cell in the nervous system—a Schwann cell—acts like a super nurturing grandmother and constantly wraps some types of axons in layers of fat blankets called **myelin sheath**. Like this (takes a second to start):²⁰



On top of its protection and insulation benefits, the myelin sheath is a major factor in the pace of communication—action potentials travel *much* faster through axons when they're covered in myelin sheath.^{14 21}



One nice example of the speed difference created by myelin: You know how when you stub your toe, your body gives you that one second of reflection time to think about what you just did and what you're about to feel, before the pain actually kicks in? What's happening is you feel both the sensation of your toe hitting against something and the sharp part of the pain right away, because sharp pain information is sent to the brain via types of axons that are myelinated. It takes a second or two for the dull pain to kick in because dull pain is sent via *unmyelinated "C fibers"*—at only around one meter/second.

Neural Networks

Neurons are similar to computer transistors in one way—they also transmit information in the binary language of 1's (action potential firing) and 0's (no action potential firing). But unlike computer transistors, the brain's neurons are constantly changing.

You know how sometimes you learn a new skill and you get pretty good at it, and then the next day you try again and you suck again? That's because what made you get good at the skill the day before was adjustments to the amount or concentration of the *chemicals* in the signaling between neurons. Repetition caused chemicals to adjust, which helped you improve, but the next day the chemicals were back to normal so the improvement went away.

But then if you *keep* practicing, you eventually get good at something in a lasting way. What's happened is you've told the brain, "this isn't just something I need in a one-off way," and the brain's neural network has responded by making *structural changes* to itself that last. Neurons have shifted shape and location and strengthened or weakened various connections in a way that has built a hard-wired set of pathways that know how to do that skill.

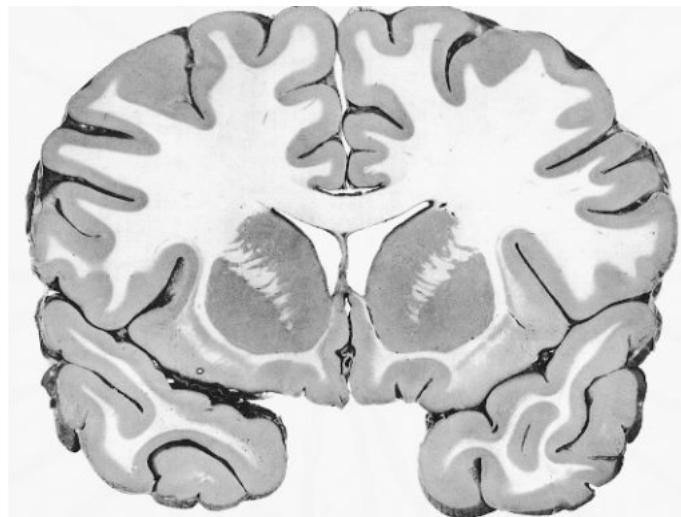
Neurons' ability to alter themselves chemically, structurally, and even functionally, allow your brain's neural network to optimize itself to the external world—a phenomenon called *neuroplasticity*. Babies' brains are the most neuroplastic of all. When a baby is born, its brain has no idea if it needs to accommodate the life of a medieval warrior who will need to become incredibly adept at sword-fighting, a 17th-century musician who will need to develop fine-tuned muscle memory for playing the harpsichord, or a modern-day intellectual who will need to store and organize a tremendous amount of information and master a complex social fabric—but the baby's brain is ready to shape itself to handle whatever life has in store for it.

Babies are the neuroplasticity superstars, but neuroplasticity remains throughout our whole lives, which is why humans can grow and change and learn new things. And it's why we can form new habits and break old ones—your habits are reflective of the existing circuitry in your brain. If you want to change your habits, you need to exert a lot of willpower to override your brain's neural pathways, but if you can keep it going long enough, your brain will eventually get the hint and alter those pathways, and the new behavior will stop requiring willpower. Your brain will have physically built the changes into a new habit.

Altogether, there are around 100 billion neurons in the brain that make up this unthinkably vast network—similar to the number of stars in the Milky Way and over 10 times the number of people in

the world. Around 15 – 20 billion of those neurons are in the cortex, and the rest are in the animal parts of your brain (surprisingly, the random cerebellum has more than three times as many neurons as the cortex).

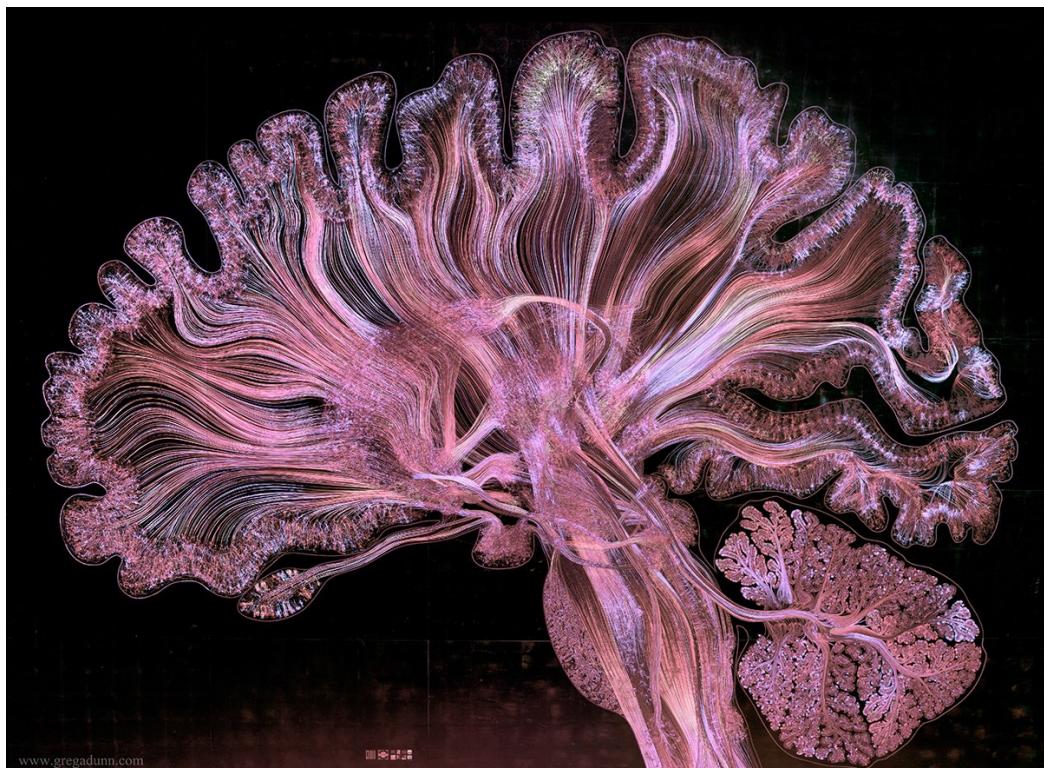
Let's zoom back out and look at another cross section of the brain—this time cut not from front to back to show a single hemisphere, but from side to side:²¹

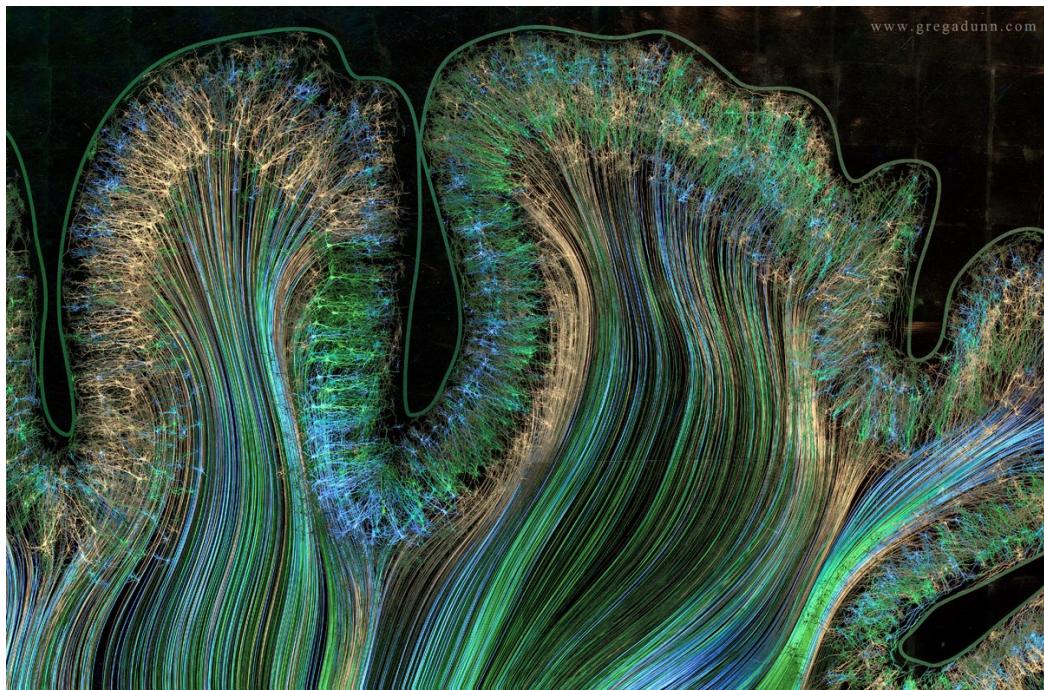


Brain material can be divided into what's called **gray matter** and **white matter**. Gray matter actually looks darker in color and is made up of the cell bodies (somas) of the brain's neurons and their thicket of dendrites and axons—along with a lot of other stuff. White matter is made up primarily of wiring—axons carrying information from somas to other somas or to destinations in the body. White matter is white because those axons are usually wrapped in myelin sheath, which is fatty white tissue.

There are two main regions of gray matter in the brain—the internal cluster of limbic system and brain stem parts we discussed above, and the nickel-thick layer of cortex around the outside. The big chunk of white matter in between is made up mostly of the axons of cortical neurons. The cortex is like a great command center, and it beams many of its orders out through the mass of axons making up the white matter beneath it.

The coolest illustration of this concept that I've come across¹⁵ is a [beautiful set of artistic representations](#) done by Dr. Greg A. Dunn and Dr. Brian Edwards. Check out the distinct difference between the structure of the outer layer of gray matter cortex and the white matter underneath it (click to view in high res):



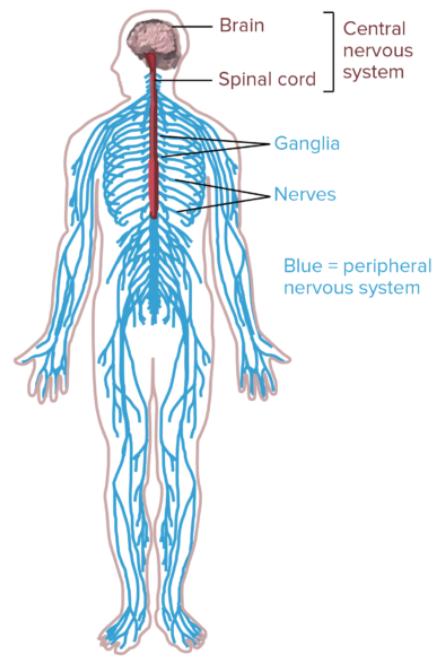


Those cortical axons might be taking information to another part of the cortex, to the lower part of the brain, or through the spinal cord—the nervous system’s superhighway—and into the rest of the body.

[16](#)

Let's look at the whole nervous system:

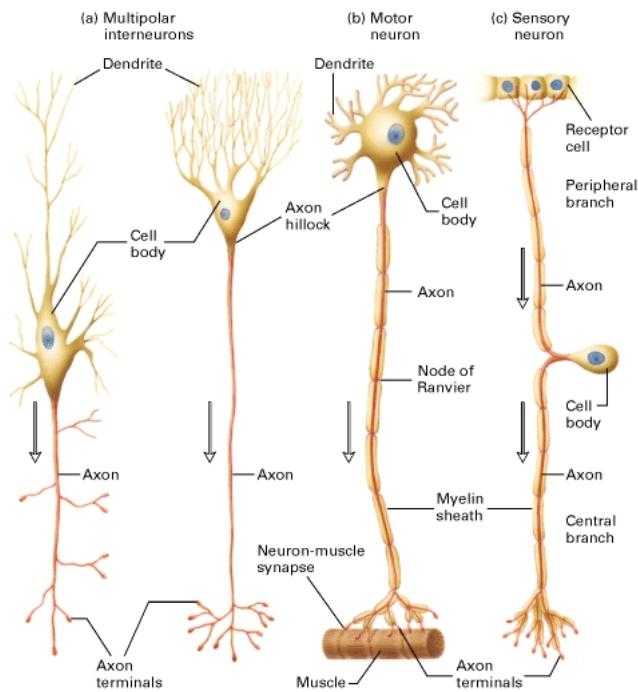
[23](#)



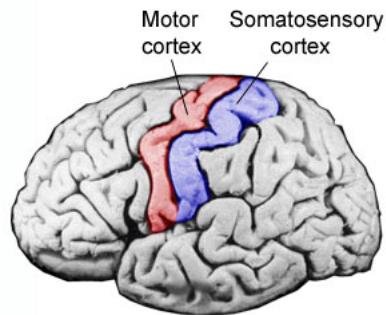
The nervous system is divided into two parts: the **central nervous system**—your brain and spinal cord—and the **peripheral nervous system**—made up of the neurons that radiate outwards from the spinal cord into the rest of the body.

Most types of neurons are **interneurons**—neurons that communicate with other neurons. When you think, it's a bunch of interneurons talking to each other. Interneurons are mostly contained to the brain.

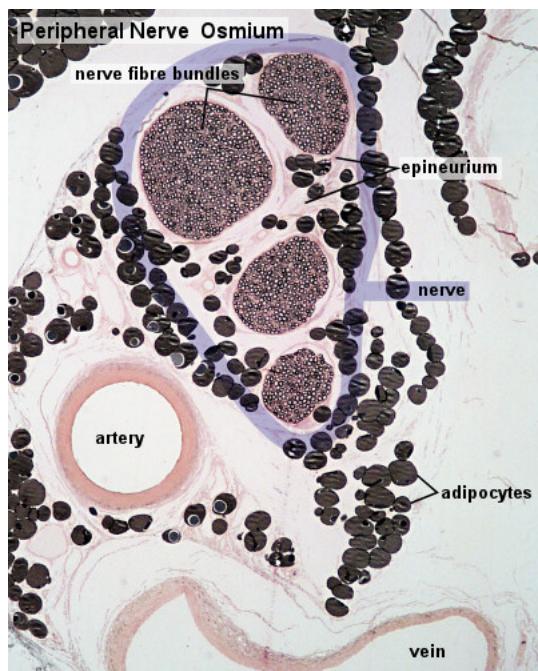
The two other kinds of neurons are **sensory neurons** and **motor neurons**—those are the neurons that head down into your spinal cord and make up the peripheral nervous system. These neurons can be up to a *meter* long.¹⁷ Here's a typical structure of each type:²⁴



Remember our two strips? ²⁵



These strips are where your peripheral nervous system originates. The axons of sensory neurons head down from the somatosensory cortex, through the brain's white matter, and into the spinal cord (which is just a massive bundle of axons). From the spinal cord, they head out to all parts of your body. Each part of your skin is lined with nerves that originate in the somatosensory cortex. A **nerve**, by the way, is a few bundles of axons wrapped together in a little cord. Here's a nerve up close: ²⁶



The nerve is the whole thing circled in purple, and those four big circles inside are bundles of many axons (here's a helpful [cartoony drawing](#)).

So if a fly lands on your arm, here's what happens:

The fly touches your skin and stimulates a bunch of sensory nerves. The axon terminals in the nerves have a little fit and start action potential-ing, sending the signal up to the brain to tell on the fly. The signals head into the spinal cord and up to the somas in the somatosensory cortex.¹⁸ The somatosensory cortex then taps the motor cortex on the shoulder and tells it that there's a fly on your arm and that it needs to deal with it (lazy). The particular somas in your motor cortex that connect to the muscles in your arm then start action potential-ing, sending the signals back into the spinal cord and then out to the muscles of the arm. The axon terminals at the end of those neurons stimulate your arm muscles, which constrict to shake your arm to get the fly off (by now the fly has already [thrown up](#) on your arm), and the fly (whose nervous system now goes through its own whole thing) flies off.

Then your amygdala looks over and realizes there was a bug on you, and it tells your motor cortex to jump embarrassingly, and if it's a spider instead of a fly, it also tells your vocal cords to yell out involuntarily and ruin your reputation.

So it seems so far like we do kind of actually understand the brain, right? But then why did that professor ask that question—*If everything you need to know about the brain is a mile, how far have we walked in this mile?*—and say the answer was *three inches*?

Well here's the thing.

You know how we totally get how an individual computer sends an email and we totally understand the broad concepts of the internet, like how many people are on it and what the biggest sites are and what the major trends are—but all the stuff in the middle—the inner workings of the internet—are pretty confusing?

And you know how economists can tell you all about how an individual consumer functions and they can also tell you about the major concepts of macroeconomics and the overarching forces at play—but no one can really tell you all the ins and outs of how the economy works or predict what will happen with the economy next month or next year?

The brain is kind of like those things. We get the little picture—we know all about how a neuron fires. And we get the big picture—we know how many neurons are in the brain and what the major lobes and structures control and how much energy the whole system uses. But the stuff in between—all that *middle* stuff about how each part of the brain actually does its thing?

Yeah we don't get that.

What really makes it clear how confounded we are is hearing a neuroscientist talk about the parts of the brain we understand *best*.

Like the visual cortex. We understand the visual cortex pretty well because it's easy to *map*.

Research scientist Paul Merolla described it to me:

The visual cortex has very nice anatomical function and structure. When you look at it, you literally see a map of the world. So when something in your visual field is in a certain region of space, you'll see a little patch in the cortex that represents that region of space, and it'll light up. And as that thing moves over, there's a topographic mapping where the neighboring cells will represent that. It's almost like having Cartesian coordinates of the real world that will map to polar coordinates in the visual cortex. And you can literally trace from your retina, through your thalamus, to your visual cortex, and you'll see an actual mapping from this point in space to this point in the visual cortex.

So far so good. But then he went on:

So that mapping is really useful if you want to interact with certain parts of the visual cortex, but there's many regions of vision, and as you get deeper into the visual cortex, it becomes a little bit more nebulous, and this topographic representation starts to break down. ... There's all these levels of things going on in the brain, and visual perception is a great example of that. We look at the world, and there's just this physical 3D world out there—like you look at a cup, and you just see a cup—but what your eyes are seeing is really just a bunch of pixels. And when you look in the visual cortex, you see that there are roughly 20-40 different maps. V1 is the first area, where it's tracking little edges and colors and things like that. And there's other areas looking at more complicated objects, and there's all these different visual representations on the surface of your brain, that you can see. And somehow all of that information is being bound together in this information stream that's being coded in a way that makes you believe you're just seeing a simple object.

And the motor cortex, another one of the best-understood areas of the brain, might be even more difficult to understand on a granular level than the visual cortex. Because even though we know which *general* areas of the motor cortex map to which areas of the body, the individual neurons in these motor cortex areas *aren't* topographically set up, and the specific way they work together to create movement in the body is anything but clear. Here's Paul again:

The neural chatter in everyone's arm movement part of the brain is a little bit different—it's not like the neurons speak English and say "move"—it's a pattern of electrical activity, and in everyone it's a little bit different. ... And you want to be able to seamlessly understand that it means "Move the arm this way" or "move the arm toward the target" or "move the arm to the left, move it up, grasp, grasp with a certain kind of force, reach with a certain speed," and so on. We don't think about these things when we move—it just happens seamlessly. So each brain has a unique code with which it talks to the muscles in the arm and hand.

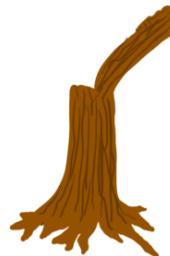
The *neuroplasticity* that makes our brains so useful to us also makes them incredibly difficult to understand—because the way each of our brains works is based on how that brain has shaped itself, based on its particular environment and life experience.

And again, those are the areas of the brain we understand the *best*. "When it comes to more sophisticated computation, like language, memory, mathematics," one expert told me, "we really don't understand how the brain works." He lamented that, for example, the concept of one's mother is coded in a different way, and in different parts of the brain, for every person. And in the frontal lobe—you know, that part of the brain where *you* really live—"there's no topography at all."

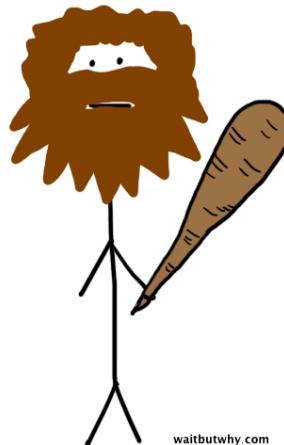
But somehow, *none of this* is why building effective brain-computer interfaces is so hard, or so daunting. What makes BMIs so hard is that the engineering challenges are *monumental*. It's *physically* working with the brain that makes BMIs among the hardest engineering endeavors in the world.

So with our brain background tree trunk built, we're ready to head up to our first branch.

Part 3: Brain-Machine Interfaces

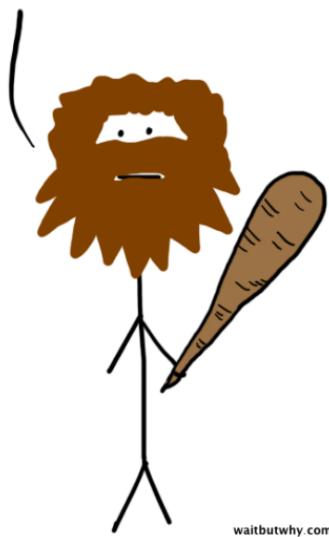


Let's zip back in time for a second to 50,000 BC and kidnap someone and bring him back here to 2017.



This is Bok. Bok, we're really thankful that you and your people invented language.

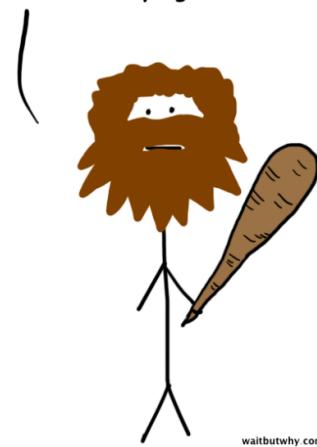
No prob.



waitbutwhy.com

As a way to thank you, we want to show you all the amazing things we were able to build because of your invention.

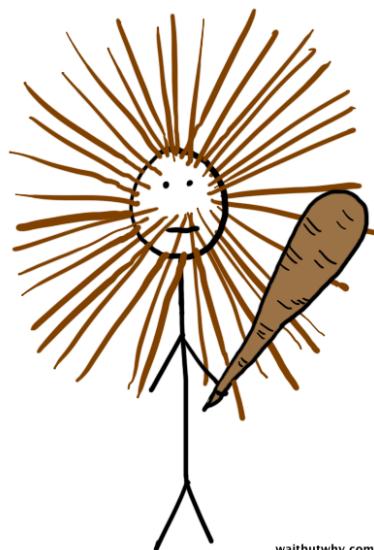
K let's see whatchya got.



waitbutwhy.com

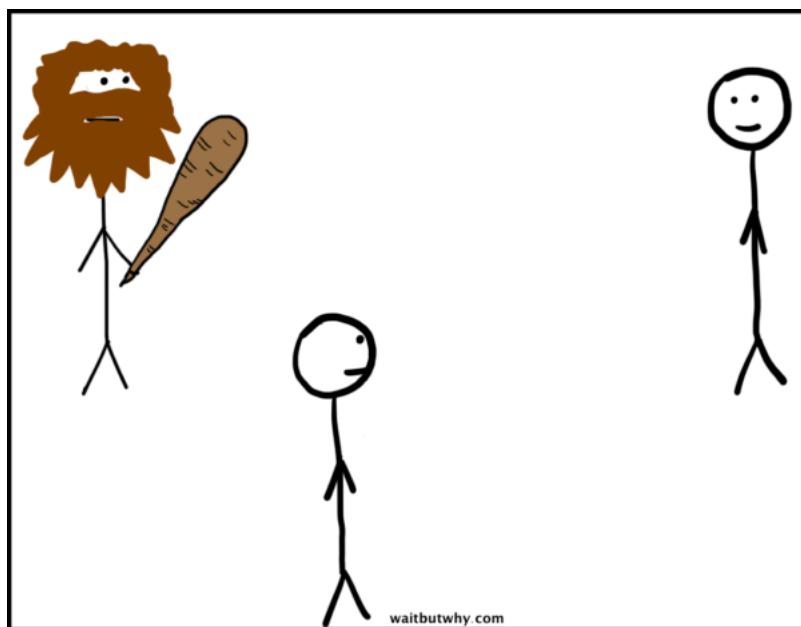
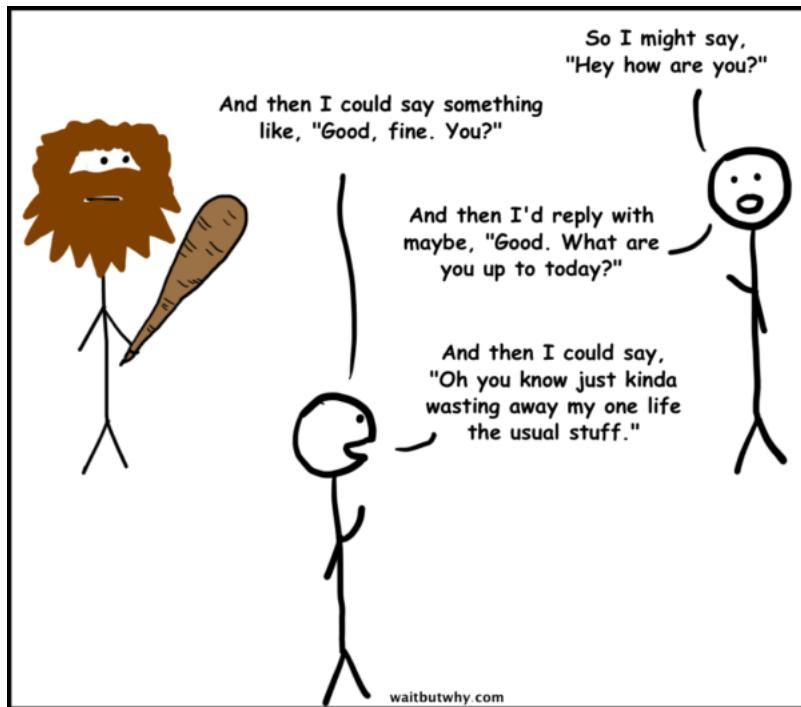
Alright, first let's take Bok on a plane, and into a submarine, and to the top of the Burj Khalifa. Now we'll show him a telescope and a TV and an iPhone. And now we'll let him play around on the internet for a while.

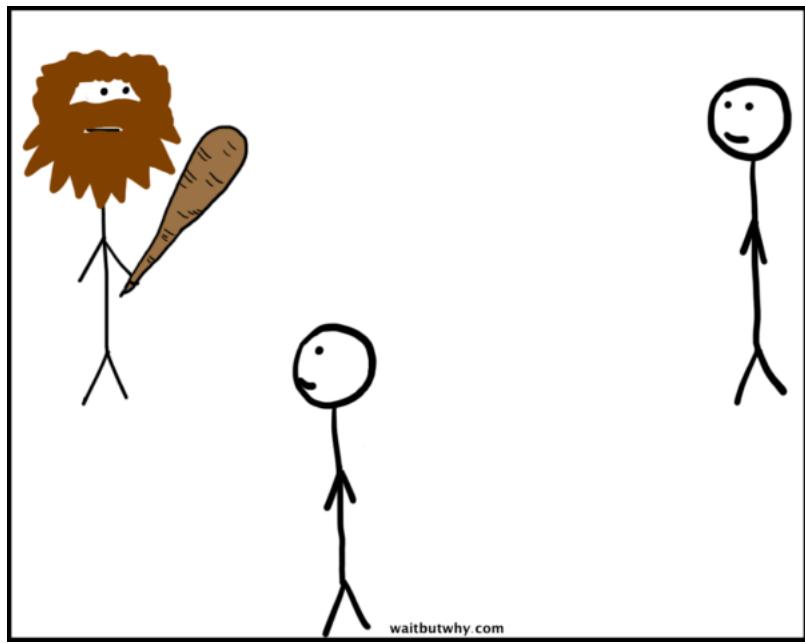
Okay that was fun. How'd it go, Bok?

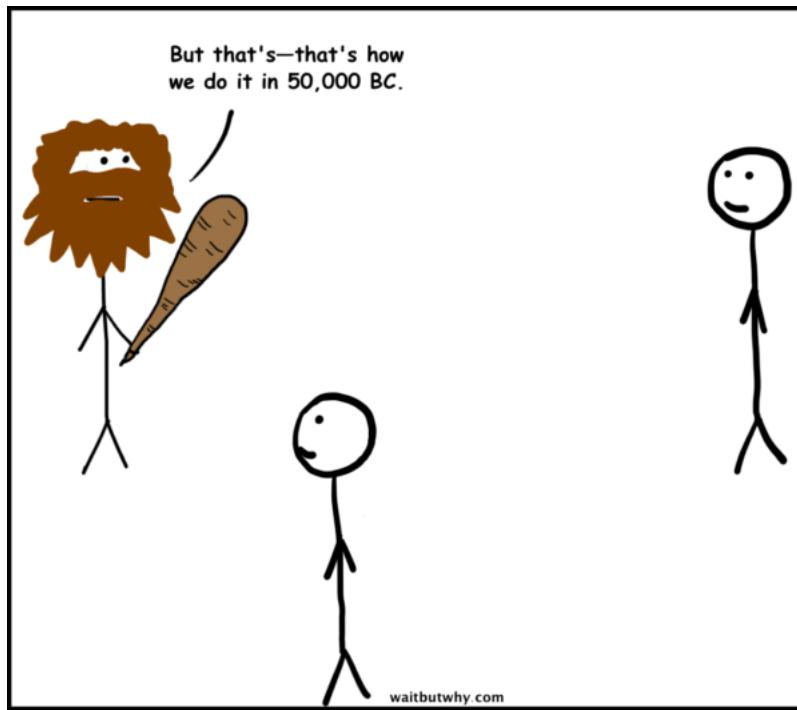


waitbutwhy.com

Yeah we figured that you'd be pretty surprised. To wrap up, let's show him how we communicate with each other.







Bok would be shocked to learn that despite all the magical powers humans have gained as a result of having learned to speak to each other, when it comes to actually speaking to each other, we're no more magical than the people of his day. When two people are together and talking, they're using 50,000-year-old technology.

Bok might also be surprised that in a world run by fancy machines, the people who made all the machines are walking around with the same biological bodies that Bok and his friends walk around with. How can that be?



This is why brain-machine interfaces—a subset of the broader field of neural engineering, which itself is a subset of biotechnology—are such a tantalizing new industry. We've conquered the world many times over with our technology, but when it comes to our brains—our most central tool—the tech world has for the most part been too daunted to dive in.

That's why we still communicate using technology Bok invented, it's why I'm typing this sentence at about a 20th of the speed that I'm thinking it, and it's why brain-related ailments still leave so many lives badly impaired or lost altogether.

But 50,000 years after the brain's great "aha!" moment, that may finally be about to change. The brain's next great frontier may be itself.

There are many kinds of potential brain-machine interface (sometimes called a brain-computer interface) that will serve many different functions. But everyone working on BMIs is grappling with either one or both of these two questions:

1) How do I get the right information out of the brain?

2) How do I send the right information into the brain?

The first is about capturing the brain's *output*—it's about **recording** what neurons are saying.

The second is about *inputting* information into the brain's natural flow or altering that natural flow in some other way—it's about **stimulating** neurons.

These two things are happening naturally in your brain all the time. Right now, your eyes are making a specific set of horizontal movements that allow you to read this sentence. That's the brain's neurons outputting information to a machine (your eyes) and the machine receiving the command and responding. And as your eyes move in just the right way, the photons from the screen are entering your retinas and stimulating neurons in the occipital lobe of your cortex in a way that allows the image of the words to enter your mind's eye. That image then stimulates neurons in another part of your brain that allows you to process the information embedded in the image and absorb the sentence's meaning.

Inputting and outputting information is what the brain's neurons *do*. All the BMI industry wants to do is get in on the action.

At first, this seems like maybe not *that* difficult a task? The brain is just a jello ball, right? And the cortex—the part of the brain in which we want to do most of our recording and stimulating—is just a napkin, located conveniently right on the outside of the brain where it can be easily accessed. Inside the cortex are around 20 billion firing neurons—20 billion oozy little transistors that if we can just learn to work with, will give us an entirely new level of control over our life, our health, and the world. Can't we figure that out? Neurons are small, but we know how to split an atom. A neuron's diameter is about 100,000 times as large as an atom's—if an atom were a marble, a neuron would be a kilometer across—so we should probably be able to handle the smallness. Right?

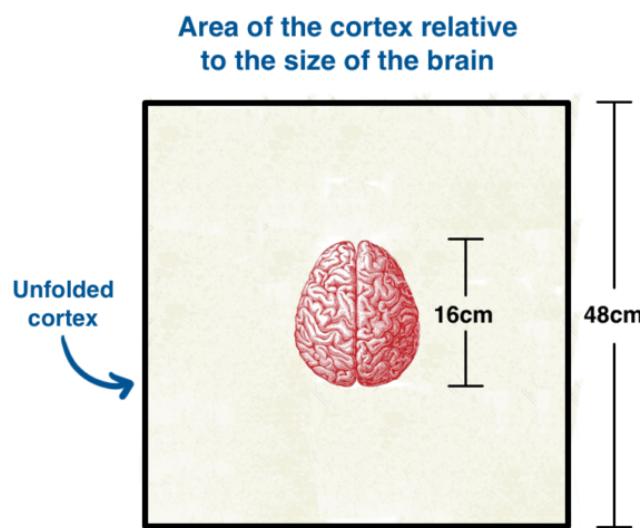
So what's the issue here?

Well on one hand, there's something to that line of thinking, in that because of those facts, this *is* an industry where immense progress can happen. We *can* do this.

But only when you understand what actually goes on in the brain do you realize why this is probably the hardest human endeavor in the world.

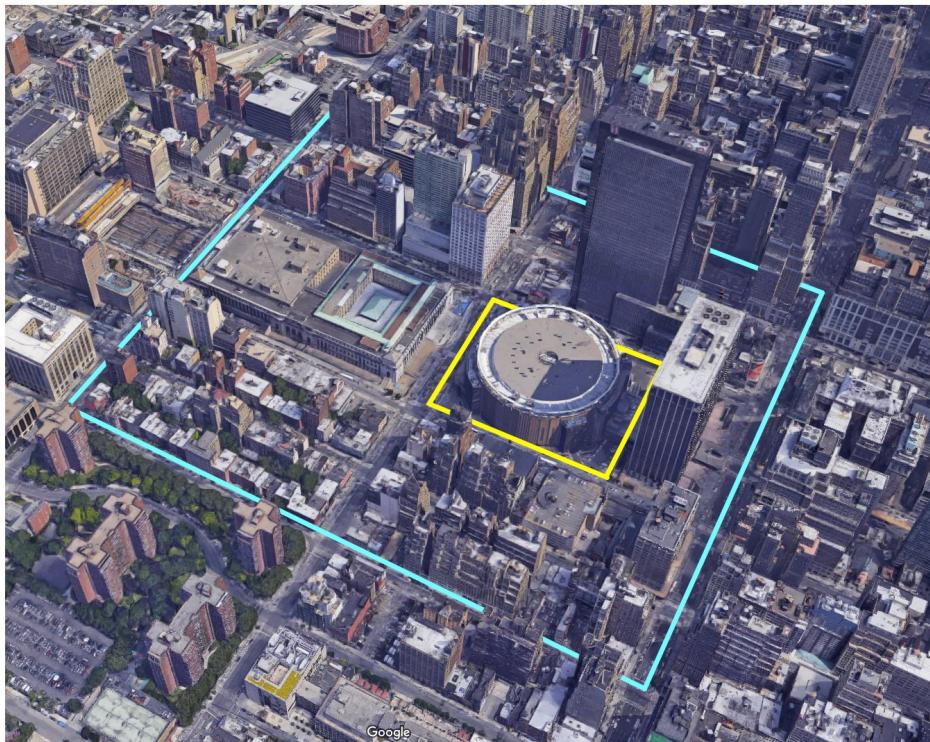
So before we talk about BMIs themselves, we need to take a closer look at what the people trying to make BMIs are dealing with here. I find that the best way to illustrate things is to scale the brain up by exactly 1,000X and look at what's going on.

Remember our cortex-is-a-napkin demonstration earlier?



Well if we scale that up by 1,000X, the cortex napkin—which was about 48cm / 19in on each side—now has a side the length of six Manhattan street blocks (or two avenue blocks). It would take you about 25 minutes to walk around the perimeter. And the brain as a whole would now fit snugly inside a two block

by two block square—just about the size of Madison Square Garden (this works in length and width, but the brain would be about double the height of MSG).

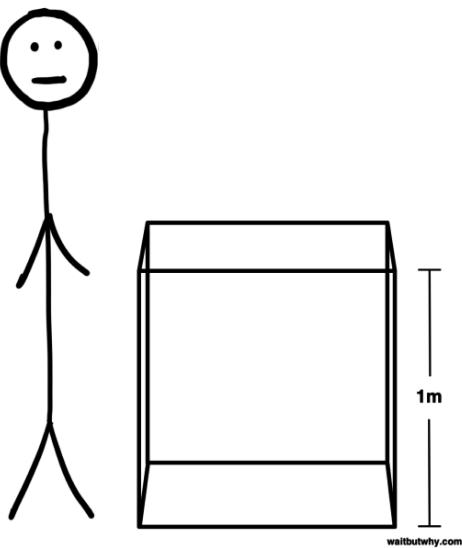


So let's lay it out in the actual city. I'm sure the few hundred thousand people who live there will understand.



I chose 1,000X as our multiplier for a couple reasons. One is that we can all instantly convert the sizes in our heads. Every millimeter of the actual brain is now a meter. And in the much smaller world of neurons, every micron is now an easy-to-conceptualize millimeter. Secondly, it conveniently brings the cortex up to human size—its 2mm thickness is now two meters—the height of a tall (6'6") man.

So we could walk up to 29th street, to the edge of our giant cortex napkin, and easily look at what was going on inside those two meters of thickness. For our demonstration, let's pull out a cubic meter of our giant cortex to examine, which will show us what goes on in a typical cubic millimeter of real cortex.



What we'd see in that cubic meter would be a *mess*. Let's empty it out and put it back together.

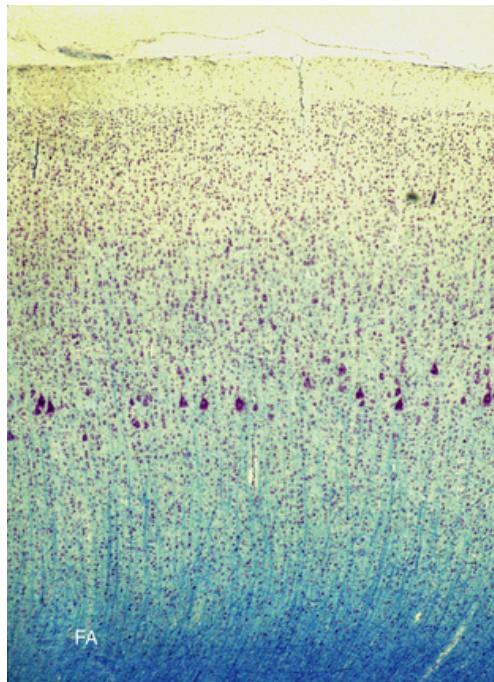
First, let's put the somas¹⁹ in—the little bodies of all the neurons that live in that cube.

Somas range in size, but the neuroscientists I spoke with said that the somas of neurons in the cortex are often around 10 or 15µm in diameter (μm = micrometer, or micron: 1/1,000th of a millimeter). That means that if you laid out 7 or 10 of them in a line, that line would be about the diameter of a human hair (which is about 100µm). On our scale, that makes a soma 1 – 1.5cm in diameter. A marble.

The volume of the whole cortex is in the ballpark of 500,000 cubic millimeters, and in that space are about 20 billion somas. That means an average cubic millimeter of cortex contains about 40,000 neurons. So there are 40,000 marbles in our cubic meter box. If we divide our box into about 40,000 cubic spaces, each with a side of 3cm (or about a cubic inch), it means each of our soma marbles is at the center of its own little 3cm cube, with other somas about 3cm away from it in all directions.

With me so far? Can you visualize our meter cube with those 40,000 floating marbles in it?

Here's a microscope image of the somas in an actual cortex, using techniques that block out the other stuff around them:²⁷



Okay not too crazy so far. But the soma is only a tiny piece of each neuron. Radiating out from each of our marble-sized somas are twisty, branchy dendrites that in our scaled-up brain can stretch out for three or four meters in many different directions, and from the other end an axon that can be over 100 meters long (when heading out laterally to another part of the cortex) or as long as a *kilometer* (when

heading down into the spinal cord and body). Each of them only about a millimeter thick, these cords turn the cortex into a dense tangle of electrical spaghetti.

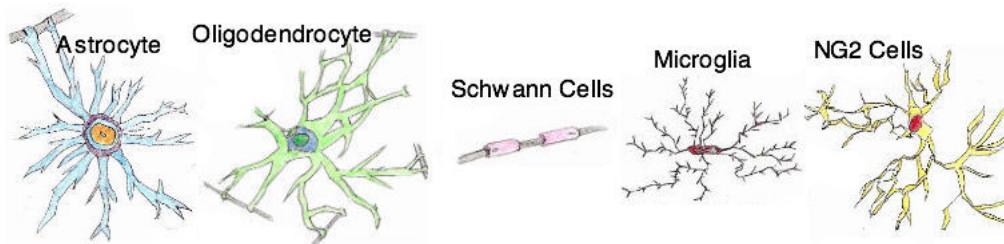
And there's a lot going on in that mash of spaghetti. Each neuron has synaptic connections to as many as 1,000—sometimes as high as 10,000—other neurons. With around 20 billion neurons in the cortex, that means there are over 20 *trillion* individual neural connections in the cortex (and as high as a quadrillion connections in the entire brain). In our cubic meter alone, there will be over 20 million synapses.

To further complicate things, not only are there many spaghetti strands coming out of *each* of the 40,000 marbles in our cube, but there are thousands of other spaghetti strings passing through our cube from *other* parts of the cortex. That means that if we were trying to record signals or stimulate neurons in this particular cubic area, we'd have a lot of difficulty, because in the mess of spaghetti, it would be very hard to figure out which spaghetti strings belonged to *our* soma marbles (and god forbid there are *Purkinje cells* in the mix).

And of course, there's the whole neuroplasticity thing. The voltages of each neuron would be constantly changing, as many as hundreds of times per second. And the tens of millions of synapse connections in our cube would be regularly changing sizes, disappearing, and reappearing.

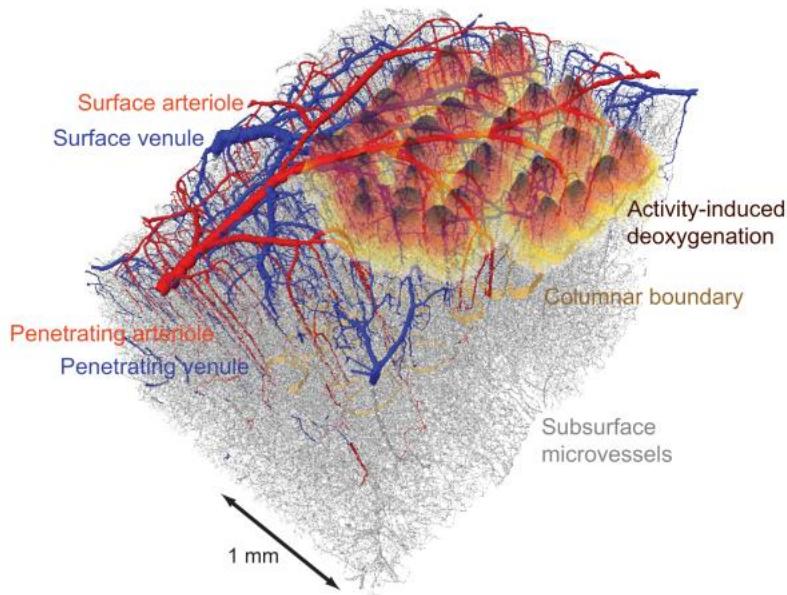
If only that were the end of it.

It turns out there are other cells in the brain called **glial cells**—cells that come in many different varieties and perform many different functions, like mopping up chemicals released into synapses, wrapping axons in myelin, and serving as the brain's immune system. Here are some common types of glial cell:²⁸



And how many glial cells are in the cortex? About the *same number as there are neurons*.²⁰ So add about 40,000 of these wacky things into our cube.

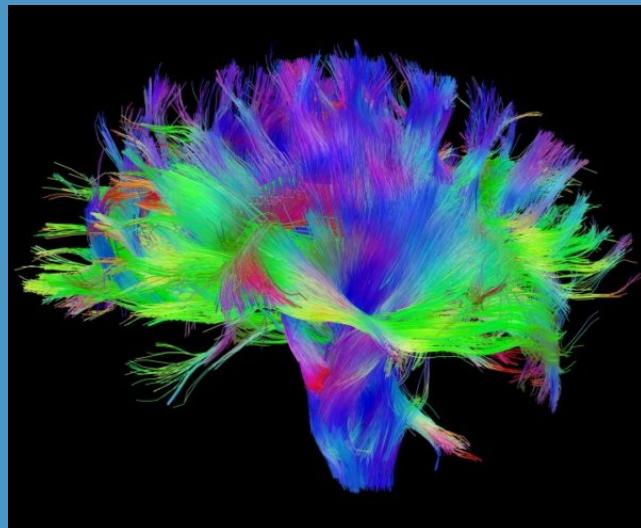
Finally, there are the blood vessels. In every cubic millimeter of cortex, there's a total of a *meter* of tiny blood vessels. On our scale, that means that in our cubic meter, there's a *kilometer* of blood vessels. Here's what the blood vessels in a space about that size look like:²⁹



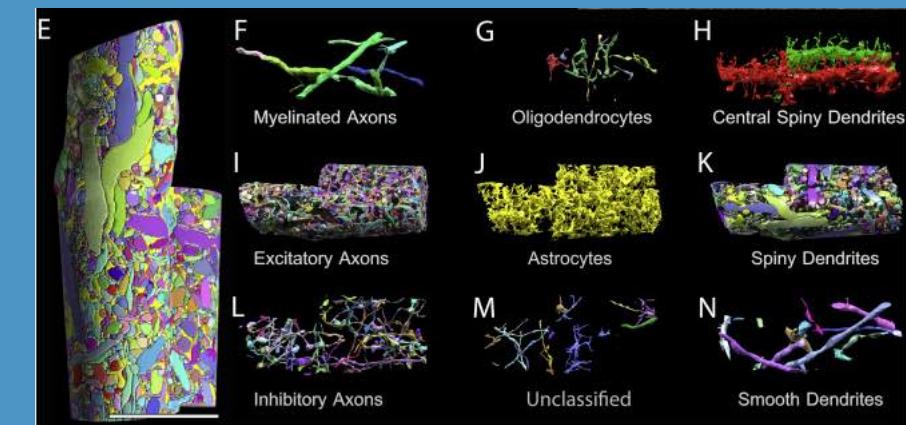
There's an amazing project going on right now in the neuroscience world called the [Human Connectome Project](#) (pronounced "connec-tome") in which scientists are trying to create a complete detailed map of the entire human brain. Nothing close to this scale of brain mapping has ever been done.^{[21](#)}

The project entails [slicing a human brain](#) into outrageously thin slices—around 30-nanometer-thick slices. That's 1/33,000th of a millimeter (here's a machine [slicing up a mouse brain](#)).

Anyway, in addition to producing some [gorgeous images](#) of the "ribbon" formations axons with similar functions often form inside white matter, like—



—the connectome project has helped people visualize just how packed the brain is with all this stuff. Here's a breakdown of all the different things going on in one tiny snippet of mouse brain (and this doesn't even include the blood vessels):^{[20](#)}



(In the image, E is the complete brain snippet, and F–N show the separate components that make up E.)

So our meter box is a jam-packed, oozy, electrified mound of dense complexity—now let's recall that in reality, everything in our box actually fits in a cubic *millimeter*.

And the brain-machine interface engineers need to figure out what the microscopic somas buried in that millimeter are saying, and other times, to stimulate *just* the right somas to get them to do what the engineers want. *Good luck with that.*

We'd have a super hard time doing that on our 1,000X brain. Our 1,000X brain that also happens to be a nice flat napkin. That's not how it normally works—usually, the napkin is up on top of our Madison Square Garden brain and full of deep folds (on our scale, between five and 30 meters deep). In fact, less than a *third* of the cortex napkin is up on the surface of the brain—most is buried inside the folds.

Also, engineers are not operating on a bunch of brains in a lab. The brain is covered with all those Russian doll layers, including the skull—which at 1,000X would be around seven meters thick. And since

most people don't really want you opening up their skull for very long—and ideally not at all—you have to try to work with those tiny marbles as non-invasively as possible.

And this is all assuming you're dealing with the cortex—but a lot of cool BMI ideas deal with the structures down below, which if you're standing on top of our MSG brain, are buried 50 or 100 meters under the surface.

The 1,000X game also hammers home the sheer scope of the brain. Think about how much was going on in our cube—and now remember that that's only *one 500,000th of the cortex*. If we broke our whole giant cortex into similar meter cubes and lined them up, they'd stretch 500km / 310mi—all the way to Boston and beyond. And if you made the trek—which would take over 100 hours of brisk walking—at any point you could pause and look at the cube you happened to be passing by and it would have all of this complexity inside of it. *All* of this is currently in your brain.

Part 3A: How Happy Are You That This Isn't Your Problem

Totes.

Back to Part 3: Brain-Machine Interfaces

So how do scientists and engineers begin to manage this situation?

Well they do the best they can with the tools they currently have—tools used to record or stimulate neurons (we'll focus on the recording side for the time being). Let's take a look at the options:

BMI Tools

With the current work that's being done, three broad criteria seem to stand out when evaluating a type of recording tool's pros and cons:

- 1) **Scale** – how many neurons can be simultaneously recorded
- 2) **Resolution** – how detailed is the information the tool receives—there are two types of resolution, **spatial** (how closely your recordings come to telling you how individual neurons are firing) and **temporal** (how well you can determine *when* the activity you record happened)
- 3) **Invasiveness** – is surgery needed, and if so, how extensively

The long-term goal is to have all three of your cakes and eat them all. But for now, it's always a question of "which one (or two) of these criteria are you willing to completely fail?" Going from one tool to another isn't an overall upgrade or downgrade—it's a tradeoff.

Let's examine the types of tools currently being used:

fMRI

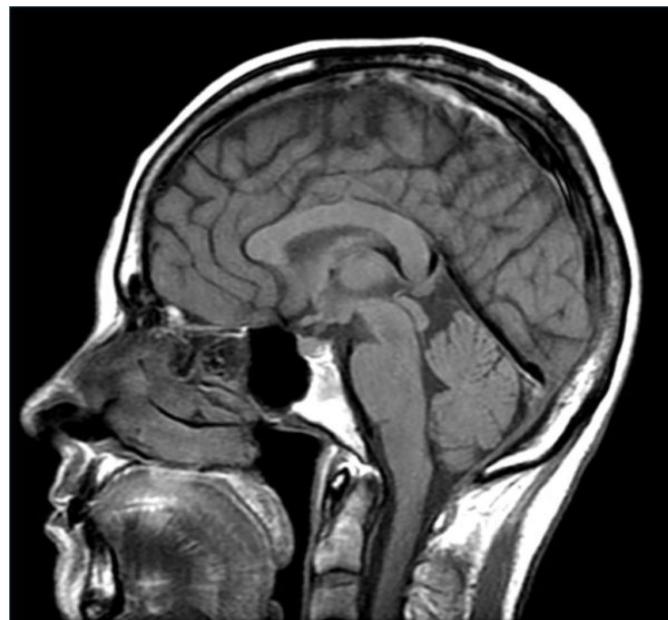
Scale: high (it shows you information across the whole brain)

Resolution: medium-low spatial, very low temporal

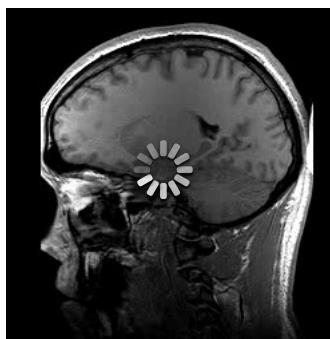
Invasiveness: non-invasive

fMRI isn't typically used for BMIs, but it is a classic *recording* tool—it gives you information about what's going on inside the brain.

fMRI uses MRI—magnetic resonance imaging—technology. MRIs, invented in the 1970s, were an evolution of the x-ray-based CAT scan. Instead of using x-rays, MRIs use magnetic fields (along with radio waves and other signals) to generate images of the body and brain. Like this:³¹

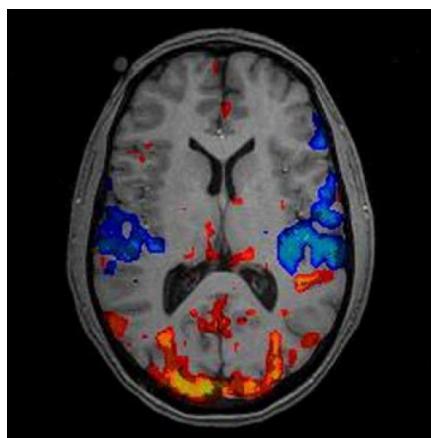


And this full set of cross sections, allowing you to see through an entire head.

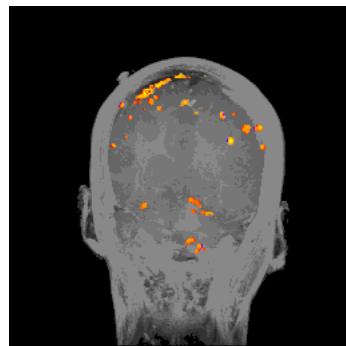


Pretty amazing technology.

fMRI (“functional” MRI) uses similar technology to track changes in *blood flow*. Why? Because when areas of the brain become more active, they use more energy, so they need more oxygen—so blood flow increases to the area to deliver that oxygen. Blood flow indirectly indicates where activity is happening. Here’s what an fMRI scan might show: ³²



Of course, there’s always blood throughout the brain—what this image shows is where blood flow has *increased* (red/orange/yellow) and where it has *decreased* (blue). And because fMRI can scan through the whole brain, results are 3-dimensional:



fMRI has many medical uses, like informing doctors whether or not certain parts of the brain are functioning properly after a stroke, and fMRI has taught neuroscientists a ton about which regions of the brain are involved with which functions. Scans also have the benefit of providing info about what's going on in the whole brain at any given time, and it's safe and totally non-invasive.

The big drawback is resolution. fMRI scans have a literal resolution, like a computer screen has with pixels, except the pixels are three-dimensional, cubic volume pixels—or “voxels.”

fMRI voxels have gotten smaller as the technology has improved, bringing the spatial resolution up. Today's fMRI voxels can be as small as a cubic millimeter. The brain has a volume of about 1,200,000mm³, so a high-resolution fMRI scan divides the brain into about one million little cubes. The problem is that on neuron scale, that's still pretty huge (the same size as our scaled-up cubic meter above)—each voxel contains tens of thousands of neurons. So what the fMRI is showing you, at best, is the average blood flow drawn in by each group of 40,000 or so neurons.

The even bigger problem is temporal resolution. fMRI tracks blood flow, which is both imprecise and comes with a delay of about a second—an eternity in the world of neurons.

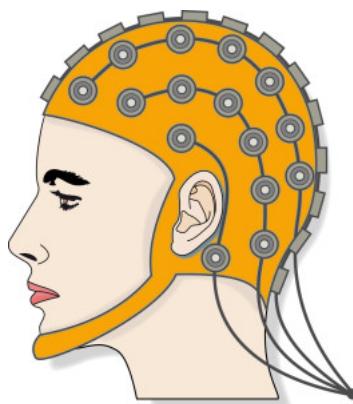
EEG

Scale: high

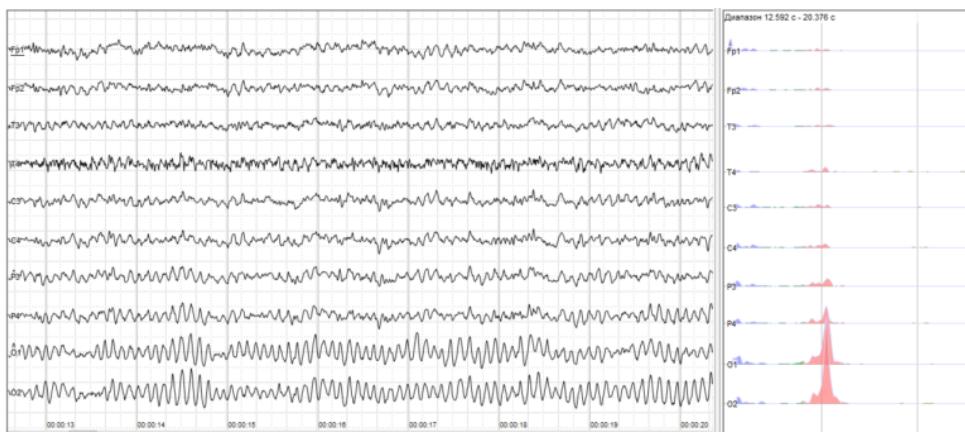
Resolution: very low spatial, medium-high temporal

Invasiveness: non-invasive

Dating back almost a century, EEG (electroencephalography) puts an array of electrodes on your head. You know, this whole thing:³³



EEG is definitely technology that will look hilariously primitive to a 2050 person, but for now, it's one of the only tools that can be used with BMIs that's totally non-invasive. EEGs record electrical activity in different regions of the brain, displaying the findings like this:³⁴



EEG graphs can uncover information about medical issues like epilepsy, track sleep patterns, or be used to determine something like the status of a dose of anesthesia.

And unlike fMRI, EEG has pretty good temporal resolution, getting electrical signals from the brain right as they happen—though the skull blurs the temporal accuracy considerably (bone is a bad conductor).

The major drawback is spatial resolution. EEG has none. Each electrode only records a broad average—a vector sum of the charges from millions or billions of neurons (and a blurred one because of the skull).

Imagine that the brain is a baseball stadium, its neurons are the members of the crowd, and the information we want is, instead of electrical activity, vocal cord activity. In that case, EEG would be like a group of microphones placed outside the stadium, against the stadium's outer walls. You'd be able to hear when the crowd was cheering and maybe predict the type of thing they were cheering about. You'd be able to hear telltale signs that it was between innings and maybe whether or not it was a close game. You could probably detect when something abnormal happened. But that's about it.

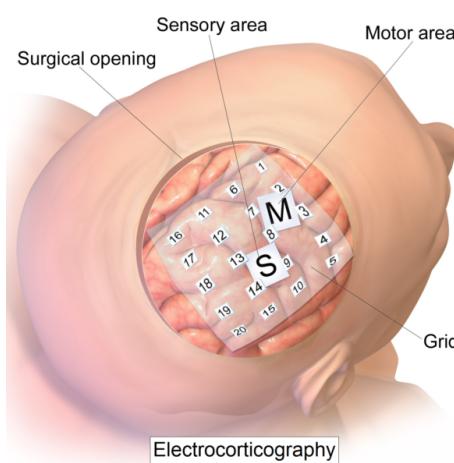
ECoG

Scale: high

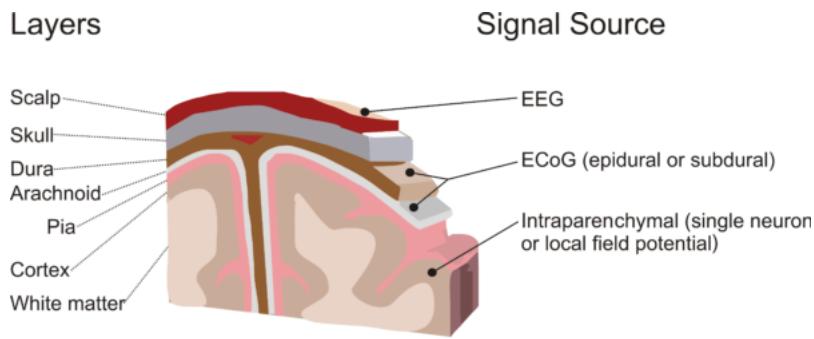
Resolution: low spatial, high temporal

Invasiveness: kind of invasive

ECoG (electrocorticography) is a similar idea to EEG, also using surface electrodes—except they put them *under* the skull, on the surface of the brain.³⁵



Ick. But effective—at least much more effective than EEG. Without the interference of the skull blurring things, ECoG picks up both higher spatial (about 1cm) and temporal resolution (5 milliseconds). ECoG electrodes can either be placed above or below the dura:³⁶



Bringing back our stadium analogy, ECoG microphones are *inside* the stadium and a bit closer to the crowd. So the sound is much crisper than what EEG mics get from outside the stadium, and ECoG mics can better distinguish the sounds of individual sections of the crowd. But the improvement comes at a cost—it requires invasive surgery. In the scheme of invasive surgeries, though, it's not so bad. As one neurosurgeon described to me, "You can slide stuff underneath the dura relatively non-invasively. You still have to make a hole in the head, but it's relatively non-invasive."

Local Field Potential

Scale: low

Resolution: medium-low spatial, high temporal

Invasiveness: very invasive

Okay here's where we shift from surface electrode discs to microelectrodes—tiny needles surgeons stick into the brain.

Brain surgeon Ben Rapoport described to me how his father (a neurologist) used to make microelectrodes:

When my father was making electrodes, he'd make them by hand. He'd take a very fine wire—like a gold or platinum or iridium wire, that was 10-30 microns in diameter, and he'd insert that wire in a glass capillary tube that was maybe a millimeter in diameter. Then they'd take that piece of glass over a flame and rotate it until the glass became soft. They'd stretch out the capillary tube until it's incredibly thin, and then take it out of the flame and break it. Now the capillary tube is flush with and pinching the wire. The glass is an insulator and the wire is a conductor. So what you end up with is a glass-insulated stiff electrode that is maybe a few 10s of microns at the tip.

Today, while some electrodes are still made by hand, newer techniques use silicon wafers and manufacturing technology borrowed from the integrated circuits industry.

The way local field potentials (LFP) work is simple—you take one of these super thin needles with an electrode tip and stick it one or two millimeters into the cortex. There it picks up the average of the electrical charges from all of the neurons within a certain radius of the electrode.

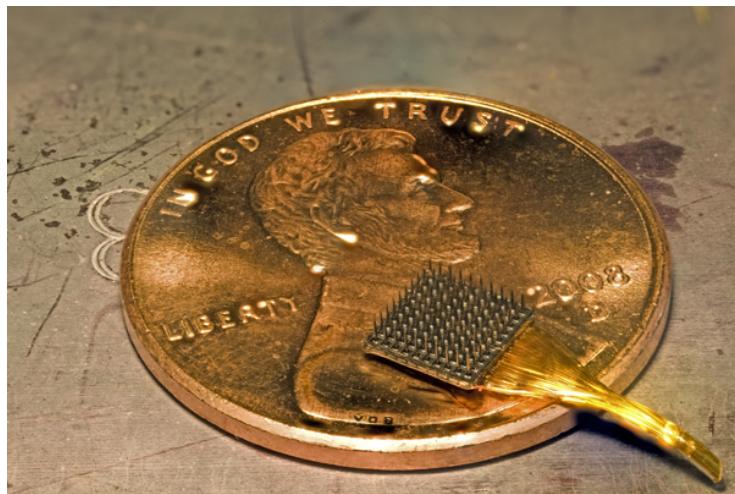
LFP gives you the not-that-bad spatial resolution of the fMRI combined with the instant temporal resolution of an ECoG. Kind of the best of all the worlds described above when it comes to resolution.

Unfortunately, it does badly on both other criteria.

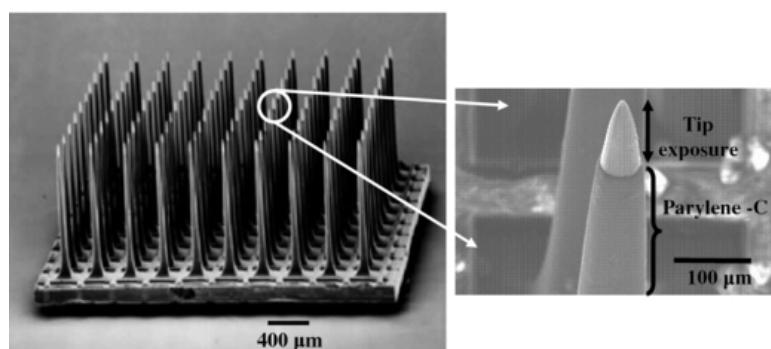
Unlike fMRI, EEG, and ECoG, microelectrode LFP does not have scale—it *only* tells you what the little sphere surrounding it is doing. And it's far more invasive, actually entering the brain.

In the baseball stadium, LFP is a single microphone hanging over a single section of seats, picking up a crisp feed of the sounds in that area, and maybe picking out an individual voice for a second here and there—but otherwise only getting the general vibe.

A more recent development is the **multielectrode array**, which is the same idea as the LFP except it's about 100 LFPs all at once, in a single area of the cortex. A multielectrode array looks like this: ³⁷



A tiny 4mm x 4mm square with 100 tiny silicon electrodes on it. Here's another image where you can see just how sharp the electrodes are—just a few microns across at the very tip:³⁸



Single-Unit Recording

Scale: tiny

Resolution: super high

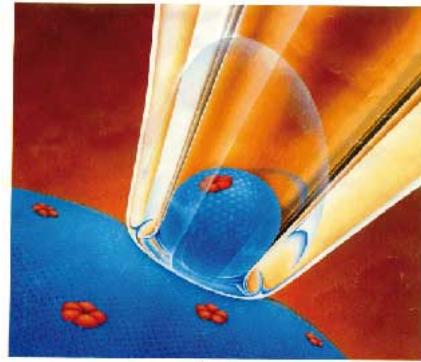
Invasiveness: very invasive

To record a broader LFP, the electrode tip is a bit rounded to give the electrode more surface area, and they turn the resistance down with the intent of allowing very faint signals from a wide range of locations to be picked up. The end result is the electrode picks up a chorus of activity from the local field.

Single-unit recording also uses a needle electrode, but they make the tip super sharp and crank up the resistance. This wipes out most of the noise and leaves the electrode picking up almost nothing—until it finds itself so close to a neuron (maybe 50μm away) that the signal from that neuron is strong enough to make it past the electrode's high resistance wall. With distinct signals from one neuron and no background noise, this electrode can now voyeur in on the private life of a single neuron. Lowest possible scale, highest possible resolution.

By the way, you can listen to a neuron fire [here](#) (what you're actually hearing is the electro-chemical firing of a neuron, converted to audio).

Some electrodes want to take the relationship to the next level and will go for a technique called the **patch clamp**, whereby it'll get rid of its electrode tip, leaving just a tiny little tube called a glass pipette,²² and it'll actually directly assault a neuron by sucking a “patch” of its membrane into the tube, allowing for even finer measurements:³⁹



A patch clamp also has the benefit that, unlike all the other methods we've discussed, because it's physically touching the neuron, it can not only record but *stimulate* the neuron,²³ injecting current or holding voltage at a set level to do specific tests (other methods can stimulate neurons, but only entire groups together).

Finally, electrodes can fully defile the neuron and actually penetrate through the membrane, which is called **sharp electrode recording**. If the tip is sharp enough, this won't destroy the cell—the membrane will actually seal around the electrode, making it very easy to stimulate the neuron or record the voltage difference between the inside and outside of the neuron. But this is a short-term technique—a punctured neuron won't survive long.

In our stadium, a single unit recording is a one-directional microphone clipped to a single crowd member's collar. A patch clamp or sharp recording is a mic in someone's throat, registering the exact movement of their vocal cords. This is a great way to learn about that person's experience at the game, but it also gives you no context, and you can't really tell if the sounds and reactions of that person are representative of what's going on in the game.

And that's about what we've got, at least in common usage. These tools are simultaneously unbelievably advanced and what will seem like Stone Age technology to future humans, who won't believe you had to choose *either* high-res *or* a wide field and that you actually had to open someone's skull to get high-quality brain readouts or write-ins.

But given their limitations, these tools have taught us worlds about the brain and led to the creation of some amazing early BMIs. Here's what's already out there—

The BMIs we already have

In 1969, a researcher named Eberhard Fetz connected a single neuron in a monkey's brain to a dial in front of the monkey's face. The dial would move when the neuron was fired. When the monkey would think in a way that fired the neuron and the dial would move, he'd get a banana-flavored pellet. Over time, the monkey started getting *better* at the game because he wanted more delicious pellets. The monkey had learned to make the neuron fire and inadvertently became the subject of the first real brain-machine interface.

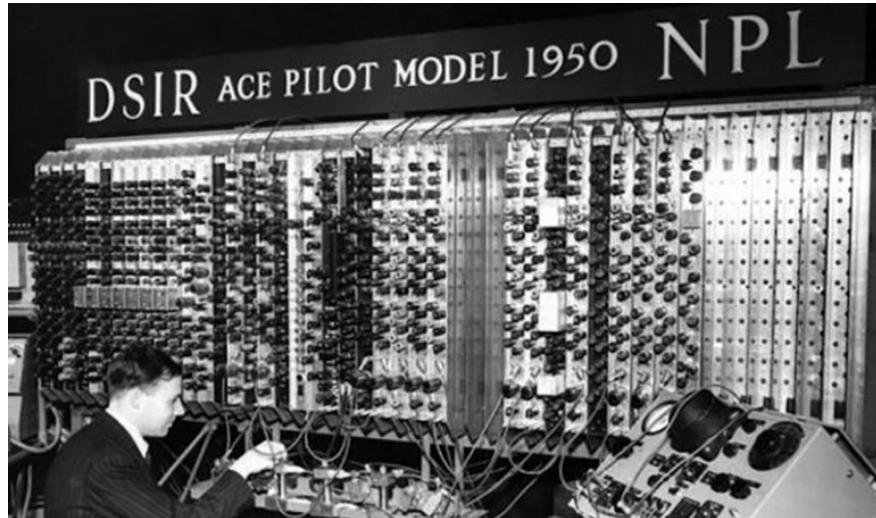
Progress was slow over the next few decades, but by the mid-90s, things had started to move, and it's been quietly accelerating ever since.

Given that both our understanding of the brain and the electrode hardware we've built are pretty primitive, our efforts have typically focused on building straightforward interfaces to be used with the areas of the brain we understand the best, like the motor cortex and the visual cortex.

And given that human experimentation is only really possible for people who are trying to use BMIs to alleviate an impairment—and because that's currently where the market demand is—our efforts have focused so far almost entirely on restoring lost function to people with disabilities.

The major BMI industries of the future that will give all humans magical superpowers and transform the world are in their fetal stage right now—and we should look at what's being worked on as a set of clues about what the mind-boggling worlds of 2040 and 2060 and 2100 might be like.

Like, check this out:



That's a computer built by Alan Turing in 1950 called the Pilot ACE. Truly cutting edge in its time.

Now check this out:



As you read through the examples below, I want you to think about this analogy—

Pilot ACE is to iPhone 7

as

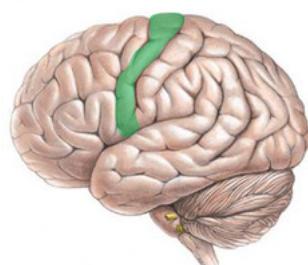
Each BMI example below is to ____

—and try to imagine what the blank looks like. And we'll come back to the blank later in the post.

Anyway, from everything I've read about and discussed with people in the field, there seem to be three major categories of brain-machine interface being heavily worked on right now:

Early BMI type #1: Using the motor cortex as a remote control

In case you forgot this from 9,000 words ago, the motor cortex is this guy:



All areas of the brain confuse us, but the motor cortex confuses us less than almost all the other areas. And most importantly, it's *well-mapped*, meaning specific parts of it control specific parts of the body (remember the upsetting [homunculus](#)?).

Also importantly, it's one of the major areas of the brain in charge of our *output*. When a human *does* something, the motor cortex is almost always the one pulling the strings (at least for the physical part of the doing). So the human brain doesn't really have to *learn* to use the motor cortex as a remote control, because the brain already uses the motor cortex as its remote control.

Lift your hand up. Now put it down. See? Your hand is like a little toy drone, and your brain just picked up the motor cortex remote control and used it to make the drone fly up and then back down.

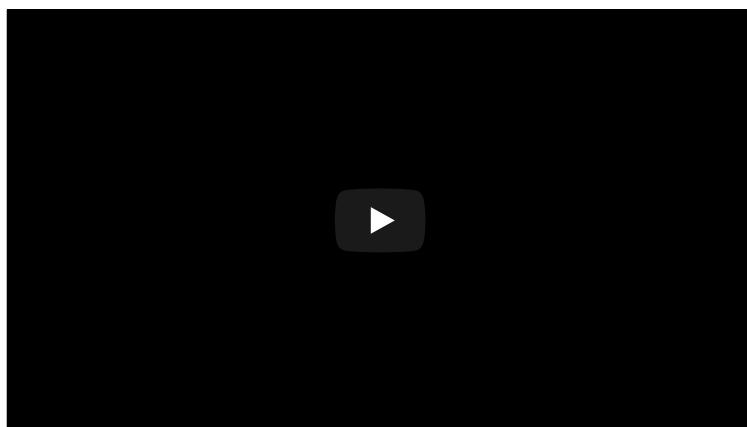
The goal of motor cortex-based BMIs is to tap into the motor cortex, and then when the remote control fires a command, to *hear* that command and then send it to some kind of machine that can respond to it the way, say, your hand would. A bundle of nerves is the middleman between your motor cortex and your hand. BMIs are the middleman between your motor cortex and a computer. Simple.

One barebones type of interface allows a human—often a person paralyzed from the neck down or someone who has had a limb amputated—to move a cursor on a screen with only their thoughts.

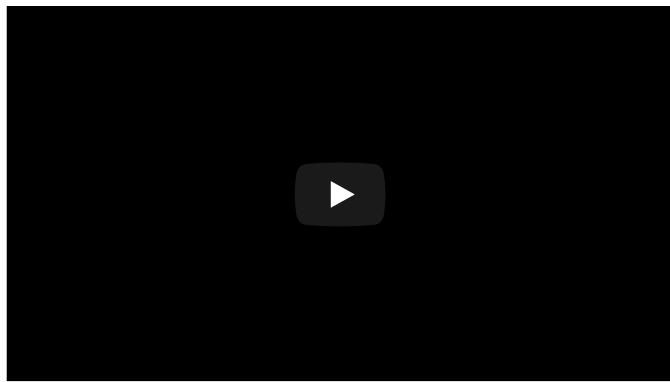
This begins with a 100-pin multielectrode array being implanted in the person's motor cortex. The motor cortex in a paralyzed person usually works just fine—it's just that the spinal cord, which had served as the middleman between the cortex and the body, stopped doing its job. So with the electrode array implanted, researchers have the person *try* to move their arm in different directions. Even though they can't do that, the motor cortex still fires normally, as if they can.

When someone moves their arm, their motor cortex bursts into a flurry of activity—but each neuron is usually only interested in one type of movement. So one neuron might fire whenever the person moves their arm to the right—but it's bored by other directions and is less active in those cases. That neuron alone, then, could tell a computer when the person wants to move their arm to the right and when they don't. But that's all. But with an electrode *array*, 100 single-unit electrodes each listen to a different neuron.²⁴ So when they do testing, they'll ask the person to try to move their arm to the right, and maybe 38 of the 100 electrodes detect their neuron firing. When the person tries to go left with their arm, maybe 41 others fire. After going through a bunch of different movements and directions and speeds, a computer takes the data from the electrodes and synthesizes it into a general understanding of which firing patterns correspond to which movement intentions on an X-Y axis.

Then when they link up that data to a computer screen, the person can use their mind, via "trying" to move the cursor, to really control the cursor. And this actually works. Through the work of motor-cortex-BMI pioneer company [BrainGate](#), here's a guy playing a video game using only his mind.



And if 100 neurons can tell you where they want to move a cursor, why couldn't they tell you when they want to pick up a mug of coffee and take a sip? That's what this quadriplegic woman did:



Another quadriplegic woman [flew an F-35 fighter jet](#) in a simulation, and a monkey recently used his mind to [ride around in a wheelchair](#).

And why stop with arms? Brazilian BMI pioneer [Miguel Nicolelis](#) and his team built an entire [exoskeleton](#) that allowed a paralyzed man to make the [opening kick](#) of the World Cup.²⁵

The Proprioception Blue Box

Moving these kinds of “neuroprosthetics” is all about the recording of neurons, but for these devices to be truly effective, this needs to not be a one-way street, but a *loop* that includes recording *and* stimulation pathways. We don’t really think about this, but a huge part of your ability to pick up an object is all of the incoming sensory information your hand’s skin and muscles send back in (called “proprioception”). In one video I saw, a woman with numbed fingers tried to light a match, and it was almost impossible for her to do it, despite having no other disabilities. And the beginning of [this video](#) shows the physical struggles of a man with a perfectly functional motor cortex but impaired proprioception. So for something like a bionic arm to really feel like an arm, and to really be useful, it needs to be able to send sensory information back in.

Stimulating neurons is even harder than recording them. As researcher Flip Sabes explained to me:

If I record a pattern of activity, it doesn’t mean I can readily recreate that pattern of activity by just playing it back. You can compare it to the planets in the Solar System. You can watch the planets move around and record their movements. But then if you jumble them all up and later want to recreate the original motion of one of the planets, you can’t just take that one planet and put it back into its orbit, because it’ll be influenced by all the other planets. Likewise, neurons aren’t just working in isolation—so there’s a fundamental irreversibility there. On top of that, with all of the axons and dendrites, it’s hard to just stimulate the neurons you want to—because when you try, you’ll hit a whole jumble of them.

Flip’s lab tries to deal with these challenges by getting the brain to help out. It turns out that if you reward a monkey with a succulent sip of orange juice when a single neuron fires, eventually the monkey will learn to make the neuron fire on demand. The neuron could then act as another kind of remote control. This means that normal motor cortex commands are only one possibility as a control mechanism. Likewise, until BMI technology gets good enough to perfect stimulation, you can use the brain’s neuroplasticity as a shortcut. If it’s too hard to make someone’s bionic fingertip touch something and send back information that feels just like the kind of sensation their own fingertip used to give them, the arm could instead send some *other* signal into the brain. At first, this would seem odd to the patient—but eventually the brain can learn to treat that signal as a new sense of touch. This concept is called “sensory substitution” and makes the brain a collaborator in BMI efforts.

In these developments are the seeds of other future breakthrough technologies—like brain-to-brain communication.

Nicolelis created an [experiment](#) where the motor cortex of one rat in Brazil was wired, via the internet, to the motor cortex of another rat in the US. The rat in Brazil was presented with two transparent boxes, each with a lever attached to it, and inside one of the boxes would be a treat. To attempt to get the treat, the rat would press the lever of the box that held the treat. Meanwhile, the rat in the US was

in a similar cage with two similar boxes, except unlike the rat in Brazil, the boxes weren't transparent and offered him no information about which of his two levers would yield a treat and which wouldn't. The only info the US rat had were the signals his brain received from the Brazil rat's motor cortex. The Brazil rat had the key knowledge—but the way the experiment worked, the rats only received treats when the US rat pressed the correct lever. If he pulled the wrong one, neither would. The amazing thing is that over time, the rats got *better* at this and began to work together, almost like a single nervous system—even though neither had any idea the other rat existed. The US rat's success rate at choosing the correct lever with no information would have been 50%. With the signals coming from the Brazil rat's brain, the success rate jumped to 64%. (Here's a [video](#) of the rats doing their thing.)

This has even worked, crudely, in people. Two people, in separate buildings, [worked together](#) to play a video game. One could *see* the game, the other had the controller. Using simple EEG headsets, the player who could see the game would, without moving his hand, *think* about moving his hand to press the "shoot" button on a controller. Because their brains' devices were communicating with each other, the player with the controller would then feel a twitch in his finger and press the shoot button.

Early BMI type #2: Artificial ears and eyes

There are a couple reasons giving sound to the deaf and sight to the blind is among the more manageable BMI categories.

The first is that like the motor cortex, the sensory cortices are parts of the brain we tend to understand pretty well, partly because they too tend to be well-mapped.

The second is that in many early applications, we don't *really* need to deal with the brain—we can just deal with the place where ears and eyes *connect* to the brain, since that's often where the impairment is based.

And while the motor cortex stuff was mostly about *recording* neurons to get information *out* of the brain, artificial senses go the other way—*stimulation* of neurons to send information *in*.

On the ears side of things, recent decades have seen the development of the groundbreaking **cochlear implant**.

The How Hearing Works Blue Box

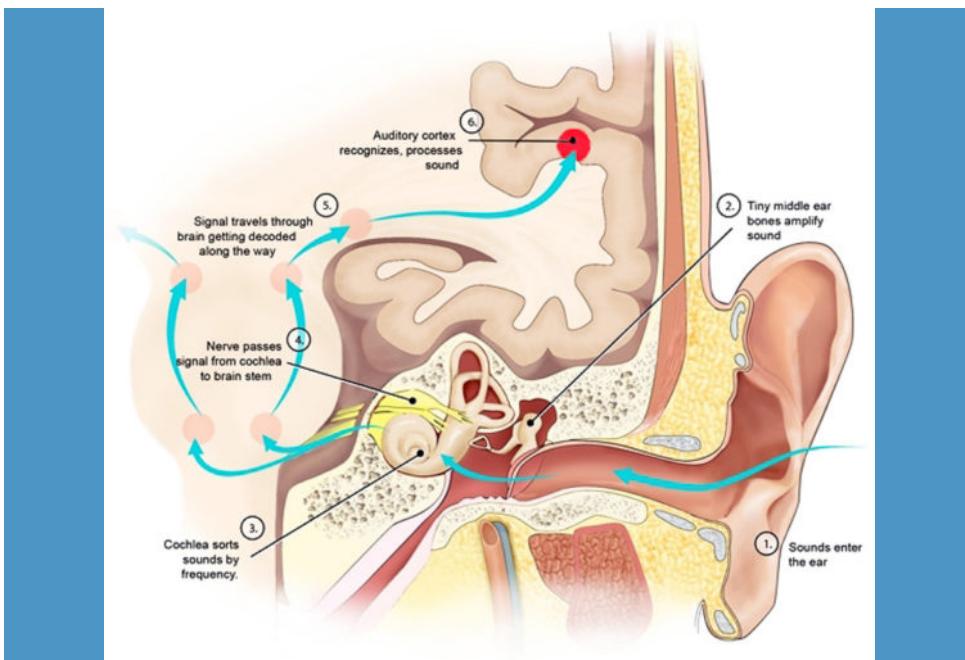
When you think you're "hearing" "sound," here's what's actually happening:

What we think of as sound is actually patterns of vibrations in the air molecules around your head. When a guitar string or someone's vocal cords or the wind or anything else makes a sound, it's because it's vibrating, which pushes nearby air molecules into a similar vibration and that pattern expands outward in a sphere, kind of like the surface of water expands outward in a circular ripple when something touches it.²⁶

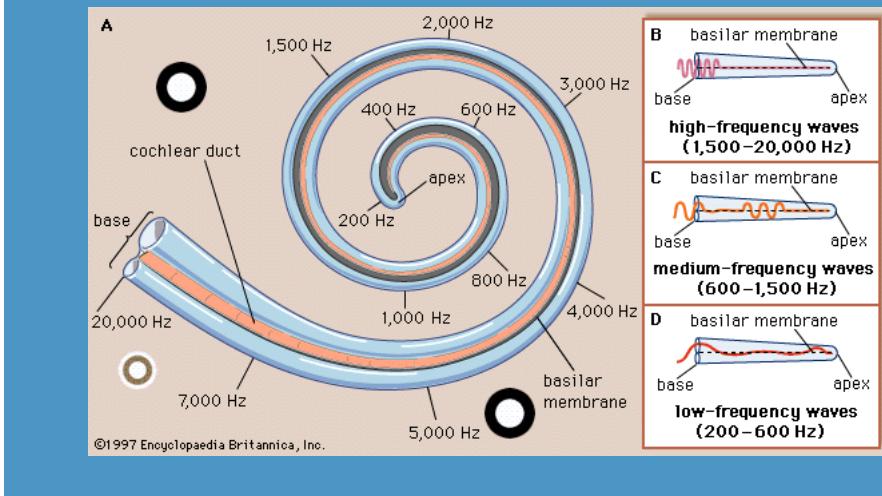
Your ear is a machine that converts those air vibrations into electrical impulses. Whenever air (or water, or any other medium whose molecules can vibrate) enters your ear, your ear translates the precise way it's vibrating into an electrical code that it sends into the nerve endings that touch it. This causes those nerves to fire a pattern of action potentials that send the code into your auditory cortex for processing. Your brain receives the information, and we call the experience of receiving that particular type of information "hearing."

Most people who are deaf or hard of hearing don't have a nerve problem or an auditory cortex problem—they usually have an ear problem. Their brain is as ready as anyone else's to turn electrical impulses into hearing—it's just that their auditory cortex isn't *receiving* any electrical impulses in the first place, because the machine that converts air vibrations into those impulses isn't doing its job.

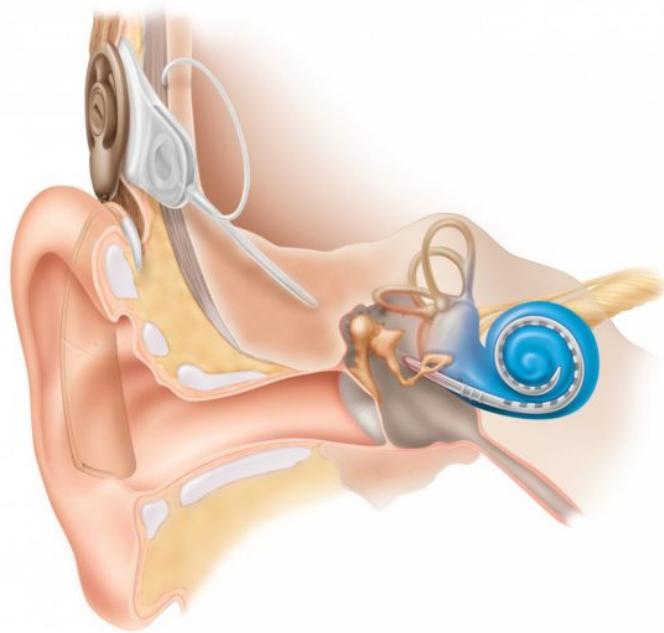
The ear has a lot of parts, but it's the cochlea in particular that makes the key conversion. When vibrations enter the fluid in the cochlea, it causes thousands of tiny hairs lining the cochlea to vibrate, and the cells those hairs are attached to transform the mechanical energy of the vibrations into electrical signals that then excite the auditory nerve. Here's what it all looks like:²⁰



The cochlea also sorts the incoming sound by frequency. Here's a cool chart that shows why lower sounds are processed at the end of the cochlea and high sounds are processed at the beginning (and also why there's a minimum and maximum frequency on what the ear can hear):⁴¹



A cochlear implant is a little computer that has a microphone coming out of one end (which sits on the ear) and a wire coming out of the other that connects to an array of electrodes that line the cochlea.



So sound comes into the microphone (the little hook on top of the ear), and goes into the brown thing, which processes the sound to filter out the less useful frequencies. Then the brown thing transmits the information through the skin, through electrical induction, to the computer's other component, which converts the info into electric impulses and sends them into the cochlea. The electrodes filter the impulses by frequency just like the cochlea and stimulate the auditory nerve just like the hairs on the cochlea do. This is what it looks like from the outside:



In other words, an artificial ear, performing the same sound-to-impulses-to-auditory-nerve function the ear does.

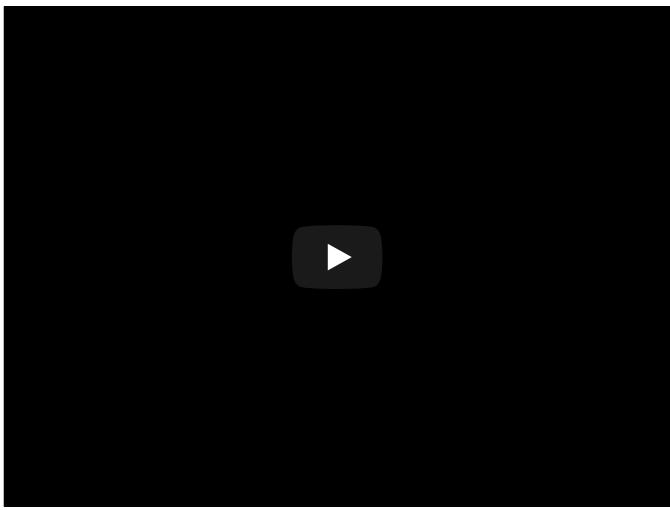
[Check out what sound sounds](#) like to someone with the implant.

Not great. Why? Because to send sound into the brain with the richness the ear hears with, you'd need 3,500 electrodes. Most cochlear implants have about 16. ²⁷ Crude.

But we're in the Pilot ACE era—so of course it's crude.

Still, today's cochlear implant allows deaf people to hear *speech* and have conversations, which is a groundbreaking development. ²⁸

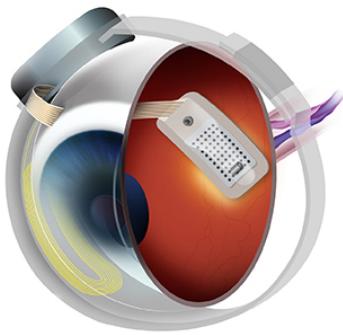
Many parents of deaf babies are now having a cochlear implant put in when the baby's about one year old. Like this baby, whose reaction to hearing for the first time is *cute*.



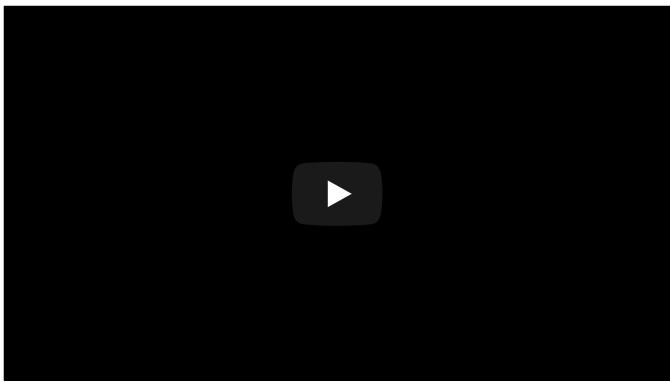
There's a similar revolution underway in the world of blindness, in the form of the **retinal implant**.

Blindness is often the result of a retinal disease. When this is the case, a retinal implant can perform a similar function for sight as a cochlear implant does for hearing (though less directly). It performs the normal duties of the eye and hands things off to nerves in the form of electrical impulses, just like the eye does.

A more complicated interface than the cochlear implant, the first retinal implant was approved by the FDA in 2011—the Argus II implant, made by [Second Sight](#). The retinal implant looks like this:⁴²



And it works like this:



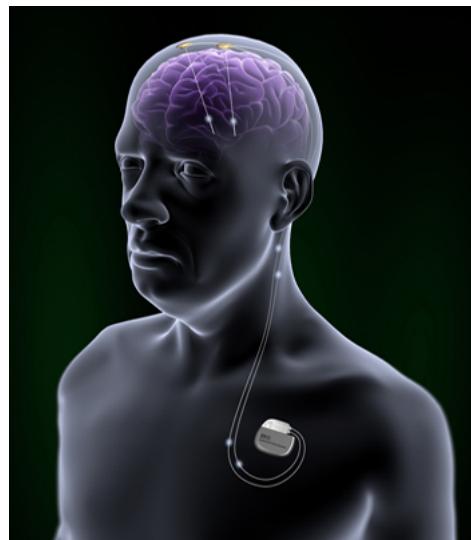
The retinal implant has 60 sensors. The retina has around a million neurons. Crude. But seeing vague edges and shapes and patterns of light and dark sure beats seeing nothing at all. What's encouraging is that you don't need a million sensors to gain a reasonable amount of sight—simulations suggest that 600-1,000 electrodes would be enough for reading and facial recognition.

Early BMI type #3: Deep brain stimulation

Dating back to the late 1980s, deep brain stimulation is yet another crude tool that is also still pretty life-changing for a lot of people.

It's also a type of category of BMI that doesn't involve communication with the outside world—it's about using brain-machine interfaces to treat or enhance yourself by altering something internally.

What happens here is one or two electrode wires, usually with [four separate electrode sites](#), are inserted into the brain, often ending up somewhere in the limbic system. Then a little pacemaker computer is implanted in the upper chest and wired to the electrodes. Like this unpleasant man:⁴³



The electrodes can then give a little zap when called for, which can do a variety of important things. Like:

- Reduce the tremors of people with Parkinson's Disease
- Reduce the severity of seizures
- Chill people with OCD out

It's also experimentally (not yet FDA approved) been able to mitigate certain kinds of chronic pain like migraines or phantom limb pain, treat anxiety or depression or PTSD, or even be combined with muscle stimulation elsewhere in the body to restore and retrain circuits that were broken down from stroke or a neurological disease.

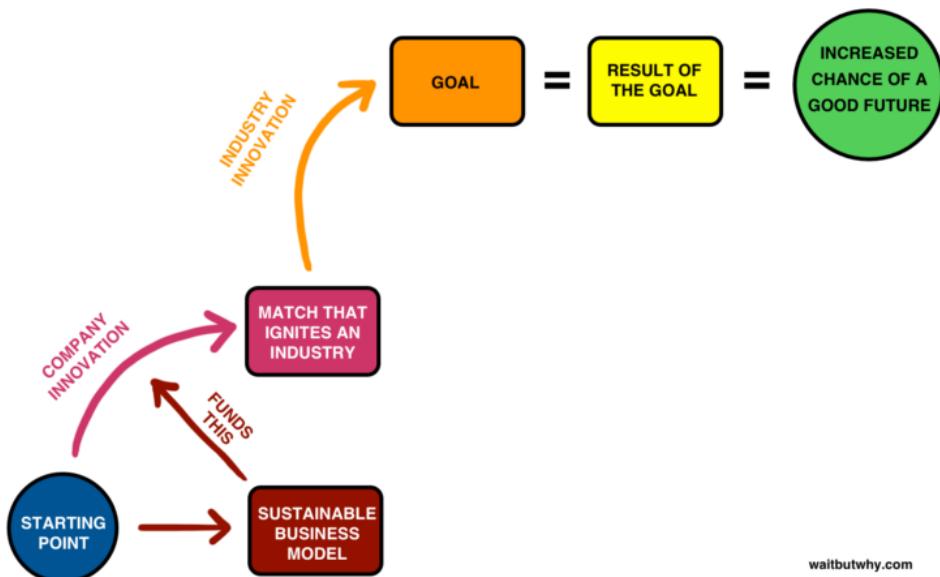
This is the state of the early BMI industry, and it's the moment when Elon Musk is stepping into it. For him, and for Neuralink, today's BMI industry is Point A. We've spent the whole post so far in the past, building up to the present moment. Now it's time to step into the future—to figure out what Point B is and how we're going to get there.

Part 4: Neuralink's Challenge



Having already written about two of Elon Musk's companies—[Tesla](#) and [SpaceX](#)—I think I understand his formula. It looks like this:

Elon Musk Company Formula



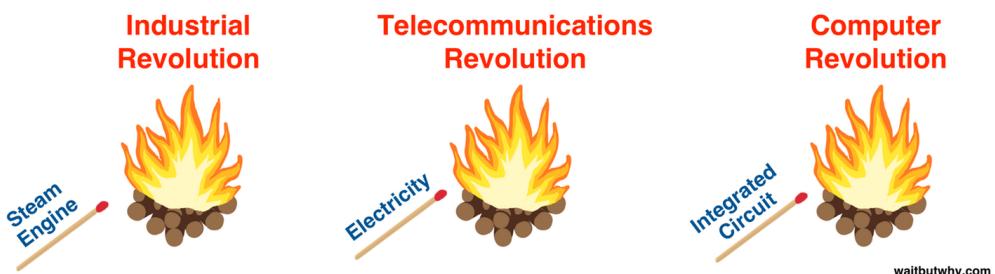
And his initial thinking about a new company always starts on the right and works its way left.

He decides that some specific change in the world will increase the likelihood of humanity having the best possible future. He knows that large-scale world change happens quickest when the whole world—the Human Colossus—is working on it. And he knows that the Human Colossus will work toward a goal if (and only if) there's an economic forcing function in place—if it's a good business decision to spend resources innovating toward that goal.

Often, before a booming industry starts booming, it's like a pile of logs—it has all the ingredients of a fire and it's ready to go—but there's no match. There's some technological shortcoming that's preventing the industry from taking off.

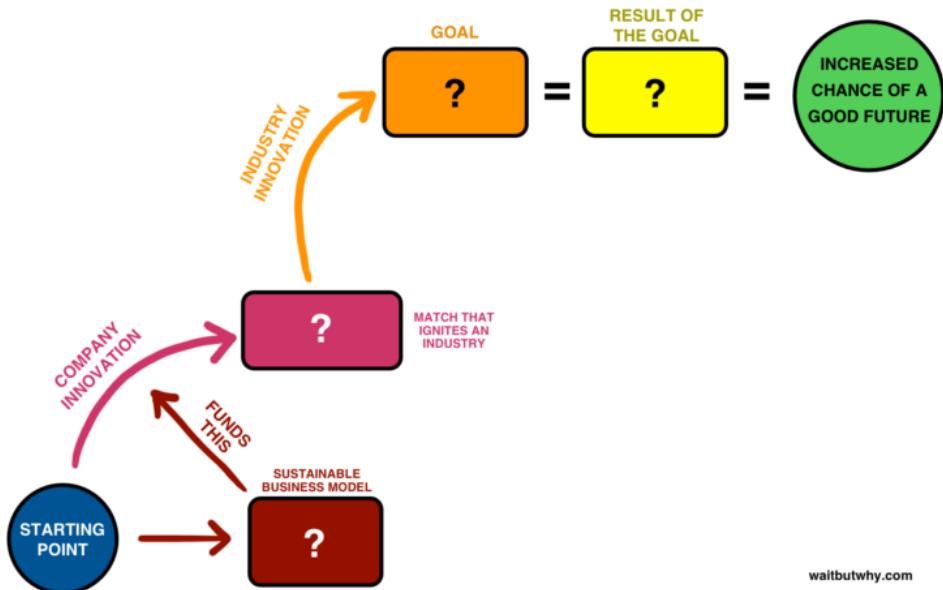
So when Elon builds a company, its core initial strategy is usually to create the match that will ignite the industry and get the Human Colossus working on the cause. This, in turn, Elon believes, will lead to developments that will change the world in the way that increases the likelihood of humanity having the best possible future. But you have to look at his companies from a zoomed-out perspective to see all of this. If you don't, you'll mistake what they do as their business for what they *do*—when in fact, what they do as their business is usually a mechanism to sustain the company while it innovates to try to make that critical match.

Back when I was working on the Tesla and SpaceX posts, I asked Elon why he went into engineering and not science, and he explained that when it comes to progress, “engineering is the limiting factor.” In other words, the progress of science, business, and industry are all at the whim of the progress of engineering. If you look at history, this makes sense—behind each of the greatest revolutions in human progress is an engineering breakthrough. A match.



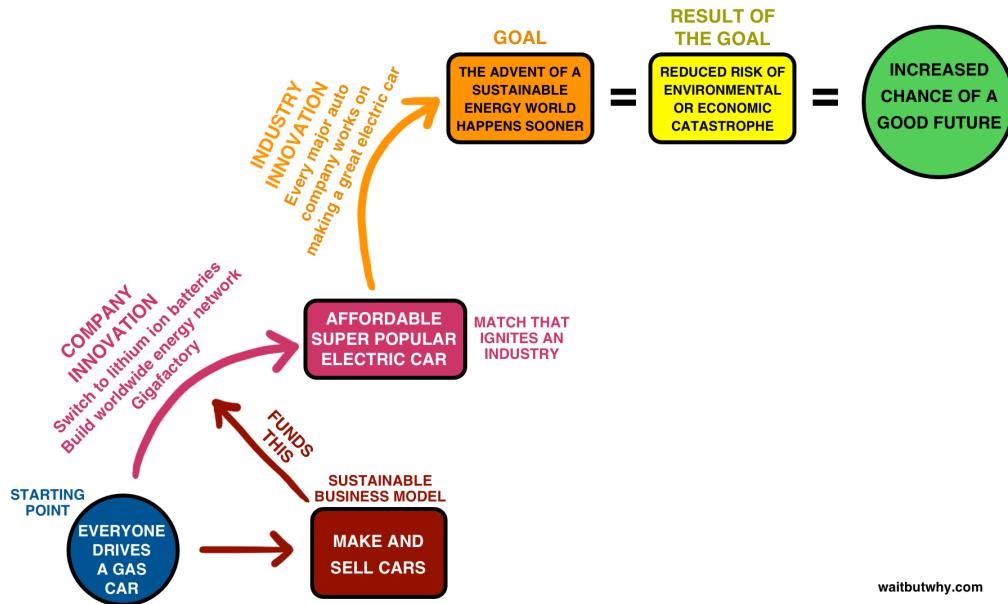
So to understand an Elon Musk company, you need to think about the match he's trying to create—along with three other variables:

Elon Musk Company Formula

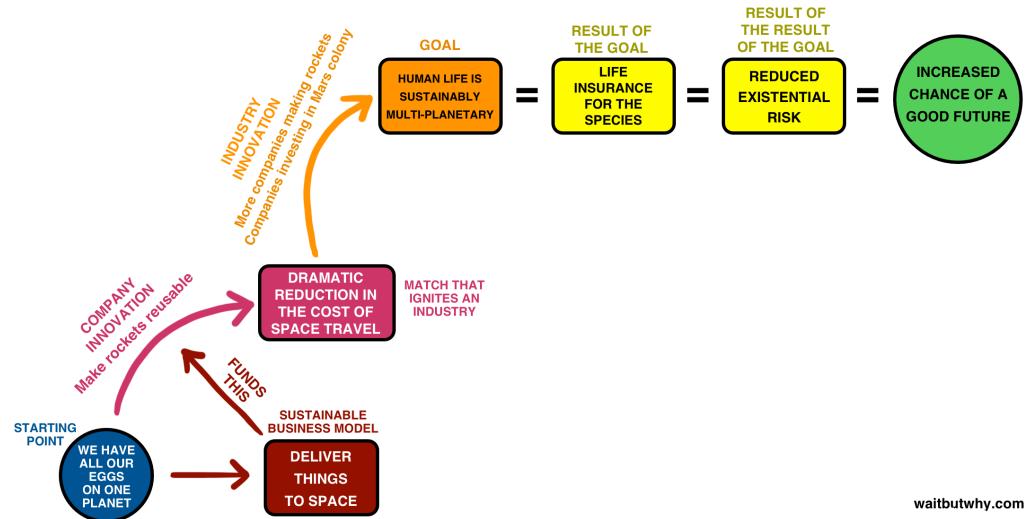


I know what's in these boxes with the other companies:

Tesla Formula

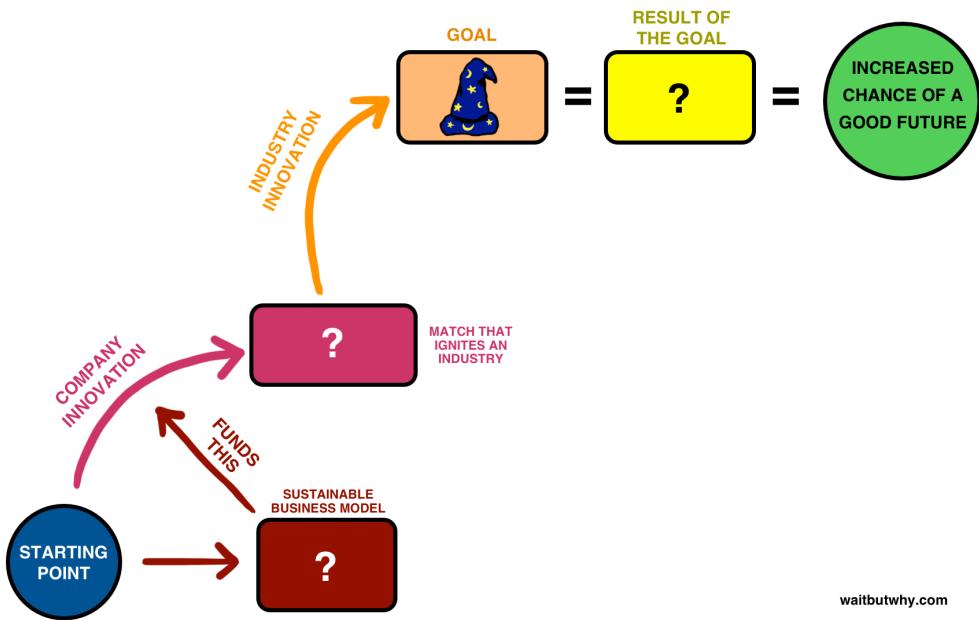


SpaceX Formula



And when I started trying to figure out what Neuralink was all about, I knew those were the variables I needed to fill in. At the time, I had only had the chance to get a very vague idea of one of the variables—that the goal of the company was “to accelerate the advent of a whole-brain interface.” Or what I’ve come to think of as a wizard hat.

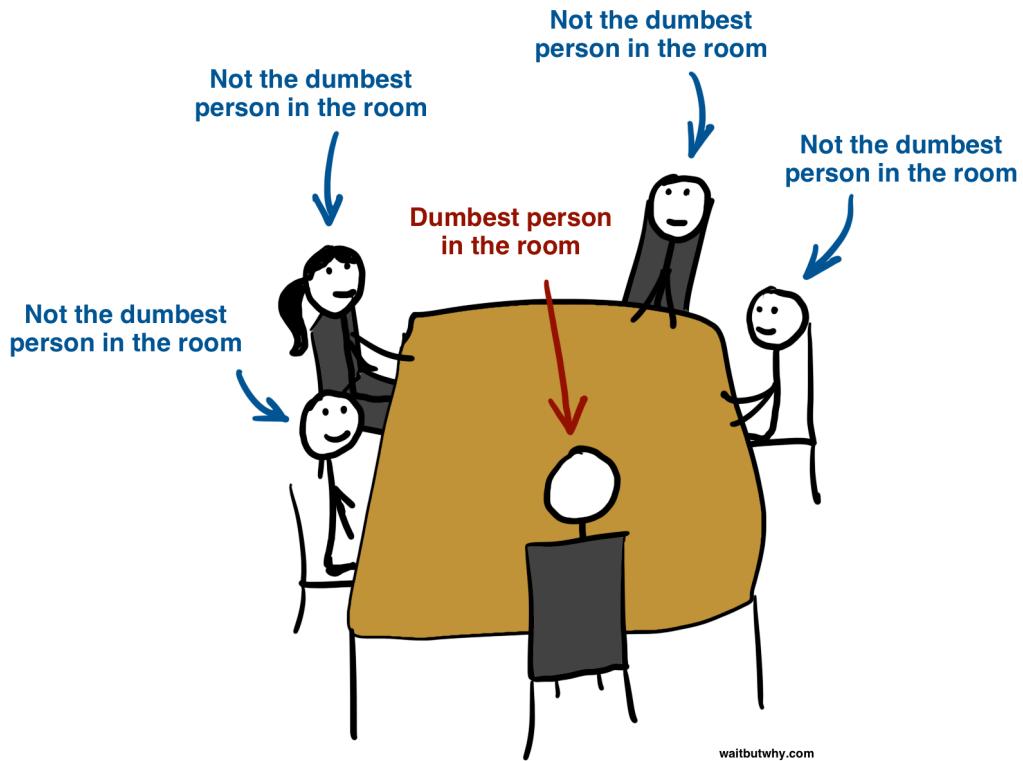
Neuralink Formula



As I understood it, a whole-brain interface was what a brain-machine interface would be in an ideal world—a super-advanced concept where essentially all the neurons in your brain are able to communicate seamlessly with the outside world. It was a concept loosely based on the science fiction idea of a “neural lace,” described in Iain Banks’ *Culture* series—a massless, volumeless, whole-brain interface that can be teleported into the brain.

I had a lot of questions.

Luckily, I was on my way to San Francisco, where I had plans to sit down with half of Neuralink’s founding team and be the dumbest person in the room.



The I'm Not Being Self-Deprecating I Really Was Definitely the Dumbest Person in the Room Just Look at This Shit Blue Box

The Neuralink team:

Paul Merolla, who spent the last seven years as the lead chip designer at IBM on their SyNAPSE program, where he led the development of the **TrueNorth chip**—one of the largest CMOS devices ever designed by transistor count nbd. Paul told me his field was called *neuromorphic*, where the goal is to design transistor circuits based on principles of brain architecture.

Vanessa Tolosa, Neuralink's microfabrication expert and one of the world's foremost researchers on biocompatible materials. Vanessa's work involves designing biocompatible materials based on principles from the integrated circuits industry.

Max Hodak, who worked on the development of some groundbreaking BMI technology at Miguel Nicolelis's lab at Duke while also commuting across the country twice a week in college to run **Transcriptic**, the "robotic cloud laboratory for the life sciences" he founded.

DJ Seo, who while at UC Berkeley in his mid-20s designed a cutting-edge new BMI concept called **neural dust**—tiny ultrasound sensors that could provide a new way to record brain activity.

Ben Rapoport, Neuralink's surgery expert, and a top neurosurgeon himself. But he also has a PhD in Electrical Engineering from MIT, allowing him to see his work as a neurosurgeon "through the lens of implantable devices."

Tim Hanson, whom a colleague described as "one of the best all-around engineers on the planet" and who self-taught himself enough about materials science and microfabrication methods to develop some of the core technology that'll be used at Neuralink.

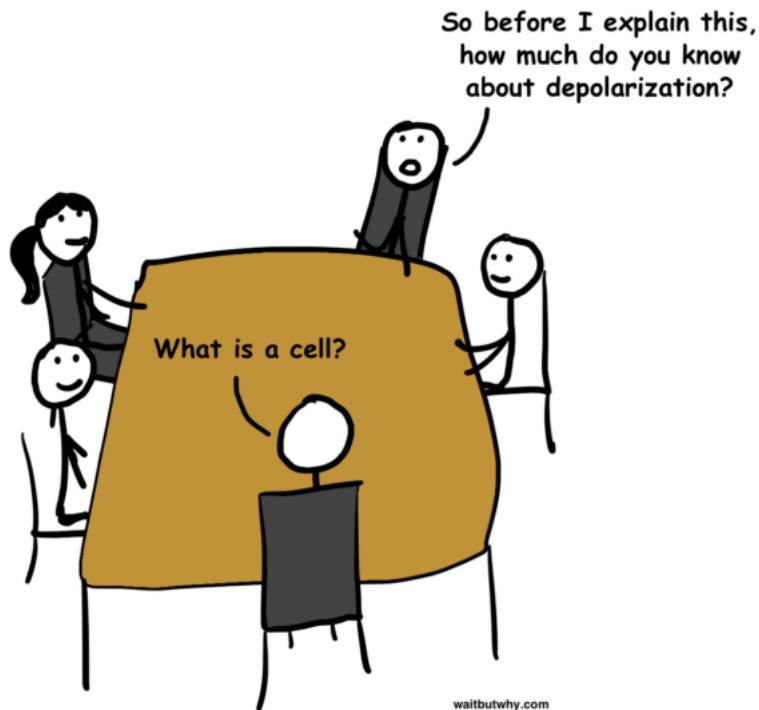
Flip Sabes, a leading researcher whose lab at UCSF has pioneered new ground in BMIs by combining "cortical physiology, computational and theoretical modeling, and human psychophysics and physiology."

Tim Gardner, a leading researcher at BU, whose lab works on implanting BMIs in *birds*, in order to study "how complex songs are assembled from elementary neural units" and learn about "the relationships between patterns of neural activity on different time-scales." Both Tim and Flip have left tenured positions to join the Neuralink team—pretty good testament to the promise they believe this company has.

And then there's Elon, both as their CEO and a fellow team member. Elon being CEO makes this different from other recent things he's started and puts Neuralink on the top tier for him, where only SpaceX and Tesla have lived. When it comes to neuroscience, Elon has the least technical knowledge on the team—but he also started SpaceX without very much technical knowledge and quickly became a certifiable rocket science expert by reading and by asking questions of the experts on the team. That'll probably happen again here. (And for good reason—he pointed out: "Without a strong technical understanding, I think it's hard to make the right decisions.")

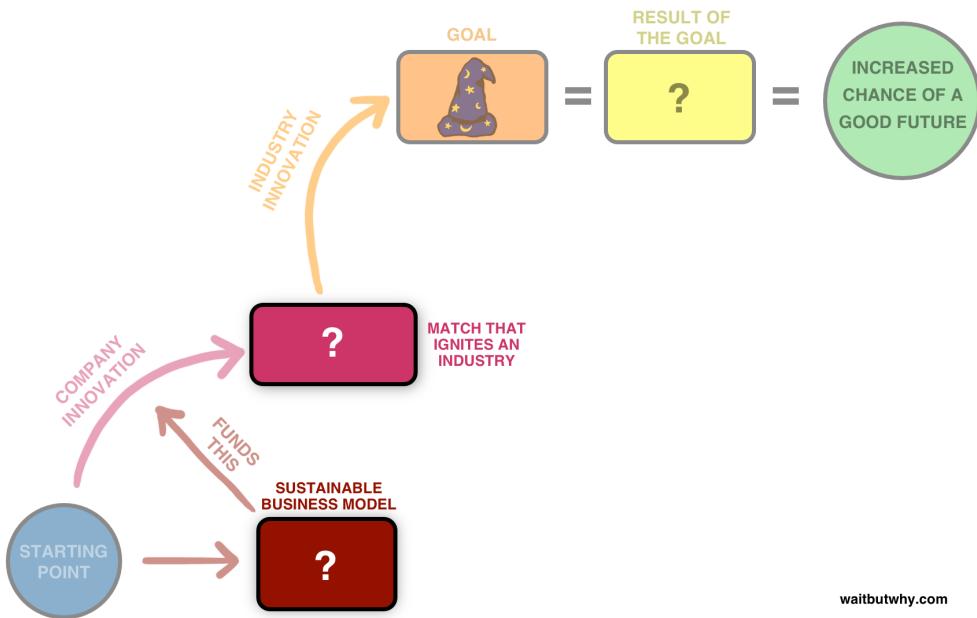
I asked Elon about how he brought this team together. He said that he met with literally over 1,000 people in order to assemble this group, and that part of the challenge was the large number of totally separate areas of expertise required when you're working on technology that involves neuroscience, brain surgery, microscopic electronics, clinical trials, etc. Because it was such a cross-disciplinary area, he looked for cross-disciplinary experts. And you can see that in those bios—everyone brings their own unique crossover combination to a group that together has the rare ability to think as a single mega-expert. Elon also wanted to find people who were totally on board with the zoomed-out mission—who were more focused on industrial results than producing white papers. Not an easy group to assemble.

But there they were, sitting around the table looking at me, as it hit me 40 seconds in that I should have done a lot more research before coming here.



They took the hint and dumbed it down about four notches, and as the discussion went on, I started to wrap my head around things. Throughout the next few weeks, I met with each of the remaining Neuralink founders as well, each time playing the role of the dumbest person in the room. In these meetings, I focused on trying to form a comprehensive picture of the challenges at hand and what the road to a wizard hat might look like. I really wanted to understand these two boxes:

Neuralink Formula



waitbutwhy.com

The first one was easy. The *business* side of Neuralink is a brain-machine interface development company. They want to create cutting-edge BMIs—what one of them referred to as “micron-sized devices.” Doing this will support the growth of the company while also providing a perfect vehicle for putting their innovations into practice (the same way SpaceX uses their launches both to sustain the company and experiment with their newest engineering developments).

As for what kind of interface they’re planning to work on first, here’s what Elon said:

We are aiming to bring something to market that helps with certain severe brain injuries (stroke, cancer lesion, congenital) in about four years.

The second box was a lot hazier. It seems obvious to us today that using steam engine technology to harness the power of fire was the thing that had to happen to ignite the Industrial Revolution. But if you talked to someone in 1760 about it, they would have had a lot less clarity—on exactly which hurdles they were trying to get past, what kinds of innovations would allow them to leap over those hurdles, or how long any of this would take. And that’s where we are here—trying to figure out what the match looks like that will ignite the neuro revolution and how to create it.

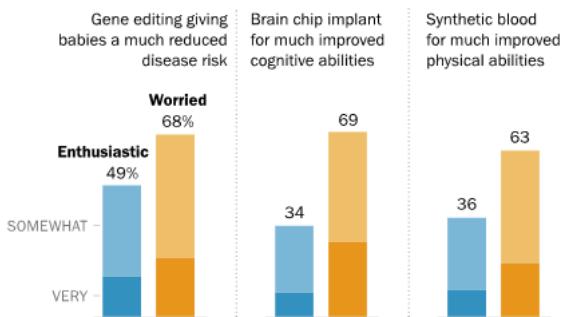
The starting place for a discussion about innovation is a discussion about hurdles—what are you even trying to innovate *past*? In Neuralink’s case, a whole lot of things. But given that, here too, engineering will likely prove to be the limiting factor, here are some seemingly large challenges that probably *won’t* end up being the major roadblock:

Public skepticism

Pew recently conducted a [survey](#) asking Americans about which future biotechnologies give them the shits the most. It turns out BMIs worry Americans even more than gene editing:⁴⁴

Public expresses more worry than enthusiasm about each of these potential human enhancements

% of U.S. adults who say they are ____ about each of these enhancements



Note: Respondents who gave other responses or who did not give an answer are not shown.

Source: Survey of U.S. adults conducted March 2-28, 2016.

"U.S. Public Wary of Biomedical Technologies to 'Enhance' Human Abilities"

PEW RESEARCH CENTER

Neuralink co-founder Flip Sabes doesn't get it.

To a scientist, to think about changing the fundamental nature of life—creating viruses, eugenics, etc.—it raises a specter that many biologists find quite worrisome, whereas the neuroscientists that I know, when they think about chips in the brain, it doesn't seem that foreign, because we already have chips in the brain. We have deep brain stimulation to alleviate the symptoms of Parkinson's Disease, we have early trials of chips to restore vision, we have the cochlear implant—so to us it doesn't seem like that big of a stretch to put devices into a brain to read information out and to read information back in.

And after learning all about chips in the brain, I agree—and when Americans eventually learn about it, I think they'll change their minds.

History supports this prediction. People were super timid about Lasik eye surgery when it first became a thing—20 years ago, 20,000 people a year had the procedure done. Then everyone got used to it and now 2,000,000 people a year get laser eye surgery. Similar story with pacemakers. And defibrillators. And organ transplants—which people at first considered a freakish Frankenstein-esque concept. Brain implants will probably be the same story.

Our non-understanding of the brain

You know, the whole "if understanding the brain is a mile, we're currently three inches in" thing. Flip weighed in on this topic too:

If it were a prerequisite to understand the brain in order to interact with the brain in a substantive way, we'd have trouble. But it's possible to decode all of those things in the brain without truly understanding the dynamics of the computation in the brain. Being able to read it out is an engineering problem. Being able to understand its origin and the organization of the neurons in fine detail in a way that would satisfy a neuroscientist to the core—that's a separate problem. And we don't need to solve all of those scientific problems in order to make progress.

If we can just use engineering to get neurons to talk to computers, we'll have done our job, and machine learning can do much of the rest. Which then, ironically, *will* teach us about the brain. As Flip points out:

The flip side of saying, "We don't need to understand the brain to make engineering progress," is that making engineering progress will almost certainly advance our scientific knowledge—kind of like the way Alpha Go ended up teaching the world's best players better strategies for the game. Then this scientific progress can lead to more engineering progress. The engineering and the science are gonna ratchet each other up here.

Angry giants

Tesla and SpaceX are both stepping on some very big toes (like the auto industry, the oil and gas industry, and the military-industrial complex). Big toes don't like being stepped on, so they'll usually do whatever they can to hinder the stepper's progress. Luckily, Neuralink doesn't really have this problem. There aren't any massive industries that Neuralink is disrupting (at least not in the foreseeable future—an eventual neuro revolution would disrupt almost *every* industry).

Neuralink's hurdles are technology hurdles—and there are many. But two challenges stand out as the largest—challenges that, if conquered, may be impactful enough to trigger all the other hurdles to fall

and totally change the trajectory of our future.

Major Hurdle 1: Bandwidth

There have never been more than a couple hundred electrodes in a human brain at once. When it comes to vision, that equals a super low-res image. When it comes to motor, that limits the possibilities to simple commands with little control. When it comes to your thoughts, a few hundred electrodes won't be enough to communicate more than the simplest spelled-out message.

We need higher bandwidth if this is gonna become a big thing. Way higher bandwidth.

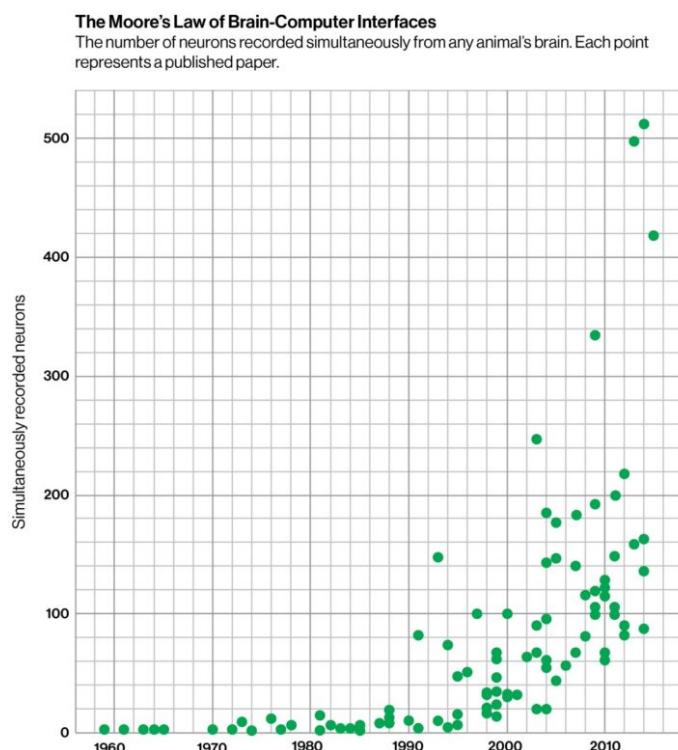
The Neuralink team threw out the number "one million simultaneously recorded neurons" when talking about an interface that could really change the world. I've also heard 100,000 as a number that would allow for the creation of a wide range of incredibly useful BMIs with a variety of applications.

Early computers had a similar problem. Primitive transistors took up a lot of space and didn't scale easily. Then in 1959 came the integrated circuit—the computer chip. Now there was a way to scale the number of transistors in a computer, and Moore's Law—the concept that the number of transistors that can fit onto a computer chip doubles every 18 months—was born.

Until the 90s, electrodes for BMIs were all made by hand. Then we started figuring out how to manufacture those little 100-electrode [multielectrode arrays](#) using conventional semiconductor technologies. Neuralink co-founder Ben Rapoport believes that "the move from hand manufacturing to Utah Array electrodes was the first hint that BMIs were entering a realm where Moore's Law could become relevant."

This is everything for the industry's potential. Our maximum today is a couple hundred electrodes able to measure about 500 neurons at once—which is either super far from a million or really close, depending on the kind of growth pattern we're in. If we add 500 more neurons to our maximum every 18 months, we'll get to a million in the year 2017. If we *double* our total every 18 months, like we do with computer transistors, we'll get to a million in the year 2034.

Currently, we seem to be somewhere in between. Ian Stevenson and Konrad Kording published a [paper](#) that looked at the maximum number of neurons that could be simultaneously recorded at various points throughout the last 50 years (in *any* animal), and put the results on this graph:⁴⁵



Source: Ian H. Stevenson, UConn

MIT Technology Review

Sometimes called Stevenson's Law, this research suggests that the number of neurons we can simultaneously record seems to consistently double every 7.4 years. If that rate continues, it'll take us till the end of this century to reach a million, and until 2225 to record every neuron in the brain and get our totally complete wizard hat.

Whatever the equivalent of the integrated circuit is for BMIs isn't here yet, because 7.4 years is too big a number to start a revolution. The breakthrough here isn't the device that can record a million neurons—it's the paradigm shift that makes the future of that graph look more like Moore's Law and less like Stevenson's Law. Once that happens, a million neurons will follow.

Major Hurdle 2: Implantation

BMIs won't sweep the world as long as you need to go in for skull-opening surgery to get involved.

This is a major topic at Neuralink. I think the word "non-invasive" or "non-invasively" came out of someone's mouth like 42 times in my discussions with the team.

On top of being both a major barrier to entry and a major safety issue, invasive brain surgery is expensive and in limited supply. Elon talked about an eventual BMI implantation process that could be automated: "The machine to accomplish this would need to be something like Lasik, an automated process—because otherwise you just get constrained by the limited number of neural surgeons, and the costs are very high. You'd need a Lasik-like machine ultimately to be able to do this at scale."

Making BMIs high-bandwidth alone would be a huge deal, as would developing a way to non-invasively implant devices. But doing *both* would start a revolution.

Other hurdles

Today's BMI patients have a wire coming out of their head. In the future, that certainly won't fly. Neuralink plans to work on devices that will be wireless. But that brings a lot of new challenges with it. You'll now need your device to be able to send and receive a lot of data wirelessly. Which means the implant also has to take care of things like signal amplification, analog-to-digital conversion, and data compression on its own. Oh and it needs to be powered inductively.

Another big one—biocompatibility. Delicate electronics tend to not do well inside a jello ball. And the human body tends to not like having foreign objects in it. But the brain interfaces of the future are intended to last forever without any problems. This means that the device will likely need to be hermetically sealed and robust enough to survive decades of the oozing and shifting of the neurons around it. And the brain—which treats today's devices like invaders and eventually covers them in scar tissue—will need to somehow be tricked into thinking the device is just a normal brain part doing its thing.²⁹

Then there's the space issue. Where exactly are you gonna put your device that can interface with a million neurons in a skull that's already dealing with making space for 100 billion neurons? A million electrodes using today's multielectrode arrays would be the size of a baseball. So further miniaturization is another dramatic innovation to add to the list.

There's also the fact that today's electrodes are mostly optimized for simple electrical recording or simple electrical stimulation. If we really want an effective brain interface, we'll need something other than single-function, stiff electrodes—something with the mechanical complexity of neural circuits, that can both record and stimulate, and that can interact with neurons chemically and mechanically as well as electrically.

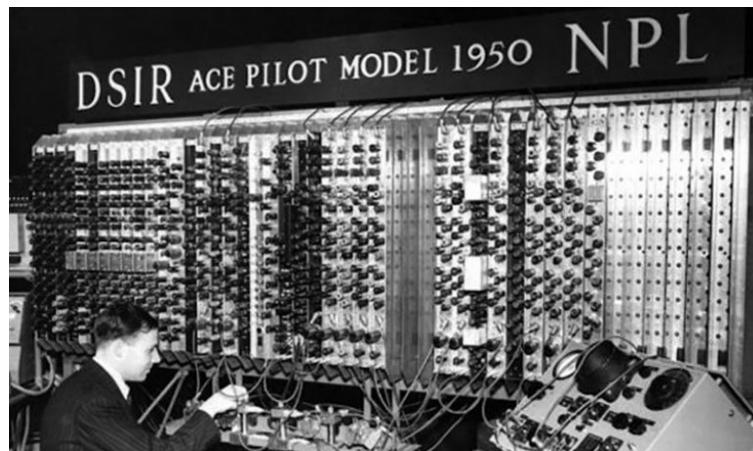
And just say all of this comes together perfectly—a high-bandwidth, long-lasting, biocompatible, bidirectional communicative, non-invasively-implanted device. Now we can speak back and forth with a million neurons at once! Except this little thing where we actually don't know how to talk to neurons. It's complicated enough to decode the static-like firings of 100 neurons, but all we're really doing is learning what a set of specific firings corresponds to and matching them up to simple commands. That won't work with millions of signals. It's like how Google Translate essentially uses two dictionaries to swap words from one dictionary to another—which is very different than *understanding language*. We'll need a pretty big leap in machine learning before a computer will be able to actually know a language, and we'll need just as big a leap for machines to understand the language of the brain—because humans certainly won't be learning to decipher the code of millions of simultaneously chattering neurons.

How easy does colonizing Mars seem right now.

But I bet the telephone and the car and the moon landing would have seemed like insurmountable technological challenges to people a few decades earlier. Just like I bet this—



—would have seemed utterly inconceivable to people at the time of this:

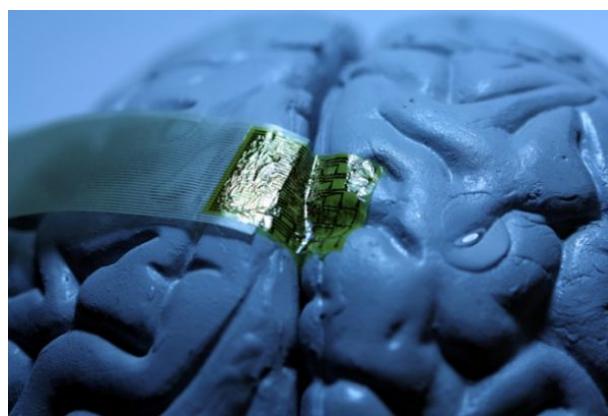


And yet, there it is in your pocket. If there's one thing we should learn from the past, it's that there will *always* be ubiquitous technology of the future that's inconceivable to people of the past. We don't know *which* technologies that seem positively impossible to us will turn out to be ubiquitous later in our lives —but there will be some. People always underestimate the Human Colossus.

If everyone you know in 40 years has electronics in their skull, it'll be because a paradigm shift took place that caused a fundamental shift in this industry. That shift is what the Neuralink team will try to figure out. Other teams are working on it too, and some cool ideas are being developed:

Current BMI innovations

A team at the University of Illinois is developing an interface made of silk:⁴⁶

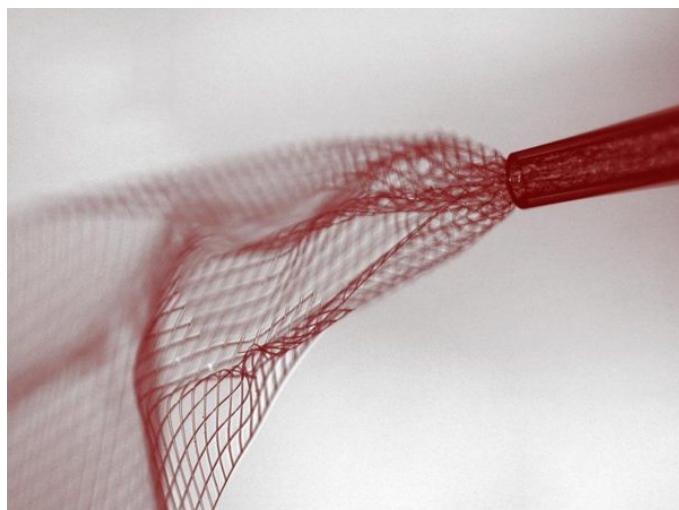


Silk can be rolled up into a thin bundle and inserted into the brain relatively non-invasively. There, it would theoretically spread out around the brain and melt into the contours like shrink wrap. On the silk would be flexible silicon transistor arrays.

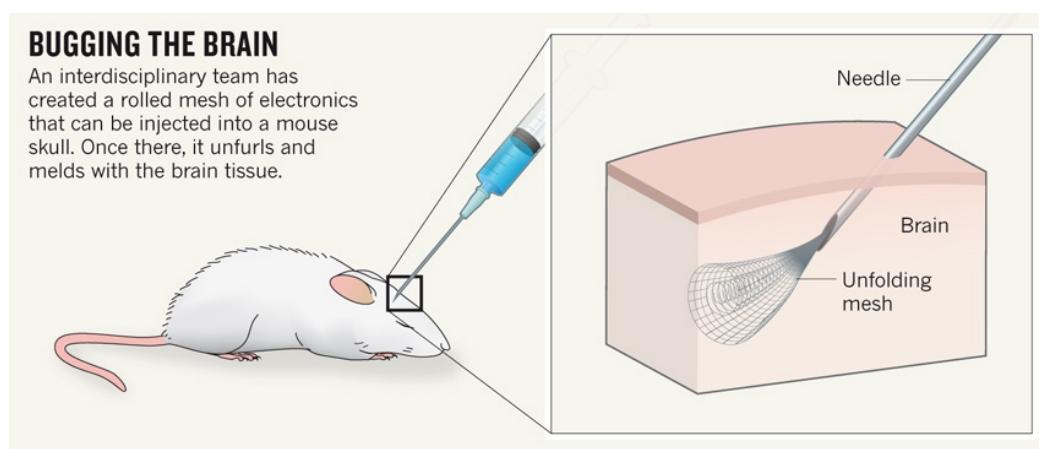
In his [TEDx Talk](#), Hong Yeo demonstrated an electrode array printed on his skin, like a temporary tattoo, and [researchers say](#) this kind of technique could potentially be used on the brain: ⁴⁷



Another group is working on a kind of nano-scale, electrode-lined **neural mesh** so tiny it can be injected into the brain with a syringe: ⁴⁸



For scale—that red tube on the right is the tip of a syringe. [Extreme Tech](#) has a nice graphic illustrating the concept:



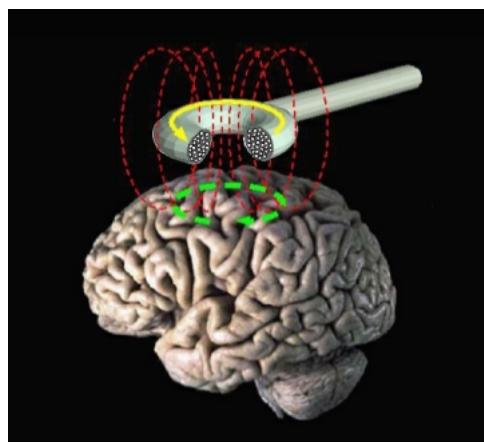
Other non-invasive techniques involve going in through veins and arteries. Elon mentioned this: "The least invasive way would be something that comes in like a hard stent like through a femoral artery and ultimately unfolds in the vascular system to interface with the neurons. Neurons use a lot of energy, so there's basically a road network to every neuron."

DARPA, the technology innovation arm of the US military, ³⁰ through their recently funded [BRAIN program](#), is working on tiny, "closed-loop" neural implants that could replace medication. ⁴⁹



A second DARPA project aims to fit a million electrodes into a device the size of two nickels stacked.

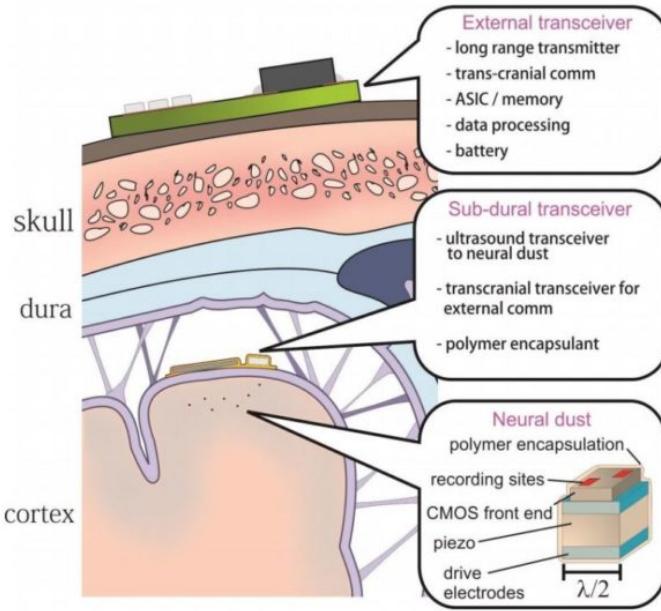
Another idea being worked on is transcranial magnetic stimulation (TMS), in which a magnetic coil outside the head can create electrical pulses inside the brain.⁵⁰



The pulses can stimulate targeted neuron areas, providing a type of deep brain stimulation that's totally non-invasive.

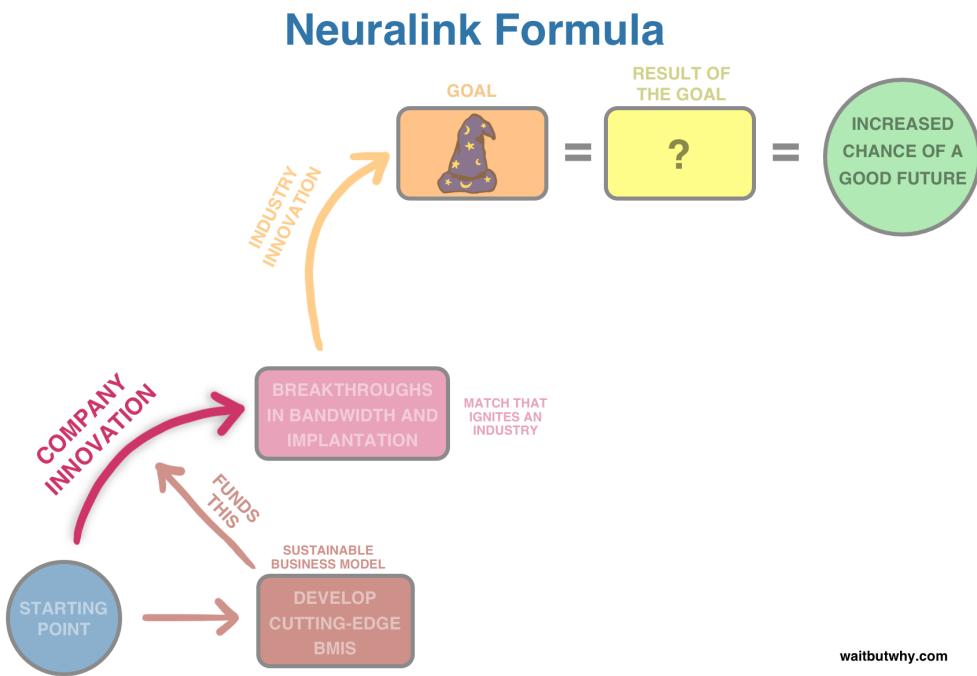
One of Neuralink's co-founders, DJ Seo, led an effort to design an even cooler interface called "[neural dust](#)." Neural dust refers to tiny, 100µm silicon sensors (about the same as the width of a hair) that would be sprinkled through the cortex. Right nearby, above the pia, would be a 3mm-sized device that could communicate with the dust sensors via ultrasound.

This is another example of the innovation benefits that come from an interdisciplinary team. DJ explained to me that "there are technologies that are not really thought about in this domain, but we can bring in some principles of their work." He says that neural dust is inspired both by microchip technology and RFID (the thing that allows hotel key cards to communicate with the door lock without making physical contact) principles. And you can easily see the multi-field influence in how it works:⁵¹



Others are working on even more out-there ideas, like **optogenetics** (where you inject a virus that attaches to a brain cell, causing it to thereafter be stimulated by light) or even using **carbon nanotubes** —a million of which could be bundled together and sent to the brain via the bloodstream.

These people are all working on this arrow:



It's a relatively small group right now, but when the breakthrough spark happens, that'll quickly change. Developments will begin to happen rapidly. Brain interface bandwidth will get better and better as the procedures to implant them become simpler and cheaper. Public interest will pick up. And when public interest picks up, the Human Colossus notices an opportunity—and then the rate of development skyrockets. Just like the breakthroughs in computer hardware caused the software industry to explode, major industries will pop up working on cutting-edge machines and intelligent apps to be used in conjunction with brain interfaces, and you'll tell some little kid in 2052 all about how when you grew up, no one could do any of the things she can do with her brain, and she'll be bored.

I tried to get the Neuralink team to talk about 2052 with me. I wanted to know what life was going to be like once this all became a thing. I wanted to know what went in the [Pilot ACE : iPhone 7 :: Early BMIs : ____] blank. But it wasn't easy—this was a team built specifically because of their focus on concrete results, not hype, and I was doing the equivalent of talking to people in the late 1700s who were feverishly trying to create a breakthrough steam engine and prodding them about when they thought there would be airplanes.

But I'd keep pulling teeth until they'd finally talk about their thoughts on the far future to get my hand off their tooth. I also focused a large portion of my talks with Elon on the far future possibilities and had other helpful discussions with Moran Cerf, a neuroscientist friend of mine who [works on BMIs](#) and thinks a lot about the long-term outlook. Finally, one reluctant-to-talk-about-his-predictions Neuralink team member told me that of course, he and his colleagues *were* dreamers—otherwise they wouldn't be doing what they're doing—and that many of them were inspired to get into this industry by science fiction. He recommended I talk to Ramez Naam, writer of the popular *Nexus Trilogy*, a series all about the future of BMIs, and also someone with a hard tech background that includes 19 software-related patents. So I had a chat with Ramez to round out the picture and ask him the 435 remaining questions I had about everything.

And I came out of all of it utterly blown away. I wrote [once](#) about how I think if you went back to 1750—a time when there was no electricity or motorized vehicles or telecommunication—and retrieved, say, George Washington, and brought him to today and showed him our world, he'd be so shocked by everything that he'd die. You'd have killed George Washington and messed everything up. Which got me thinking about the concept of how many years one would need to go into the future such that the ensuing shock from the level of progress would kill you. I called it a Die Progress Unit, or DPU.

Ever since the Human Colossus was born, our world has had a weird property to it—it gets more *magical* as time goes on. That's why DPUs are a thing. And because advancement begets more rapid advancement, the trend is that as time passes, the DPUs get *shorter*. For George Washington, a DPU was a couple hundred years, which is outrageously short in the scheme of human history. But we now live in a time where things are moving *so* fast that we might experience one or even multiple DPUs *in our lifetime*. The amount that changed between 1750 and 2017 might happen again between now and another time when you're still alive. This is a *ridiculous* time to be alive—it's just hard for us to notice because we live life so zoomed in.

Anyway, I think about DPUs a lot and I always wonder what it would feel like to go forward in a time machine and experience what George would experience coming here. What kind of future could blow my mind so hard that it would kill me? We can talk about things like AI and gene editing—and I have no doubt that progress in those areas could make me die of shock—but it's always, "Who knows what it'll be like!" Never a descriptive picture.

I think I might finally have a descriptive picture of a piece of our shocking future. Let me paint it for you.

Part 5: The Wizard Era



The budding industry of brain-machine interfaces is the seed of a revolution that will change just about everything. But in many ways, the brain-interface future isn't really a new *thing* that's happening. If you take a step back, it looks more like the next big chapter in a trend that's been going on for a long time. Language took forever to turn into writing, which then took forever to turn into printing, and that's where things were when George Washington was around. Then came electricity and the pace picked up. Telephone. Radio. Television. Computers. And just like that, everyone's homes became magical. Then phones became cordless. Then mobile. Computers went from being devices for work and games to windows into a digital world we all became a part of. Then phones and computers merged into an everything device that brought the magic out of our homes and put it into our hands. And on our wrists. We're now in the early stages of a virtual and augmented reality revolution that will wrap the magic around our eyes and ears and bring our whole being into the digital world.

You don't need to be a futurist to see where this is going.

Magic has worked its way from industrial facilities to our homes to our hands and soon it'll be around our heads. And then it'll take the next natural step. The magic is heading into our brains.

It will happen by way of a “whole-brain interface,” or what I’ve been calling a wizard hat—a brain interface so complete, so smooth, so biocompatible, and so high-bandwidth that it feels as much a part of you as your cortex and limbic system. A whole-brain interface would give your brain the ability to communicate wirelessly with the cloud, with computers, and with the brains of anyone with a similar interface in their head. This flow of information between your brain and the outside world would be so effortless, it would feel similar to the thinking that goes on in your head today. And though we’ve used the term brain-machine interface so far, I kind of think of a BMI as a specific brain interface to be used for a specific purpose, and the term doesn’t quite capture the everything-of-everything concept of the whole-brain interface. So I’ll call that a wizard hat instead.

Now, to fully absorb the implications of having a wizard hat installed in your head and what that would change about you, you’ll need to wrap your head around (no pun intended) two things:

1) The intensely mind-bending idea

2) The super ridiculously intensely mind-bending idea

We’ll tackle #1 in this section and save #2 for the last section after you’ve had time to absorb #1.

Elon calls the whole-brain interface and its many capabilities a “digital tertiary layer,” a term that has two levels of meaning that correspond to our two mind-bending ideas above.

The first meaning gets at the idea of physical brain parts. We discussed three layers of brain parts—the brain stem (run by the frog), the limbic system (run by the monkey), and the cortex (run by the rational thinker). We were being thorough, but for the rest of this post, we’re going to leave the frog out of the discussion, since he’s entirely functional and lives mostly behind the scenes.

When Elon refers to a “digital tertiary layer,” he’s considering our existing brain having two layers—our animal limbic system (which could be called our primary layer) and our advanced cortex (which could be called our secondary layer). The wizard hat interface, then, would be our tertiary layer—a new physical brain part to complement the other two.

If thinking about this concept is giving you the willies, Elon has news for you:

We already have a digital tertiary layer in a sense, in that you have your computer or your phone or your applications. You can ask a question via Google and get an answer instantly. You can access any book or any music. With a spreadsheet, you can do incredible calculations. If you had an Empire State building filled with people—even if they had calculators, let alone if they had to do it with a pencil and paper—one person with a laptop could outdo the Empire State Building filled with people with calculators. You can video chat with someone in freaking Timbuktu for free. This would've gotten you burnt for witchcraft in the old days. You can record as much video with sound as you want, take a zillion pictures, have them tagged with who they are and when it took place. You can broadcast communications through social media to millions of people simultaneously for free. These are incredible superpowers that the President of the United States didn't have twenty years ago.

The thing that people, I think, don't appreciate right now is that they are already a cyborg. You're already a different creature than you would have been twenty years ago, or even ten years ago. You're already a different creature. You can see this when they do surveys of like, “how long do you want to be away from your phone?” and—particularly if you're a teenager or in your 20s—even a day hurts. If you leave your phone behind, it's like missing limb syndrome. I think people—they're already kind of merged with their phone and their laptop and their applications and everything.

This is a hard point to really absorb, because we don’t *feel* like cyborgs. We feel like humans who use devices to do things. But think about your digital self—you when you’re interacting with someone on the internet or over FaceTime or when you’re in a YouTube video. Digital you is fully you—as much as in-person you is you—right? The only difference is that you’re not there in person—you’re using magic powers to send yourself to somewhere far away, at light speed, through wires and satellites and electromagnetic waves. The difference is the medium.

Before language, there wasn’t a good way to get a thought from your brain into my brain. Then early humans invented the technology of language, transforming vocal cords and ears into the world’s first communication devices and air as the first communication medium. We use these devices every time we talk to each other in person. It goes:



Then we built upon that with another leap, inventing a second layer of devices, with its own medium, allowing us to talk long distance:

YOUR BRAIN > YOUR VOCAL CORDS > AIR > YOUR PHONE > TELECOMMUNICATIONS NETWORK > MY PHONE > AIR > MY EAR > MY BRAIN

Or maybe:

YOUR BRAIN > YOUR THUMBS > YOUR PHONE > TELECOMMUNICATIONS NETWORK > MY PHONE > AIR > MY EYES > MY BRAIN

In that sense, your phone is as much “you” as your vocal cords or your ears or your eyes. All of these things are simply tools to move thoughts from brain to brain—so who cares if the tool is held in your hand, your throat, or your eye sockets? The digital age has made us a dual entity—a physical creature who interacts with its physical environment using its biological parts and a digital creature whose digital devices—whose digital *parts*—allow it to interact with the digital world.

But because we don’t think of it like that, we’d consider someone with a phone in their head or throat a cyborg and someone else with a phone in their hand, pressed up against their head, not a cyborg. Elon’s point is that the thing that makes a cyborg a cyborg is their capabilities—not from which side of the skull those capabilities are generated.

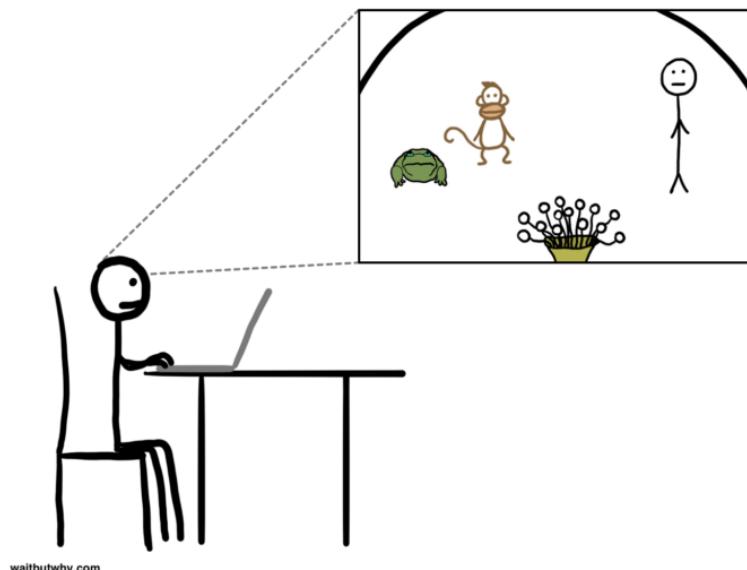
We’re already a cyborg, we already have superpowers, and we already spend a huge part of our lives in the digital world. And when you think of it like that, you realize how obvious it is to want to upgrade the medium that connects us to that world. This is the change Elon believes is actually happening when the magic goes into our brains:

You’re already digitally superhuman. The thing that would change is the interface—having a high-bandwidth interface to your digital enhancements. The thing is that today, the interface all necks down to this tiny straw, which is, particularly in terms of output, it’s like poking things with your meat sticks, or using words—either speaking or tapping things with fingers. And in fact, output has gone backwards. It used to be, in your most frequent form, output would be ten-finger typing. Now, it’s like, two-thumb typing. That’s crazy slow communication. We should be able to improve that by many orders of magnitude with a direct neural interface.

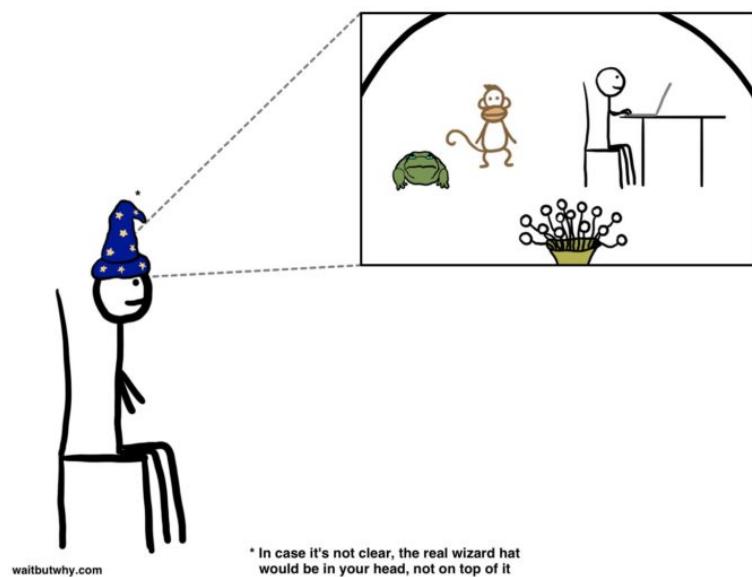
In other words, putting our technology into our brains isn’t about whether it’s good or bad to become cyborgs. It’s that we *are* cyborgs and we will *continue* to be cyborgs—so it probably makes sense to upgrade ourselves from primitive, low-bandwidth cyborgs to modern, high-bandwidth cyborgs.

A whole-brain interface is that upgrade. It changes us from creatures whose primary and secondary layers live inside their heads and whose tertiary layer lives in their pocket, in their hand, or on their desk

—



—to creatures whose three layers all live together.

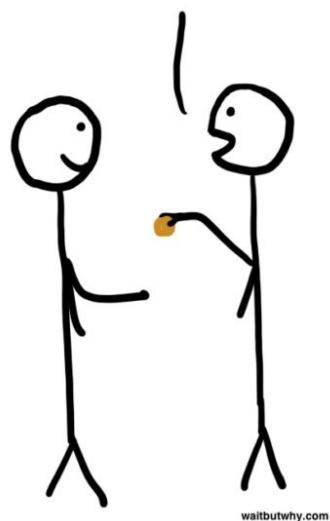


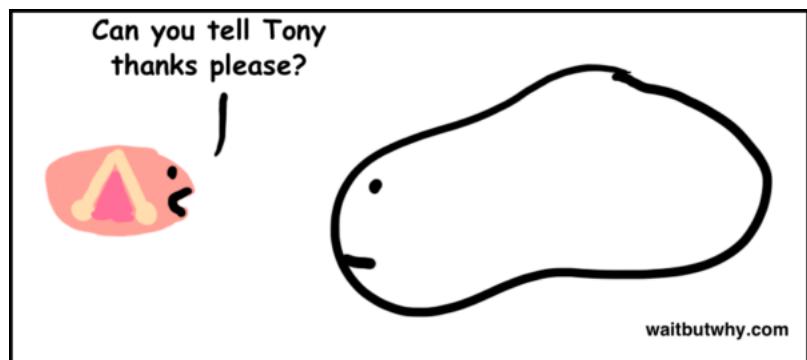
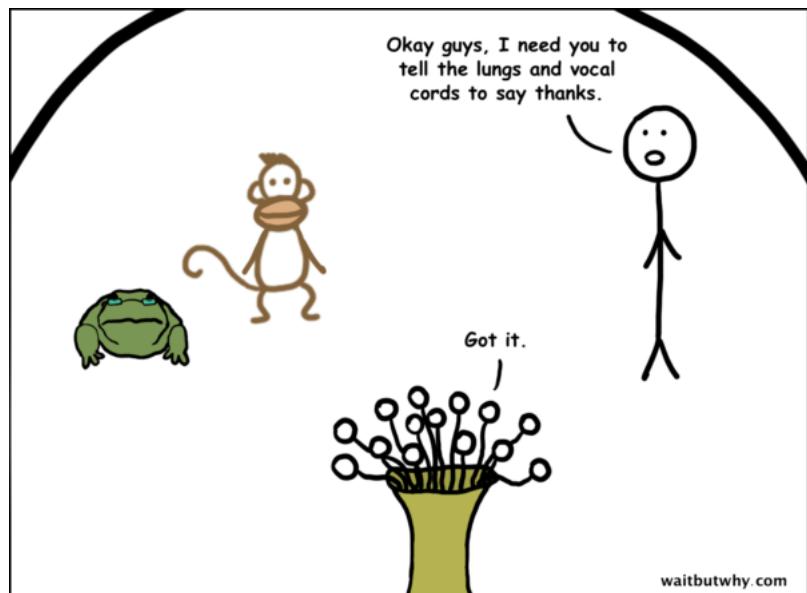
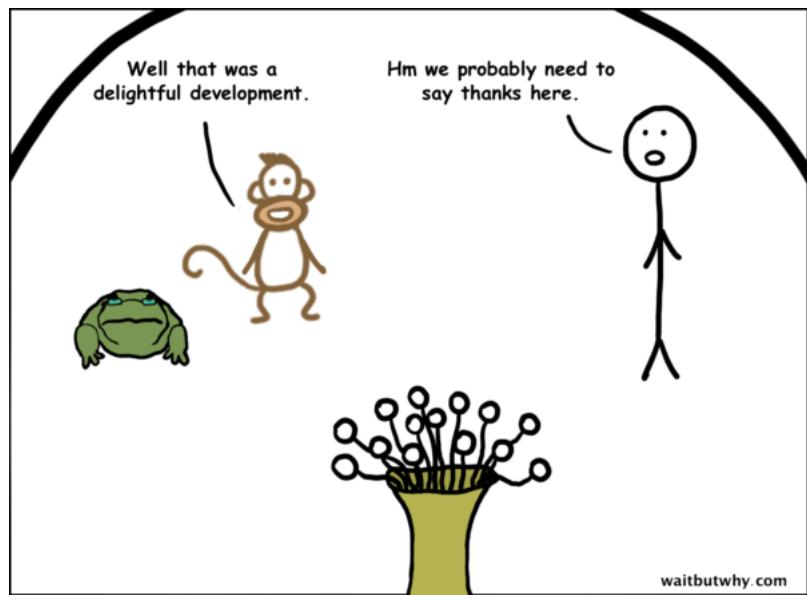
Your life is full of devices, including the one you're currently using to read this. A wizard hat makes your *brain* into the device, allowing your thoughts to go straight from your head into the digital world.

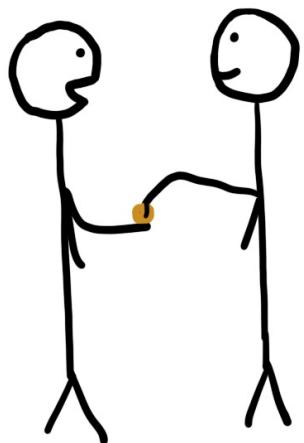
Which doesn't only revolutionize human-computer communication.

Right now humans communicate with each other like this:

Here have a caramel.



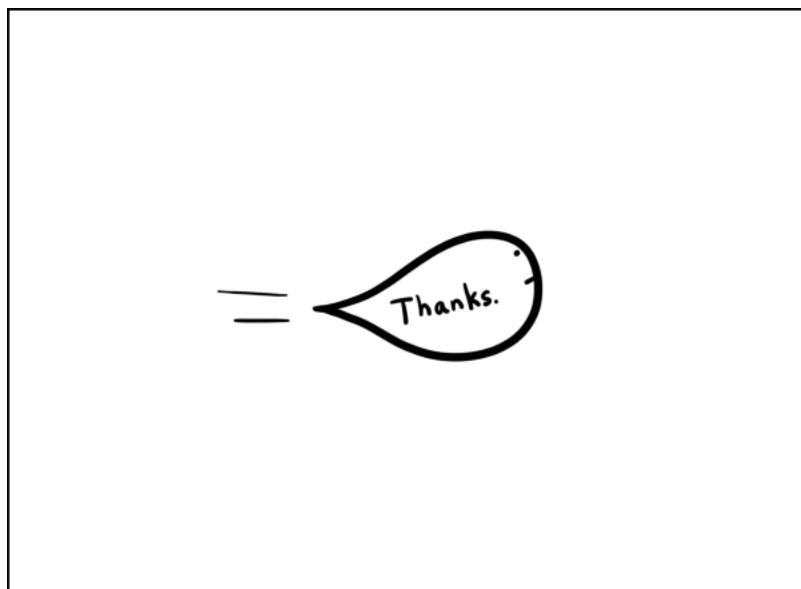




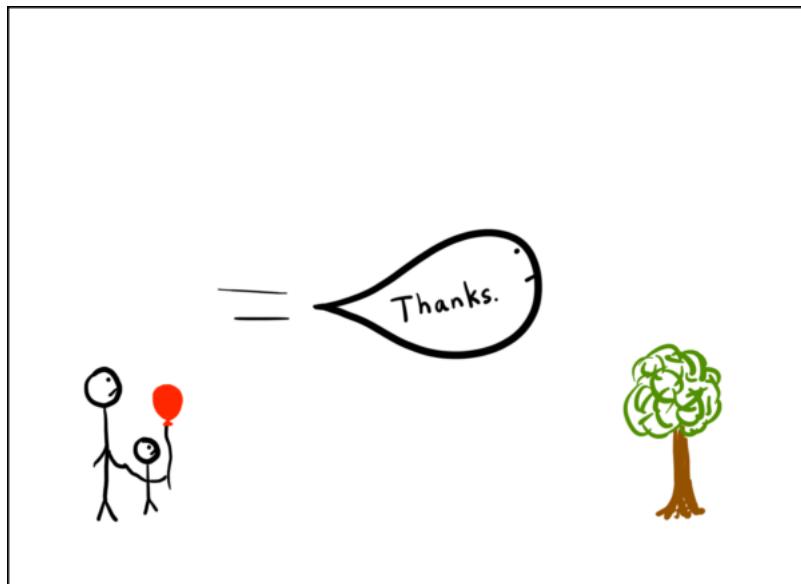
waitbutwhy.com



waitbutwhy.com

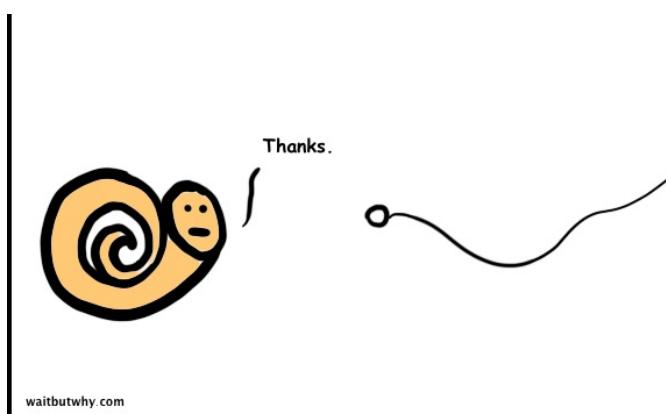
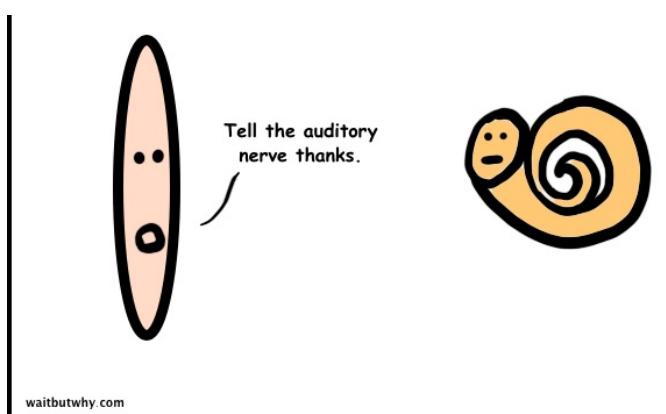
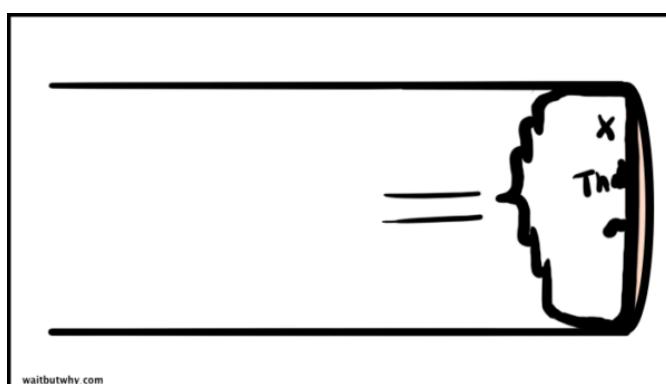
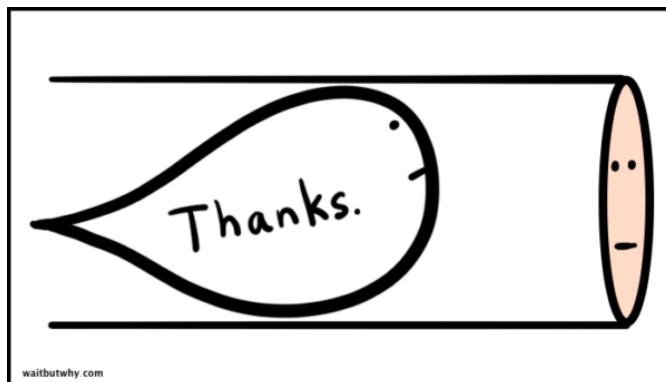
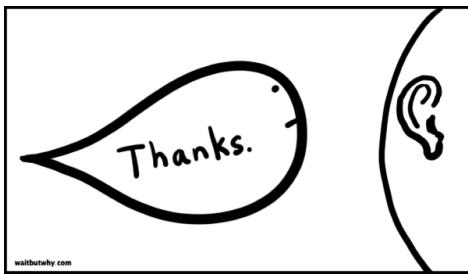


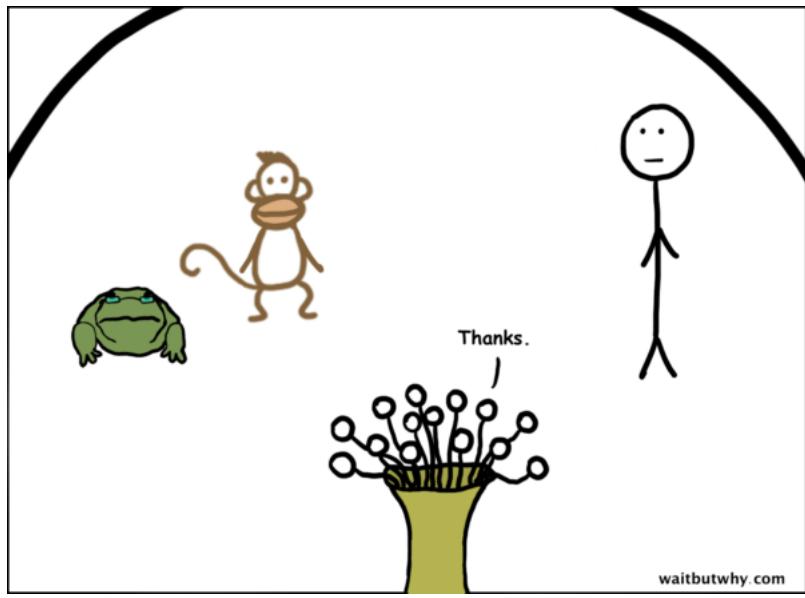
waitbutwhy.com



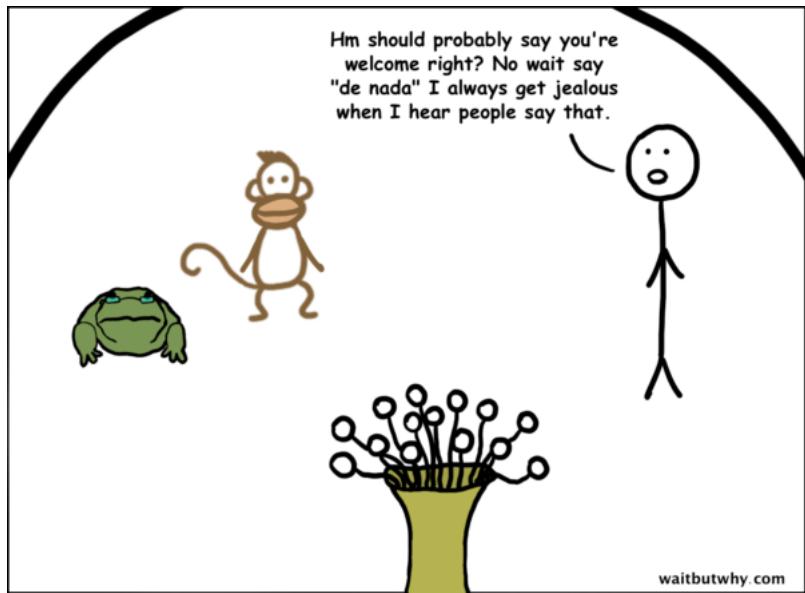
waitbutwhy.com







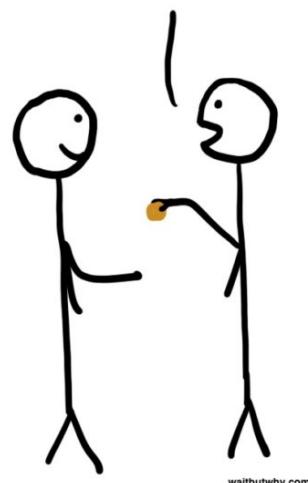
waitbutwhy.com



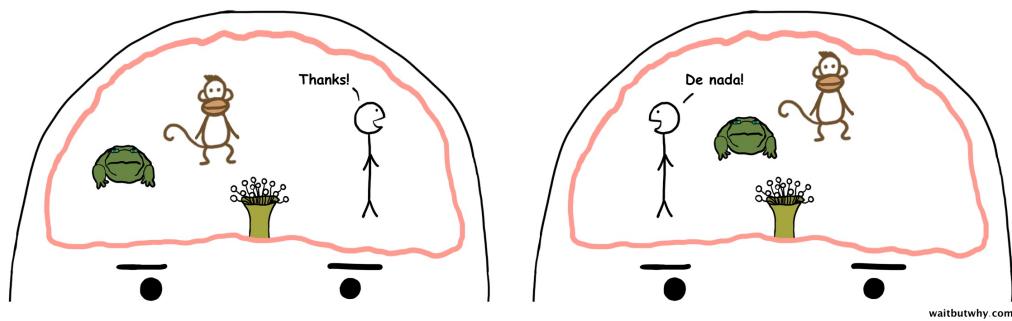
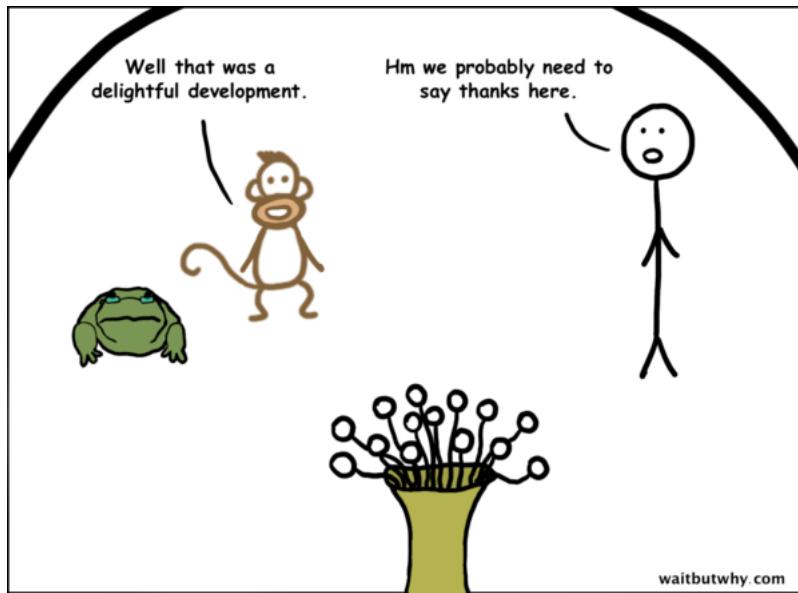
waitbutwhy.com

And that's how it's been ever since we could communicate. But in a wizard hat world, it would look more like this:

Here have a caramel.

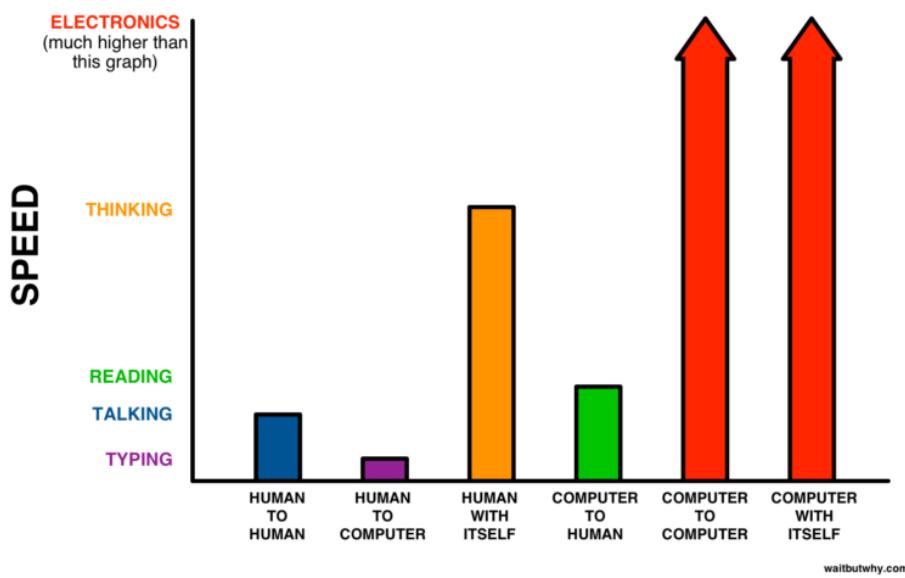


waitbutwhy.com



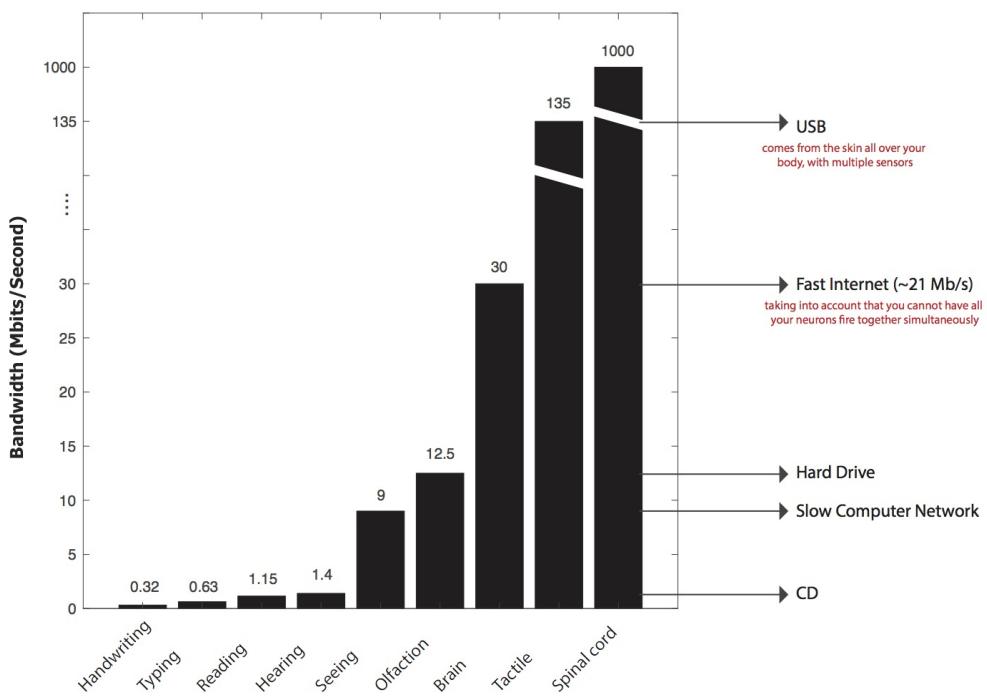
Elon always emphasizes bandwidth when he talks about Neuralink's wizard hat goals. Interface bandwidth allows incoming images to be HD, incoming sound to be hi-fi, and motor movement commands to be tightly controlled—but it's also a huge factor in communication. If information were a milkshake, bandwidth would be the width of the straw. Today, the bandwidth-of-communication graph looks something like this:

Communication Speed Today



So computers can suck up the milkshake through a giant pipe, a human thinking would be using a large, pleasant-to-use straw, while language would be a frustratingly tiny coffee stirrer straw and typing (let alone texting) would be like trying to drink a milkshake through a syringe needle—you might be able to get a drop out once a minute.

Moran Cerf has gathered data on the actual bandwidth of different parts of the nervous system and on this graph, he compares them to equivalent bandwidths in the computer world:



You can see here on Moran's graph that the disparity in bandwidth between the ways we communicate and our thinking (which is at 30 bits/second on this graph) is even starker than my graph above depicts.

But making our brains the device cuts out those tiny straws, turning all of these:

YOUR BRAIN > YOUR VOCAL CORDS > AIR > MY EAR > MY BRAIN

YOUR BRAIN > YOUR VOCAL CORDS > AIR > YOUR PHONE > TELECOMMUNICATIONS NETWORK > MY PHONE > AIR > MY EAR > MY BRAIN

YOUR BRAIN > YOUR THUMBS > YOUR PHONE > TELECOMMUNICATIONS NETWORK > MY PHONE > AIR > MY EYES > MY BRAIN

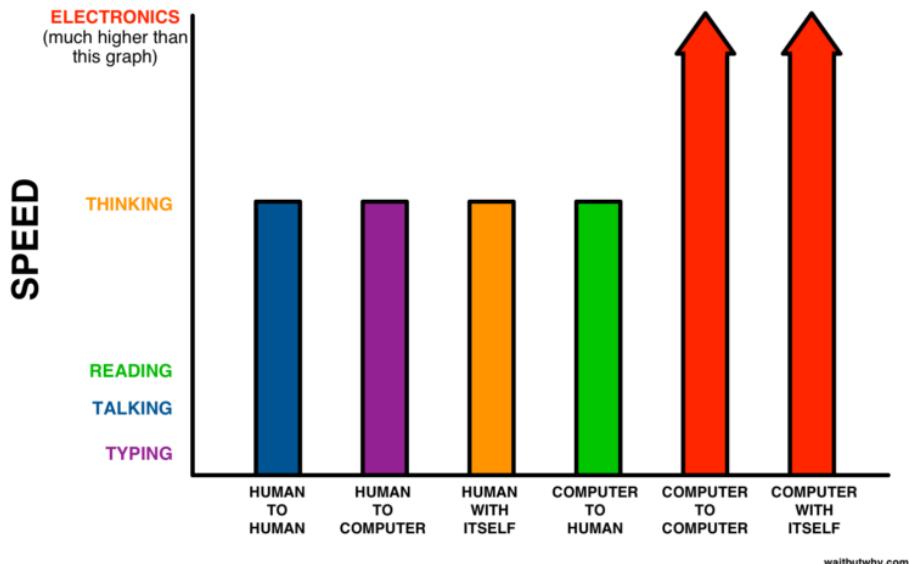
waitbutwhy.com

To this:

YOUR BRAIN > MY BRAIN

Which preserves all the meaning with none of the fuss—and changes the graph to this:

Communication Speed in the Digital Brain Era



We'd still be using straws, but *far* bigger, more effective ones.

But it's not just about the speed of communication. As Elon points out, it's about the *nuance* and accuracy of communication as well:

There are a bunch of concepts in your head that then your brain has to try to compress into this incredibly low data rate called speech or typing. That's what language is—your brain has executed a compression algorithm on thought, on concept transfer. And then it's got to listen as well, and decompress what's coming at it. And this is very lossy as well. So, then when you're doing the decompression on those, trying to understand, you're simultaneously trying to model the other person's mind state to understand where they're coming from, to recombine in your head what concepts they have in their head that they're trying to communicate to you. ... If you have two brain interfaces, you could actually do an uncompressed direct conceptual communication with another person.

This makes sense—nuance is like a *high-resolution thought*, which makes the file simply too big to transfer quickly through a coffee straw. The coffee straw gives you two bad options when it comes to nuance: take a *lot* of time saying a lot of words to really depict the nuanced thought or imagery you want to convey to me, or save time by using succinct language—but inevitably fail to transfer over the nuance. Compounding the effect is the fact that language itself is a low-resolution medium. A word is simply an *approximation* of a thought—buckets that a whole category of similar-but-distinct thoughts can all be shoved into. If I watch a horror movie and want to describe it to you in words, I'm stuck with a few simple low-res buckets—"scary" or "creepy" or "chilling" or "intense." My actual impression of that movie is very specific and not exactly like any other movie I've seen—but the crude tools of language force my brain to "round to the nearest bucket" and choose the word that most closely resembles my actual impression, and that's the information you'll receive from me. You won't receive the thought—you'll receive the *bucket*—and now you'll have to guess which of the *many* nuanced impressions that all approximate to that bucket is the most similar to my impression of the movie. You'll decompress my description—"scary as shit"—into a high-res, nuanced thought that *you* associate with "scary as shit," which will inevitably be based on your own experience watching other horror movies, and your own personality. The end result is that a lot has been lost in translation—which is exactly what you'd expect when you try to transfer a high-res file over a low-bandwidth medium, quickly, using low-res tools. That's why Elon calls language data transfer "lossy."

We do the best we can with these limitations—and over time, we've supplemented language with slightly higher-resolution formats like video to better convey nuanced imagery, or music to better convey nuanced emotion. But compared to the richness and uniqueness of the ideas in our heads, and the large-bandwidth straw our internal *thoughts* flow through, all human-to-human communication is very lossy.

Thinking about the phenomenon of communication as what it is—brains trying to share things with each other—you see the history of communication not as this:

Isolated Brains



Language



Writing



Telephones



Internet



Mobile / Wearable Devices



Digitally Communicating Brains

waitbutwhy.com

As much as this:

Isolated Brains



A bunch of brain communication hacks using different kinds of middlemen



Digitally Communicating Brains

waitbutwhy.com

Or it could be put this way:

Era of Isolated Brains



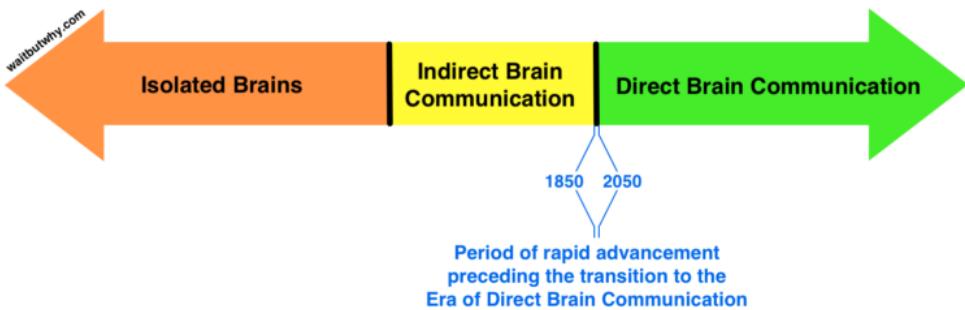
Era of Indirect Brain Communication



Era of Direct Brain Communication

waitbutwhy.com

It really may be that the second major era of communication—the 100,000-year Era of Indirect Communication—is in its very last moments. If we zoom out on the timeline, it's possible the entire last 150 years, during which we've suddenly been *rapidly* improving our communication media, will look to far-future humans like one concept: the transition from Era 2 to Era 3. We might be living *on* the line that divides timeline sections.



And because indirect communication requires third-party body parts or digital parts, the end of Era 2 may be looked back upon as the era of physical devices. In an era where your brain is the device, there will be no need to carry anything around. You'll have your body and, if you want, clothes—and that's it.

When Elon thinks about wizard hats, this is usually the stuff he's thinking about—communication bandwidth and resolution. And we'll explore why in Part 6 of this post.

First, let's dig into the mind-boggling concept of your brain becoming a device and talk about what a wizard hat world might be like.

One thing to keep in mind as we think about all of this is that none of it will take you by surprise. You won't go from having nothing in your brain to a digital tertiary layer in your head, just like people didn't go from the Apple IIGS to using Tinder overnight. The Wizard Era will come gradually, and by the time the shift actually begins to happen, we'll all be very used to the technology, and it'll seem normal.

Supporting this point is the fact the staircase up to the Wizard Era has already started, and you haven't even noticed. But there are thousands of people currently walking around with electrodes in their brain, like those with cochlear implants, retinal implants, and deep brain implants—all benefiting from early BMIs.

The next few steps on the staircase will continue to focus on restoring lost function in different parts of the body—the first people to have their lives transformed by digital brain technology will be the disabled. As specialized BMIs serve more and more forms of disability, the concept of brain implants will work its way in from the fringes and become something we're all used to—just like no one blinks an eye when you say your friend just got Lasik surgery or your grandmother just got a pacemaker installed.

Elon talks about some types of people early BMIs could help:

The first use of the technology will be to repair brain injuries as a result of stroke or cutting out a cancer lesion, where somebody's fundamentally lost a certain cognitive element. It could help with people who are quadriplegics or paraplegics by providing a neural shunt from the motor cortex down to where the muscles are activated. It can help with people who, as they get older, have memory problems and can't remember the names of their kids, through memory enhancement, which could allow them to function well to a much later time in life—the medically advantageous elements of this for dealing with mental disablement of one kind or another, which of course happens to all of us when we get old enough, are very significant.

As someone who lost a grandfather to dementia five years before losing him to death, I'm excited to hear this.

And as interface bandwidth improves, disabilities that hinder millions today will start to drop like flies. The concepts of complete blindness and deafness—whether centered in the sensory organs or in the brain³¹—are already on the way out. And with enough time, perfect vision or hearing will be restorable.

Prosthetic limbs—and eventually sleek, full-body exoskeletons underneath your clothes—will work so well, providing both outgoing motor functions and an incoming sense of touch, that paralysis or amputations will only have a minor long-term effect on people's lives.

In Alzheimer's patients, memories themselves are often not lost—only the bridge to those memories. Advanced BMIs could help restore that bridge or serve as a new one.

While this is happening, BMIs will begin to emerge that people *without* disabilities want. The very early adopters will probably be pretty rich. But so were the early cell phone adopters.⁵²



That's Gordon Gekko, and that 1983, two-pound cell phone cost almost \$9,000 in today's dollars. And now over half of living humans own a mobile phone—all of them far less shitty than Gordon Gekko's.

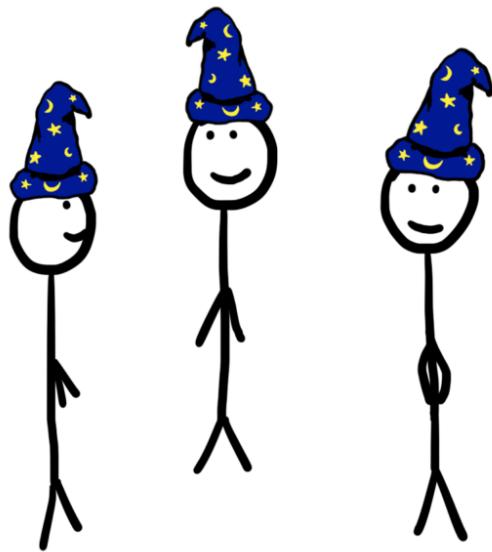
As mobile phones got cheaper, and better, they went from new and fancy and futuristic to ubiquitous. As we go down the same road with brain interfaces, things are going to get *really* cool.

Based on what I learned from my conversations with Elon, Ramez, and a dozen neuroscientists, let's look at what the world might look like in a few decades. The timeline is uncertain, including the order in which the below developments may become a reality. And, of course, some of the below predictions are sure to be way off the mark, just as there will be other developments in this field that won't be mentioned here because people today literally can't imagine them yet.

But some version of a lot of this stuff probably *will* happen, at *some* point, and a lot of it could be in your lifetime.

Looking at all the predictions I heard, they seemed to fall into two broad categories: communication capabilities and internal enhancements.

The Wizard Era: Communication



Motor communication

"Communication" in this section can mean human-to-human or human-to-computer. Motor communication is all about human-to-computer—the whole "motor cortex as remote control" thing from earlier, but now the unbelievably rad version.

Like many future categories of brain interface possibility, motor communication will start with restoration applications for the disabled, and as those development efforts continually advance the possibilities, the technology will begin to be used to create *augmentation* applications for the non-disabled as well. The same technologies that will allow a quadriplegic to use their thoughts as a remote control to move a bionic limb can let anyone use their thoughts as a remote control...to move anything. Well not *anything*—I'm not talking about telekinesis—anything built to be used with a brain remote. But in the Wizard Era, *lots* of things will be built that way.

Your car (or whatever people use for transportation at that point) will pull up to your house and your mind will open the car door. You'll walk up to the house and your mind will unlock and open the front door (all doors at that point will be built with sensors to receive motor cortex commands). You'll think about wanting coffee and the coffee maker will get that going. As you head to the fridge the door will open and after getting what you need it'll close as you walk away. When it's time for bed, you'll decide you want the heat turned down and the lights turned off, and those systems will feel you make that decision and adjust themselves.

None of this stuff will take any effort or thought—we'll all get very good at it and it'll feel as automatic and subconscious as moving your eyes to read this sentence does to you now.

People will play the piano with their thoughts. And do building construction. And steer vehicles. In fact, today, if you're driving somewhere and something jumps out in the road in front of you, what neuroscientists know is that your brain sees it and begins to react well before your consciousness knows what's going on or your arms move to steer out of the way. But when your brain is the one steering the car, you'll have swerved out of the way before you even realize what happened.

Thought communication

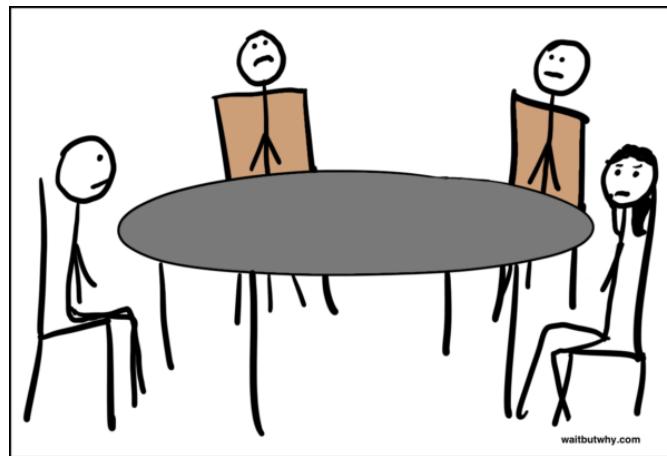
This is what we discussed up above—but you have to resist the natural instinct to equate a thought conversation with a normal language conversation where you simply hear each other's voices in your head. As we discussed, words are compressed approximations of uncompressed thoughts, so why would you ever bother with any of that, or deal with lossiness, if you didn't have to? When you watch a movie, your head is buzzing with thoughts—but do you have a compressed spoken word dialogue going on in your head? Probably not—you're just *thinking*. Thought conversations will be like *that*.

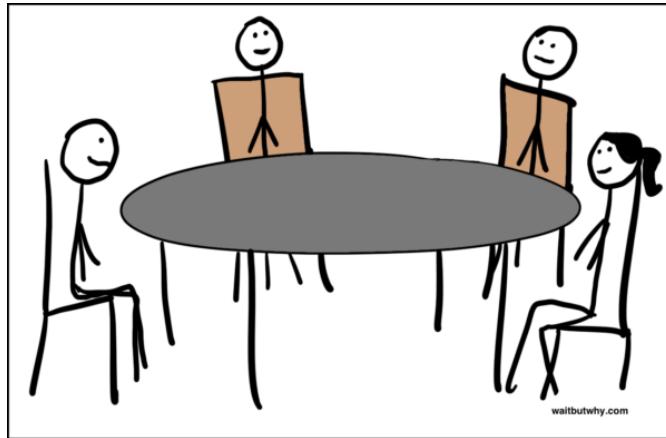
Elon says:

If I were to communicate a concept to you, you would essentially engage in consensual telepathy. You wouldn't need to verbalize unless you want to add a little flair to the conversation or something (laughs), but the conversation would be conceptual interaction on a level that's difficult to conceive of right now.

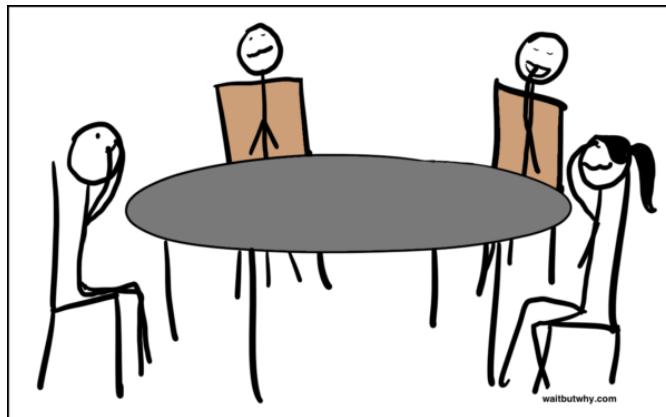
That's the thing—it's difficult to really understand what it would be like to *think* with someone. We've never been able to try. We communicate with ourselves through thought and with everyone else through symbolic representations of thought, and that's all we can imagine.

Even weirder is the concept of a *group* thinking together. This is what a group brainstorm could look like in the Wizard Era.

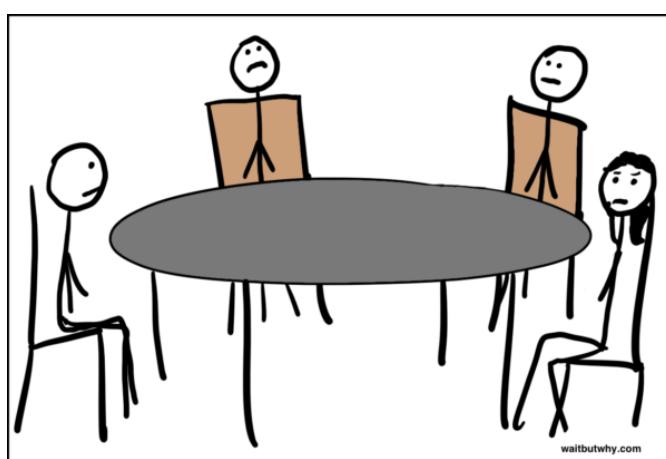




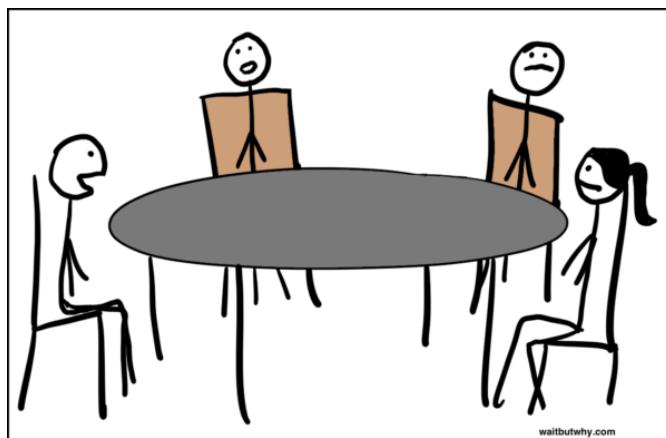
waitbutwhy.com



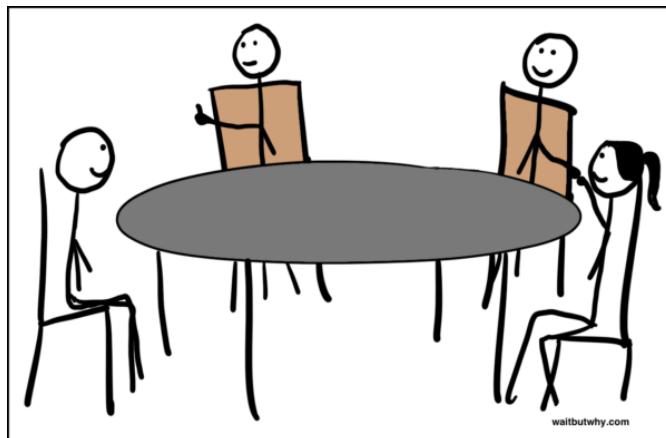
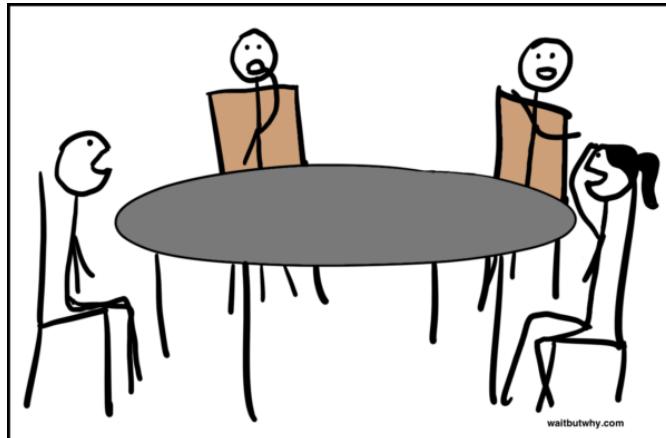
waitbutwhy.com



waitbutwhy.com



waitbutwhy.com



And of course, they wouldn't need to be in the same room. This group could have been in four different countries while this was happening—with no external devices in sight.

Ramez has [written about](#) the effect group thinking might have on the world:

That type of communication would have a huge impact on the pace of innovation, as scientists and engineers could work more fluidly together. And it's just as likely to have a transformative effect on the public sphere, in the same way that email, blogs, and Twitter have successively changed public discourse.

The idea of collaboration today is *supposed* to be two or more brains working together to come up with things none of them could have on their own. And a lot of the time, it works pretty well—but when you consider the “lost in transmission” phenomenon that happens with language, you realize how much more effective group thinking would be.

I asked Elon a question that pops into everyone’s mind when they first hear about thought communication:

“So, um, will everyone be able to know what I’m thinking?”

He assured me they would not. “People won’t be able to read your thoughts—you would have to will it. If you don’t will it, it doesn’t happen. Just like if you don’t will your mouth to talk, it doesn’t talk.” Phew.

You can also think with a computer. Not just to issue a command, but to actually brainstorm something with a computer. You and a computer could strategize something together. You could compose a piece of music together. Ramez talked about using a computer as an imagination collaborator: “You could imagine something, and the computer, which can better forward predict or analyze physical models, could fill in constraints—and that allows you to get feedback.”

One concern that comes up when people hear about thought communication in particular is a potential loss of individuality. Would this make us one great hive mind with each individual brain as just another bee? Almost across the board, the experts I talked to believed it would be the opposite. We could act as one in a collaboration when it served us, but technology has thus far enhanced human individuality. Think of how much easier it is for people today to express their individuality and customize life to themselves than it was 50 or 100 or 500 years ago. There’s no reason to believe that trend won’t continue with more progress.

Multimedia communication

Similar to thought communication, but imagine how much easier it would be to describe a dream you had or a piece of music stuck in your head or a memory you're thinking about if you could just beam the thing into someone's head, like showing them on your computer screen. Or as Elon said, "I could think of a bouquet of flowers and have a very clear picture in my head of what that is. It would take a lot of words for you to even have an approximation of what that bouquet of flowers looks like."

How much faster could a team of engineers or architects or designers plan out a new bridge or a new building or a new dress if they could beam the vision in their head onto a screen and others could adjust it with their minds, versus sketching things out—which not only takes far longer, but probably is inevitably lossy?

How many symphonies could Mozart have written if he had been able to think the music in his head onto the page? How many Mozarts are out there right now who never learned how to play instruments well enough to get their talent out?

I watched this [delightful animated short movie](#) the other day, and below the video the creator, Felix Colgrave, said the video took him two years. How much of that time was spent dreaming up the art versus painstakingly getting it from his head into the software? Maybe in a few decades, I'll be able to watch animation streaming live out of Felix's head.

Emotional communication

Emotions are the quintessential example of a concept that words are poorly-equipped to accurately describe. If ten people say, "I'm sad," it actually means ten different things. In the Wizard Era, we'll probably learn pretty quickly that the specific emotions people feel are as unique to people as their appearance or sense of humor.

This could work as communication—when one person communicates just what they're feeling, the other person would be able to access the feeling in their own emotional centers. Obvious implications for a future of heightened empathy. But emotional communication could also be used for things like entertainment, where a movie, say, could also project out to the audience—directly into their limbic systems—certain feelings it wants the audience to feel as they watch. This is already what the film score does—another hack—and now it could be done directly.

Sensory communication

This one is intense.

Right now, the only two microphones that can act as inputs for the "speaker" in your head—your auditory cortex—are your two ears. The only two cameras that can be hooked up to the projector in your head—your visual cortex—are your two eyes. The only sensory surface that you can feel is your skin. The only thing that lets you experience taste is your tongue.

But in the same way we can currently hook an implant, for example, into someone's cochlea—which connects a *different* mic to their auditory cortex—down the road we'll be able to let sensory input information stream into your wizard hat wirelessly, from anywhere, and channel right into your sensory cortices the same way your bodily sensory organs do today. In the future, sensory organs will be only one set of inputs into your senses—and compared to what our senses will have access to, not a very exciting one.

Now what about output?

Currently, the only speaker your ear inputs can play *out of* is your auditory cortex. Only *you* can see what your eye cameras capture and only you can feel what touches your skin—because only you have access to the particular cortices those inputs are wired to. With a wizard hat, it would be a breeze for your brain to beam those input signals out of your head.

So you'll have sensory input capabilities and sensory output capabilities—or both at the same time. This will open up all *kinds* of amazing possibilities.

Say you're on a beautiful hike and you want to show your husband the view. No problem—just think out to him to request a brain connection. When he accepts, connect your retina feed to his visual cortex. Now his vision is filled with exactly what your eyes see, as if he's there. He asks for the other senses to get the full picture, so you connect those too and now he hears the waterfall in the distance and feels the breeze and smells the trees and jumps when a bug lands on your arm. You two share the equivalent of a five-minute discussion about the scene—your favorite parts, which other places it reminds you of, etc. along with a shared story from his day—in a 30-second thought session. He says he has to get back to what he was working on, so he cuts off the sense connections except for vision, which he reduces to a little picture-in-picture window on the side of his visual field so he can check out more of the hike from time to time.

A surgeon could control a machine scalpel with her motor cortex instead of holding one in her hand, and she could receive sensory input from that scalpel so that it would feel like an 11th finger to her. So it would be as if one of her fingers *was* a scalpel and she could do the surgery without holding any tools, giving her much finer control over her incisions. An inexperienced surgeon performing a tough operation could bring a couple of her mentors into the scene as she operates to watch her work through her eyes and think instructions or advice to her. And if something goes really wrong, one of them could “take the wheel” and connect their motor cortex to her outputs to take control of her hands.

There would be no more need for screens of course—because you could just make a virtual screen appear in your visual cortex. Or jump into a VR movie with all your senses. Speaking of VR—Facebook, the maker of the Oculus Rift, is diving into this too. In an interview with Mark Zuckerberg about VR (for an upcoming post), the conversation at one point turned to BMIs. He said: “Touch gives you input and it’s a little bit of haptic feedback. Over the long term, it’s not clear that we won’t just like to have our hands in no controller, and maybe, instead of having buttons that we press, we would just *think* something.”

The ability to record sensory input means you can also record your memories, or share them—since a memory in the first place is just a not-so-accurate playback of previous sensory input. Or you could play them back as live experiences. In other words, that [Black Mirror episode](#) will probably actually happen.

An NBA player could send out a livestream invitation to his fans before a game, which would let them see and hear through his eyes and ears while he plays. Those who miss it could jump into the recording later.

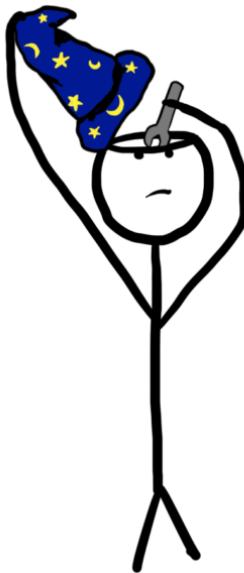
You could save a great sex experience in the cloud to enjoy again later—or, if you’re not too private a person, you could send it over to a friend to experience. (Needless to say, the porn industry will thrive in the digital brain world.)

Right now, you can go on YouTube and watch a first-hand account of almost anything, for free. This would have blown George Washington’s mind—but in the Wizard Era, you’ll be able to actually *experience* almost anything for free. The days of fancy experiences being limited to rich people will be long over.

Another idea, via the imagination of Moran Cerf: Maybe player brain injuries will drive the NFL to alter the rules so that the players’ biological bodies stay on the sidelines, while they play the game with an artificial body whose motor cortex they control and whose eyes and ears they see and hear through. I like this idea and think it would be closer to the current NFL than it seems at first. In one way, you’ll still need to be a great athlete to play, since most of what makes a great athlete great is their motor cortex, their muscle memory, and their decision-making. But the other component of being a great athlete—the physical body itself—would now be artificial. The NFL could make all of the artificial playing bodies identical—this would be a cool way to see whose skills were actually best—or they could insist that artificial body matches in every way the biological body of the athlete, to mimic as closely as possible how the game would go if players used their biological bodies like in the old days. Either way, if this rule change happened, you can imagine how *crazy* it would seem to people that players used to have their actual, fragile brains *on the field*.

I could go on. The communication possibilities in a wizard hat world, especially when you combine them with each other, are endless—and damn fun to think about.

The Wizard Era: Internal Control



Communication—the flow of information into and out of your brain—is only one way your wizard hat will be able to serve you.

A whole-brain interface can stimulate any part of your brain in any way—it has to have this capability for the input half of all the communication examples above. But that capability also gives you a whole new level of control over your brain. Here are some ways people of the future might take advantage of that:

Win the battle in your head for both sides

Often, the battle in our heads between our prefrontal cortex and limbic system comes down to the fact that both parties are trying to do what's best for us—it's just that our limbic system is wrong about what's best for us because it thinks we live in a tribe 50,000 years ago.

Your limbic system isn't making you eat your ninth Starburst candy in a row because it's a dick—it's making you eat it because it thinks that A) any fruit that sweet and densely chewy must be *super* rich in calories and B) you might not find food again for the next four days so it's a good idea to load up on a high-calorie food whenever the opportunity arises.

Meanwhile, your prefrontal cortex is just watching in horror like "WHY ARE WE DOING THIS."

But Moran believes that a good brain interface could fix this problem:⁵³

Consider eating a chocolate cake. While eating, we feed data to our cognitive apparatus. These data provide the enjoyment of the cake. The enjoyment isn't in the cake, per se, but in our neural experience of it. Decoupling our sensory desire (the experience of cake) from the underlying survival purpose (nutrition) will soon be within our reach.

This concept of "sensory decoupling" would make *so* much sense if we could pull it off. You could get the *enjoyment* of eating like shit without actually putting shit in your body. Instead, Moran says, what would go in your body would be "nutrition inputs customized for each person based on genomes, microbiomes or other factors. Physical diets released from the tyranny of desire."⁵⁴

The same principle could apply to things like sex, drugs, alcohol, and other pleasures that get people into trouble, healthwise or otherwise.

Ramez Naam talks about how a brain interface could also help us win the discipline battle when it comes to time:⁵⁵

We know that stimulating the right centers in the brain can induce sleep or alertness, hunger or satiation, ease or stimulation, as quick as the flip of a switch. Or, if you're running code, on a schedule. (Siri: Put me to sleep until 7:30, high priority interruptions only. And let's get hungry for lunch around noon. Turn down the sugar cravings, though.)

Take control of mood disorders

Ramez also emphasized that a great deal of scientific evidence suggests that moods and disorders are tied to what the chemicals in your brain are doing. Right now, we take drugs to alter those chemicals, and Ramez explains why direct neural stimulation is a far better option:⁵⁶

Pharmaceuticals enter the brain and then spread out randomly, hitting whatever receptor they work on all across your brain. Neural interfaces, by contrast, can stimulate just one area at a time, can be tuned in real-time, and can carry information out about what's happening.

Depression, anxiety, OCD, and other disorders may be easy to eradicate once we can take better control of what goes on in our brain.

Mess with your senses

Want to hear what a dog hears? That's easy. The pitch range we can hear is limited by the dimensions of our cochlea—but pitches out of the ear's range can be sent straight into our auditory nerve.³²

Or maybe you want a new sense. You love bird watching and want to be able to sense when there's a bird nearby. So you buy an infrared camera that can detect bird locations by their heat signals and you link it to your brain interface, which stimulates neurons in a certain way to alert you to the presence of a bird and tell you its location. I can't describe what you'd experience when it alerts you, so I'll just say words like "feel" or "see," because I can only imagine the five senses we have. But in the future, there will be more words for new, useful types of senses.

You could also dim or shut off parts of a sense, like pain perhaps. Pain is the body's way of telling us we need to address something, but in the future, we'll elect to get that information in much less unpleasant formats.³³

Increase your knowledge

There's evidence from experiments with rats that it's possible to boost how fast a brain can learn—sometimes by 2x or even 3x—just by priming certain neurons to prepare to make a long-term connection.

Your brain would also have access to all the knowledge in the world, at all times. I talked to Ramez about how accessing information in the cloud might work. We parsed it out into four layers of capability, each requiring a more advanced brain interface than the last:

Level 1: I want to know a fact. I call on the cloud for that info—like Googling something with my brain—and the answer, in text, appears in my mind's eye. Basically what I do now except it all happens in my head.

Level 2: I want to know a fact. I call on the cloud for that info, and then a second later I just *know* it. No reading was involved—it was more like the way I'd recall something from memory.

Level 3: I just know the fact I want to know the second I want it. I don't even know if it came from the cloud or if it was stored in my brain. I can essentially treat the whole cloud like my brain. I don't know all the info—my brain could never fit it all—but any time I want to know something it downloads into my consciousness so seamlessly and quickly, it's as if it were there all along.

Level 4: Beyond just knowing facts, I can deeply *understand* anything I want to, in a complex way. We discussed the example of *Moby Dick*. Could I download *Moby Dick* from the cloud into my memory and then suddenly have it be the same as if I had read the whole book? Where I'd have thoughts and opinions and I could cite passages and have discussions about the themes?

Ramez thinks all four of these are possible with enough time, but that the fourth in particular will take a very long time to happen, if ever.

So there are about 50 delightful potential things about putting a wizard hat on your brain. Now for the undelightful part.

The scary thing about wizard hats

As is always the case with the advent of new technologies, when the Wizard Era rolls around, the dicks of the world will do their best to ruin everything.

And this time, the stakes are extra high. Here are some things that could suck:

Trolls can have an even fielder day. The troll-type personalities of the world have been having a field day ever since the internet came out. They literally can't believe their luck. But with brain interfaces, they'll have an even fielder day. Being more connected to each other means a lot of good things—like empathy going up as a result of more exposure to all kinds of people—but it also means a lot of bad things. Just like the internet. Bad guys will have more opportunity to spread hate or build hateful coalitions. The internet has been a godsend for ISIS, and a brain-connected world would be an even more helpful recruiting tool.

Computers crash. And they have bugs. And normally that's not the end of the world, because you can try restarting, and if it's really being a piece of shit, you can just get a new computer. You can't get a new head. There will have to be a way *way* higher number of precautions taken here.

Computers can be hacked. Except this time they have access to your thoughts, sensory input, and memories. Bad times.

Holy shit computers can be hacked. In the last item I was thinking about bad guys using hacking to *steal* information from my brain. But brain interfaces *can also put information in*. Meaning a clever hacker might be able to change your thoughts or your vote or your identity or make you want to do something terrible you normally wouldn't ever consider. And you wouldn't know it ever happened. You could feel strongly about voting for a candidate and a little part of you would wonder if someone manipulated your thoughts so you'd feel that way. The darkest possible scenario would be an ISIS-type organization actually influencing millions of people to join their cause by altering their thoughts. This is definitely the scariest paragraph in this post. Let's get out of here.

Why the Wizard Era will be a good thing anyway even though there are a lot of dicks

Physics advancements allow bad guys to make nuclear bombs. Biological advancements allow bad guys to make bioweapons. The invention of cars and planes led to crashes that kill over a million people a year. The internet enabled the spread of fake news, made us vulnerable to cyberattack, made terrorist recruiting efforts easier, and allowed predators to flourish.

And yet—

Would people choose to reverse our understanding of science, go back to the days of riding horses across land and boats across the ocean, or get rid of the internet?

Probably not.

New technology also comes along with real dangers and it always *does* end up harming a lot of people. But it also always seems to help a lot *more* people than it harms. Advancing technology almost always proves to be a net positive.

People also love to hate the *concept* of new technology—because they worry it's unhealthy and makes us less human. But those same people, if given the option, usually wouldn't consider going back to George Washington's time, when half of children died before the age of 5, when traveling to other parts of the world was impossible for almost everyone, when a far greater number of humanitarian atrocities were being committed than there are today, when women and ethnic minorities had *far* fewer rights across the world than they do today, when far more people were illiterate and far more people were living under the poverty line than there are today. They wouldn't go back 250 years—a time right before *the biggest explosion of technology in human history happened*. Sounds like people who are immensely *grateful* for technology. And yet their opinion holds—our technology is ruining our lives, people in the old days were much wiser, our world's going to shit, etc. I don't think they've thought about it hard enough.

So when it comes to what will be a long list of dangers of the Wizard Era—they suck, and they'll continue to suck as some of them play out into sickening atrocities and catastrophes. But a vastly larger group of good guys will wage war back, as they always do, and a giant "brain security" industry will be born. And I bet, if given the option, people in the Wizard Era wouldn't for a second consider coming back to 2017.

The Timeline

I always know when humanity doesn't know what the hell is going on with something when all the experts are contradicting each other about it.³⁴

The timeline for our road to the Wizard Era is one of those times—in large part because no one knows to what extent we'll be able to make Stevenson's Law look more like Moore's Law.

My conversations yielded a wide range of opinions on the timeline. One neuroscientist predicted that I'd have a whole-brain interface in my lifetime. Mark Zuckerberg said: "I would be pretty disappointed if in 25 years we hadn't made some progress towards thinking things to computers." One prediction on the longer end came from Ramez Naam, who thought the time of people beginning to install BMIs for reasons other than disability might not come for 50 years and that mass adoption would take even longer.

"I hope I'm wrong," he said. "I hope that Elon bends the curve on this."

When I asked Elon about his timeline, he said:

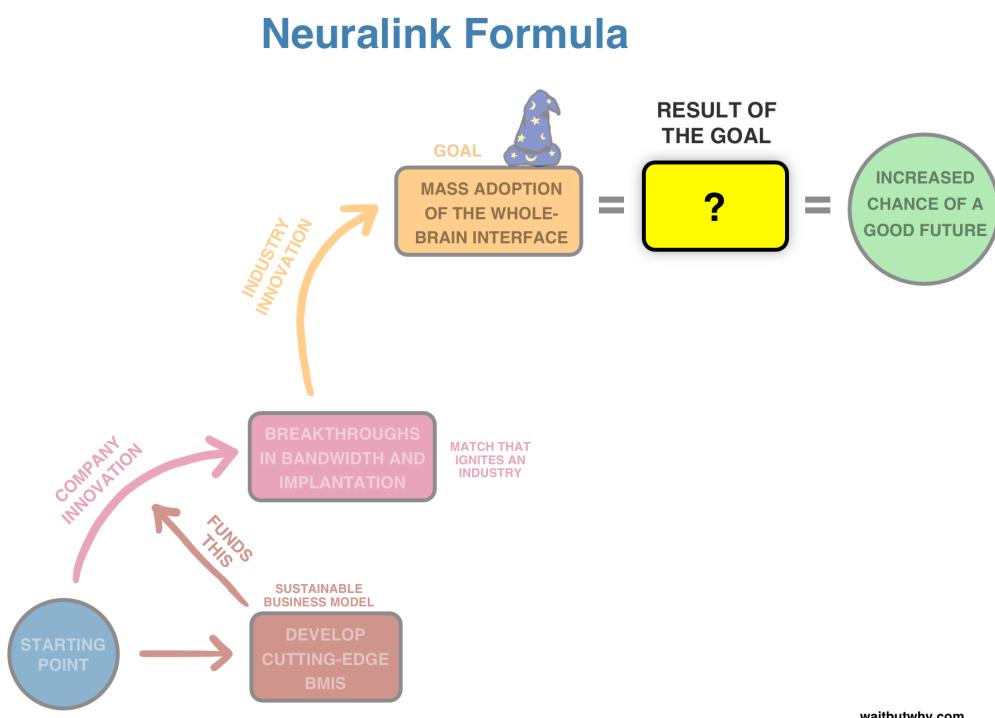
I think we are about 8 to 10 years away from this being usable by people with no disability ... It is important to note that this depends heavily on regulatory approval timing and how well our devices work on people with disabilities.

During another discussion, I had asked him about why he went into this branch of biotech and not into genetics. He responded:

Genetics is just too slow, that's the problem. For a human to become an adult takes twenty years. We just don't have that amount of time.

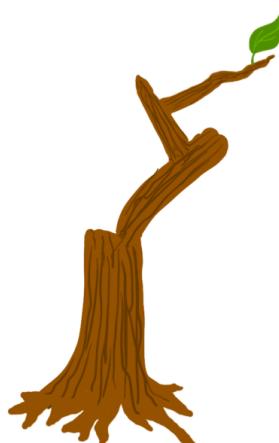
A lot of people working on this challenge have a lot of different motivations for doing so, but rarely did I talk to people who felt motivated by *urgency*.

Elon's urgency to get us into the Wizard Era is the final piece of the Neuralink puzzle. Our last box to fill in:



With Elon's companies, there's always some "result of the goal" that's his real reason for starting the company—the piece that ties the company's goal into humanity's better future. In the case of Neuralink, it's a piece that takes a lot of tree climbing to understand. But with the view from all the way up here, we've got everything we need for our final stretch of the road.

Part 6: The Great Merger



Imagine an alien explorer is visiting a new star and finds three planets circling it, all with life on them. The first happens to be identical to the way Earth was in 10 million BC. The second happens to be identical to Earth in 50,000 BC. And the third happens to be identical to Earth in 2017 AD.

The alien is no expert on primitive biological life but circles around all three planets, peering down at each with his telescope. On the first, he sees lots of water and trees and mountains and some little signs of animal life. He makes out a herd of elephants on an African plain, a group of dolphins skipping along the ocean's surface, and a few other scattered critters living out their Tuesday.

He moves on to the second planet and looks around. More critters, not too much different. He notices one new thing—occasional little points of flickering light dotting the land.

Bored, he moves on to the third planet. Whoa. He sees planes crawling around above the land, vast patches of gray land with towering buildings on them, ships of all kinds sprinkled across the seas, long railways stretching across continents, and he has to jerk his spaceship out of the way when a satellite soars by him.

When he heads home, he reports on what he found: "Two planets with primitive life and one planet with intelligent life."

You can understand why that would be his conclusion—but he'd be wrong.

In fact, it's the first planet that's the odd one out. Both the second and third planets have intelligent life on them—equally intelligent life. So equal that you could kidnap a newborn baby from Planet 2 and swap it with a newborn on Planet 3 and both would grow up as normal people on the other's planet, fitting in seamlessly. Same people.

And yet, how could that be?



The Human Colossus. That's how.

Ever wonder why you're so often unimpressed by humans and yet so blown away by the accomplishments of humanity?

It's because *humans* are still, deep down, those people on Planet 2.

Plop a baby human into a group of chimps and ask them to raise him, Tarzan style, and the human as an adult will know how to run around the forest, climb trees, find food, and masturbate. That's who each of us actually is.

Humanity, on the other hand, is a superintelligent, tremendously-knowledgeable, millennia-old Colossus, with 7.5 billion neurons. And that's who built Planet 3.

The invention of language allowed each human brain to dump its knowledge onto a pile before its death, and the pile became a tower and grew taller and taller until one day, it became the brain of a great Colossus that built us a civilization. The Human Colossus has been inventing things ever since, getting continually better at it with time. Driven only by the desire to create value, the Colossus is now moving at an unprecedented pace—which is why we live in an unprecedented and completely anomalous time in history.

You know how I said we might be living literally *on* the line between two vast eras of communication?

Well the truth is, we seem to be on a *lot* of historic timeline boundaries. After 1,000 centuries of human life and 3.8 billion years of Earthly life, it seems like *this* century will be the one where Earth life makes the leap from the Single-Planetary Era to the Multi-Planetary Era. This century may be the one when an Earthly species finally manages to wrest the genetic code from the forces of evolution and learns to

reprogram itself. People alive today could witness the moment when biotechnology finally frees the human lifespan from the will of nature and hands it over to the will of each individual.

The Human Colossus has reached an entirely new level of power—the kind of power that can overthrow 3.8-billion-year eras—positioning us on the verge of *multiple* tipping points that will lead to unimaginable change. And if our alien friend finds a fourth planet one day that happens to be identical to Earth in 2100, you can be pretty damn sure it'll look nothing to him like Planet 3.

I hope you enjoyed Planet 3, because we're leaving it. Planet 4 is where we're headed, whether we like it or not.

PLANET 4



waitbutwhy.com

If I had to sum up the driving theme behind everything Elon Musk does, it would be pretty simple:

He wants to prepare us for Planet 4.

He lives in the big picture, and his only lens is the maximum zoom-out. That's why he's such an unusual visionary. It's also why he's so worried.

It's not that he thinks Planet 4 is definitely a bad place—it's that he thinks it *could be* a bad place, and he recognizes that the generations alive today, whether they realize it or not, are the first in history to face real, hardcore existential risk.

At the same time, the people alive today also are the first who can live with the actually realistic hope for a genuinely utopian future—one that defies even death and taxes. Planet 4 could be our promised land.

When you zoom way out, you realize how unfathomably high the stakes actually are.

And the outcome isn't at the whim of chance—it's at the whim of the Human Colossus. Planet 4 is only coming because the Colossus is building it. And whether that future is like heaven or hell depends on what the Colossus does—maybe over the next 150 years, maybe over only the next 50. Or 25.

But the unfortunate thing is that the Human Colossus isn't optimized to maximize the chances of a safe transition to the best possible Planet 4 for the most possible humans—it's optimized to build Planet 4, in any way possible, as quickly as possible.

Understanding all of this, Elon has dedicated his life to trying to influence the Human Colossus to bring its motivation more in line with the long-term interests of humans. He knows it's not possible to rewire the Human Colossus—not unless existential risk were suddenly directly in front of each human's face, which normally doesn't happen until it's already too late—so he treats the Colossus like a pet.

If you want your dog to sit, you correlate sitting on command with getting a treat. For the Human Colossus, a treat is a ripe new industry simultaneously exploding in both supply and demand.

Elon saw the Human Colossus dog peeing on the floor in the form of continually adding ancient, deeply-buried carbon into the carbon cycle—and rather than plead with the Colossus to stop peeing on the floor (which a lot of people waste their breath doing) or try to threaten the Colossus into behaving (which governments try to do, with limited success), he's creating an electric car so rad that everyone will want one. The auto industry sees the shift in consumer preferences this is beginning to create, and in the nine years since Tesla released its first car, the number of major car companies with an electric

car in their line went from zero to almost all of them. The Colossus seems to be taking the treat, and a change in behavior may follow.

Elon saw the Human Colossus dog running into traffic in the form of humanity keeping all of its eggs on one planet, despite all of those tipping points on the horizon, so he built SpaceX to learn to land a rocket, which will cut the cost of space travel by about 99% and make dedicating resources to the space industry a much tastier morsel for the Colossus. His plan with Mars isn't to try to convince humanity that it's a good idea to build a civilization there in order to buy life insurance for the species—it's to create an affordable regular cargo and human transit route to Mars, knowing that once that happens, there will be enough value-creation opportunity in Mars development that the Colossus will become determined to make it happen.

But to Elon, the scariest thing the Human Colossus is doing is teaching the Computer Colossus to think. To Elon, and many others, the development of superintelligent AI poses by far the greatest existential threat to humanity. It's not that hard to see why. Intelligence gives us godlike powers over all other creatures on Earth—which has not been a fun time for the creatures. If any of their body parts are possible value creators, we have major industries processing and selling those body parts. We sometimes kill them for sport. But we're probably the least fun all the times we're just doing our thing, for our own reasons, with no hate in our hearts or desire to hurt anyone, and there are creatures, or ecosystems, that just *happen* to be in our way or in the line of fire of the side effects of what we're doing. People like to get all mad at humanity about this, but really, we're just doing what species do—being selfish, first and foremost.

The issue for other creatures isn't our selfishness—it's the immense damage our selfishness can do because of the tremendous power we have over them. Power that comes from our intelligence advantage.

So it's pretty logical to be apprehensive about the prospect of intentionally creating something that will have (perhaps far) more intelligence than we do—especially since every human on the planet is an amateur at creating something like that, because no one has ever done it before.

And things are progressing quickly. Elon talked about the rapid progress made by Google's game-playing AI:

I mean, you've got these two things where AlphaGo crushes these human players head-on-head, beats Lee Sedol 4 out of 5 games and now it will beat a human every game all the time, while playing the 50 best players, and beating them always, all the time. You know, that's like one year later.

And it's on a harmless thing like AlphaGo right now. But the degrees of freedom at which the AI can win are increasing. So, Go has many more degrees of freedom than Chess, but if you take something like one of the real-time strategy competitive games like League of Legends or Dota 2, that has vastly more degrees of freedom than Go, so it can't win at that yet. But it will be able to. And then there's reality, which has the ultimate number of degrees of freedom. ³⁵

And for reasons discussed above, that kind of thing worries him:

What I came to realize in recent years—the last couple years—is that AI is obviously going to surpass human intelligence by a lot. ... There's some risk at that point that something bad happens, something that we can't control, that humanity can't control after that point—either a small group of people monopolize AI power, or the AI goes rogue, or something like that. It may not, but it could.

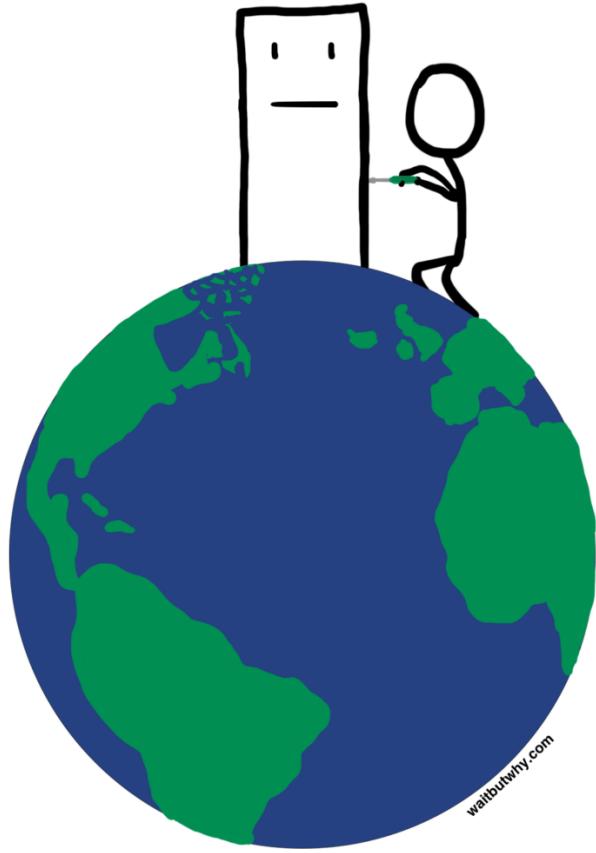
But in typical Human Colossus form, "the collective will is not attuned to the danger of AI."

When I interviewed Elon in 2015, I asked him if he would ever join the effort to build superintelligent AI. He said, "My honest opinion is that we shouldn't build it." And when I later commented that building something smarter than yourself did seem like a basic Darwinian error (a phrase I stole from Nick Bostrom), Elon responded, "We're gonna win the Darwin Award, collectively."

Now, two years later, here's what he says:

I was trying to really sound the alarm on the AI front for quite a while, but it was clearly having no impact (laughs) so I was like, "Oh fine, okay, then we'll have to try to help develop it in a way that's good."

He's accepted reality—the Human Colossus is not going to quit until the Computer Colossus, one day, wakes up. This is happening.



No matter what anyone tells you, no one knows what will happen when the Computer Colossus learns to think. In my [long AI explainer](#), I explored the reasoning of both those who are convinced that superintelligent AI will be the solution to every problem we have, and those who see humanity as a bunch of kids playing with a bomb they don't understand. I'm personally still torn about which camp I find more convincing, but it seems pretty rational to plan for the worst and do whatever we can to increase our odds. Many experts agree with that logic, but there's little consensus on the best strategy for creating superintelligent AI safely—just a whole lot of ideas from people who acknowledge they don't really know the answer. How could anyone know how to take precautions for a future world they have no way to understand?

Elon also acknowledges he doesn't know the answer—but he's working on a plan he thinks will give us our best shot.

Elon's Plan

Abraham Lincoln was pleased with himself when he came up with the line:

—and that government of the people, by the people, for the people, shall not perish from the earth.

Fair—it's a good line.

The whole idea of "of the people, by the people, for the people" is the centerpiece of democracy.

Unfortunately, "the people" are unpleasant. So democracy ends up being unpleasant. But unpleasant tends to be a dream compared to the alternatives. Elon talked about this:

I think that the protection of the collective is important. I think it was Churchill who said, "Democracy's the worst of all systems of government, except for all the others." It's fine if you have Plato's incredible philosopher king as the king, sure. That would be fine. Now, most dictators do not turn out that way. They tend to be quite horrible.

In other words, democracy is like escaping from a monster by hiding in a sewer.

There are plenty of times in life when it's a good strategy to take a risk in order to give yourself a chance for the best possible outcome, but when the stakes are at their absolute highest, the right move is usually to play it safe. Power is one of those times. That's why, even though democracy essentially guarantees a certain level of mediocrity, Elon says, "I think you're hard-pressed to find many people in the United States who, no matter what they think of any given president, would advocate for a dictatorship."

And since Elon sees AI as the ultimate power, he sees AI development as the ultimate “play it safe” situation. Which is why his strategy for minimizing existential AI risk seems to essentially be that AI power needs to be of the people, by the people, for the people.

To try to implement that concept in the realm of AI, Elon has approached the situation from multiple angles.

For the **by the people** and **for the people** parts, he and Sam Altman created [OpenAI](#)—a self-described “non-profit AI research company, discovering and enacting the path to safe artificial general intelligence.”

Normally, when humanity is working on something new, it starts with the work of a few innovative pioneers. When they succeed, an industry is born and the Human Colossus jumps on board to build upon what the pioneers started, en masse.

But what if the thing those pioneers were working on was a magic wand that might give whoever owned it immense, unbreakable power over everyone else—including the power to prevent anyone else from making a magic wand? That would be kinda stressful, right?

Well that’s how Elon views today’s early AI development efforts. And since he can’t stop people from trying to make a magic wand, his solution is to create an open, collaborative, transparent magic wand development lab. When a new breakthrough innovation is discovered in the lab, instead of making it a tightly-kept secret like the other magic wand companies, the lab publishes the innovation for anyone to see or borrow for their own magic-wand-making efforts.

On one hand, this could have drawbacks. Bad guys are out there trying to make a magic wand too, and you *really* don’t want the first magic wand to end up in the hands of a bad guy. And now the bad guys’ development efforts can benefit from all of the innovations being published by the lab. This is a serious concern.

But the lab also boosts the efforts of millions of other people trying to create magic wands. This generates a ton of competition for the secretive early pioneers, and it becomes less likely that any one inventor can create a magic wand long before others also do. More likely is that when the first magic wand is eventually created, there are thousands of others near completion as well—different wands, with different capabilities, made by different people, for different reasons. If we have to have magic wands on Earth, Elon thinks, let’s at least make sure they’re in the hands of a large number of people across the world—not one all-powerful sorcerer. Or as he puts it:

Essentially, if everyone’s from planet Krypton, that’s great. But if only one of them is Superman and Superman also has the personality of Hitler, then we’ve got a problem.

More broadly, a single pioneer’s magic wand would likely have been built to serve that inventor’s own needs and purposes. But by turning the future magic wand industry into a collective effort, a wide variety of needs and purposes will have a wand made for them, making it more likely that the capabilities of the world’s aggregate mass of magic wands will overarchingly represent the needs of the masses.

You know, like democracy.

It worked fine for Nikola Tesla and Henry Ford and the Wright Brothers and Alan Turing to jump-start revolutions by jumping way out ahead of the pack. But when you’re dealing with the invention of something *unthinkably* powerful, you can’t sit back and let the pioneers kick things off—it’s leaving too much to chance.

OpenAI is an effort to democratize the creation of AI, to get the entire Human Colossus working on it during its pioneer phase. Elon sums it up:

AI is definitely going to vastly surpass human abilities. To the degree that it is linked to human will, particularly the sum of a large number of humans, it would be an outcome that is desired by a large number of humans, because it would be a function of their will.

So now you’ve maybe got early human-level-or-higher AI superpower being made by the people, for the people—which brings down the likelihood that the world’s AI ends up in the hands of a single bad guy or a tightly-controlled monopoly.

Now all we’ve got left is **of the people**.

This one should be easy. Remember, the Human Colossus is creating superintelligent AI for the same reason it created cars, factory machines, and computers—to serve as an extension of itself to which it can outsource work. Cars do our walking, factory machines do our manufacturing, and computers take care of information storage, organization, and computation.

Creating computers that can think will be our greatest invention yet—they'll allow us to outsource our most important and high-impact work. Thinking is what built everything we have, so just imagine the power that will come from building ourselves a superintelligent thinking *extension*. And extensions of the people by definition belong to the people—they're *of* the people.

There's just this one thing—

High-caliber AI isn't quite like those other inventions. The rest of our technology is great at the thing it's built to do, but in the end, it's a mindless machine with narrow intelligence. The AI we're trying to build will be *smart*, like a person—like a *ridiculously* smart person. It's a fundamentally different thing than we've ever made before—so why would we expect normal rules to apply?

It's always been an automatic thing that the technology we make inherently belongs to us—it's such an obvious point that it almost seems silly to make it. But could it be that if we make something smarter than a person, it might not be so easy to control?

Could it be that a creation that's better at thinking than any human on Earth might not be fully content to serve as a human extension, even if that's what it was built to do?

We don't know how issues will actually manifest—but it seems pretty safe to say that yes, these possibilities *could be*.

And if what could be turns out to *actually* be, we may have a serious problem on our hands.

Because, as the human history case study suggests, when there's something on the planet way smarter than everyone else, it can be a really bad thing for everyone else. And if AI becomes the new thing on the planet that's way smarter than everyone else, and it turns out *not* to clearly belong to us—it means that it's its *own thing*. Which drops us into the category of "everyone else."

So *people* gaining monopolistic control of AI is its own problem—and one that OpenAI is hoping to solve. But it's a problem that may pale in comparison to the prospect of AI being uncontrollable.

This is what keeps Elon up at night. He sees it as only a matter of time before superintelligent AI rises up on this planet—and when that happens, he believes that it's critical that we don't end up as part of "everyone else."

That's why, in a future world made up of AI and everyone else, he thinks we have only one good option:

To be AI.

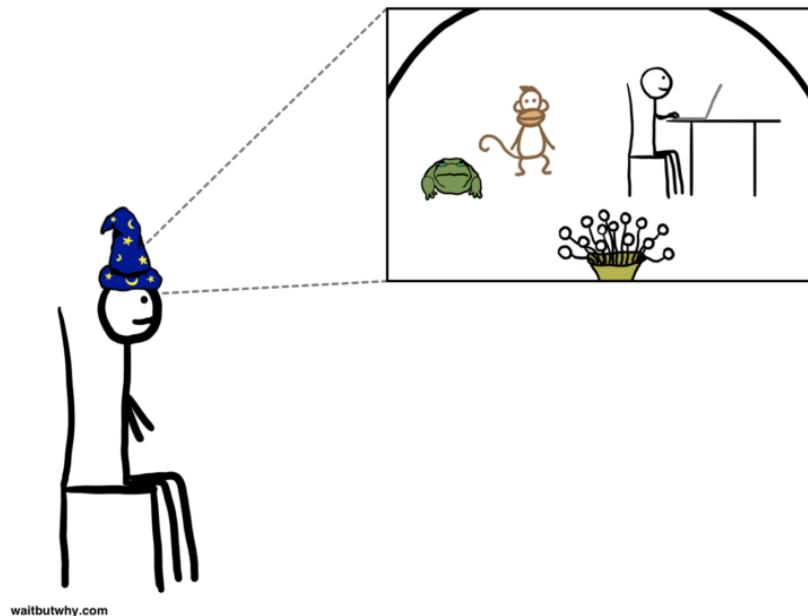
Remember before when I said that there were two things about wizard hats we had to wrap our heads around?

1) The intensely mind-bending idea

2) The super ridiculously intensely mind-bending idea

This is where #2 comes in.

These two ideas are the two things Elon means when he refers to the wizard hat as a digital tertiary layer in our brains. The first, as we discussed, is the concept that a whole-brain interface is kind of the same thing as putting our devices in our heads—effectively making your *brain* the device. Like this:



Your devices give you cyborg superpowers and a window into the digital world. Your brain's wizard hat electrode array is a new brain structure, joining your limbic system and cortex.

But your limbic system, cortex, and wizard hat are just the hardware systems. When you experience your limbic system, it's not the physical system you're interacting with—it's the *information flow* within it. It's the *activity* of the physical system that bubbles up in your consciousness, making you feel angry, scared, horny, or hungry.

Same thing for your cortex. The napkin wrapped around your brain stores and organizes information, but it's the information itself that you experience when you think something, see something, hear something, or feel something. The visual cortex in itself does nothing for you—it's the stream of photon information flowing through it that gives you the experience of having a visual cortex. When you dig in your memory to find something, you're not searching for *neurons*, you're searching for information stored in the neurons.

The limbic system and cortex themselves are just gray matter. The flow of activity within the gray matter is what forms your familiar internal characters, the monkey brain and the rational human brain.

So what does that mean about your digital tertiary layer?

It means that while what's actually in your brain is the physical device—the electrode array itself—the component of the tertiary layer that you'll experience and get to know as a character is the information that flows through the array.

And just like the feelings and urges of the limbic system and the thoughts and chattering voice of the cortex all feel to you like *parts of you*—like your inner essence—the activity that flows through your wizard hat will feel like *you* every bit as much as the others do. He says:

I think that, conceivably, there's a way for there to be a tertiary layer that feels like it's part of you. It's not some thing that you offload to, it's you.

This makes sense on paper. You do most of your “thinking” with your cortex, but then when you get hungry, you don't say, “My limbic system is hungry,” you say, “I'm hungry.” Likewise, Elon thinks, when you're trying to figure out the solution to a problem and your AI comes up with the answer, you won't say, “My AI got it,” you'll say, “Aha! I got it.” When your limbic system wants to procrastinate and your cortex wants to work, a situation I *might* be familiar with, it doesn't feel like you're arguing with some external being, it feels like a singular *you* is struggling to be disciplined. Likewise, when you think up a strategy at work and your AI disagrees, that'll be a genuine disagreement and a debate will ensue—but it will feel like an *internal* debate, not a debate between you and someone else that just happens to take place in your thoughts. The debate will feel like *thinking*.

It makes sense on paper.

But when I first heard Elon talk about this concept, it didn't really feel right. No matter how hard I tried to get it, I kept framing the idea as something familiar—like an AI system whose *voice* I could hear in my head, or even one that I could think together with. But in those instances, the AI still seemed like an external system I was communicating with. It didn't seem like *me*.

But then, one night while working on the post, I was rereading some of Elon's quotes about this, and it suddenly *clicked*. The AI *would* be me. *Fully. I got it.*

Then I lost it. The next day, I tried to explain the epiphany to a friend and I left us both confused. I was back in "Wait, but it kind of wouldn't really *be me*, it would be *communicating* with me" land. Since then, I've dipped into and out of the idea, never quite able to hold it for long. The best thing I can compare it to is having a moment when it actually makes *sense* that time is relative and space-time is a single fabric. For a second, it seems *intuitive* that time moves slower when you're moving really fast. And then I lose it. As I typed those sentences just now, it did *not* seem intuitive.

The idea of *being* AI is especially tough because it combines *two* mind-numbing concepts—the brain interface and the abilities it would give you, and artificial general intelligence. Humans today are simply not equipped to understand either of those things, because as imaginative as we think we are, our imaginations only really have our life experience as their toolkit, and these concepts are both totally novel. It's like trying to imagine a color you've never seen.

That's why when I hear Elon talk with conviction about this stuff, I'm somewhere in between deeply believing it myself and taking his word for it. I go back and forth. But given that he's someone who probably found space-time intuitive when he was seven, and given that he's someone who knows how to colonize Mars, I'm inclined to listen hard to what he says.

And what he says is that this is all about bandwidth. It's obvious why bandwidth matters when it comes to making a wizard hat useful. But Elon believes that when it comes to interfacing with AI, high bandwidth isn't just *preferred*, but actually *fundamental* to the prospect of being AI, versus simply *using* AI. Here he is walking me through his thoughts:

The challenge is the communication bandwidth is extremely slow, particularly output. When you're outputting on a phone, you're moving two thumbs very slowly. That's crazy slow communication. ... If the bandwidth is too low, then your integration with AI would be very weak. Given the limits of very low bandwidth, it's kind of pointless. The AI is just going to go by itself, because it's too slow to talk to. The faster the communication, the more you'll be integrated—the slower the communication, the less. And the more separate we are—the more the AI is "other"—the more likely it is to turn on us. If the AIs are all separate, and vastly more intelligent than us, how do you ensure that they don't have optimization functions that are contrary to the best interests of humanity? ... If we achieve tight symbiosis, the AI wouldn't be "other"—it would be you and with a relationship to your cortex analogous to the relationship your cortex has with your limbic system.

Elon sees communication bandwidth as the key factor in determining our level of integration with AI, and he sees that level of integration as the key factor in how we'll fare in the AI world of our future:

We're going to have the choice of either being left behind and being effectively useless or like a pet—you know, like a house cat or something—or eventually figuring out some way to be symbiotic and merge with AI.

Then, a second later:

A house cat's a good outcome, by the way.

Without really understanding what kinds of AI will be around when we reach the age of superintelligent AI, the idea that human-AI integration will lend itself to the protection of the species makes intuitive sense. Our vulnerabilities in the AI era will come from bad people in control of AI or rogue AI not aligned with human values. In a world in which millions of people control a little piece of the world's aggregate AI power—people who can think with AI, can defend themselves with AI, and who fundamentally understand AI because of their own integration with it—humans are less vulnerable. People will be a lot more powerful, which is scary, but like Elon said, if everyone is Superman, it's harder for any one Superman to cause harm on a mass scale—there are lots of checks and balances. And we're less likely to lose control of AI in general because the AI on the planet will be so widely distributed and varied in its goals.

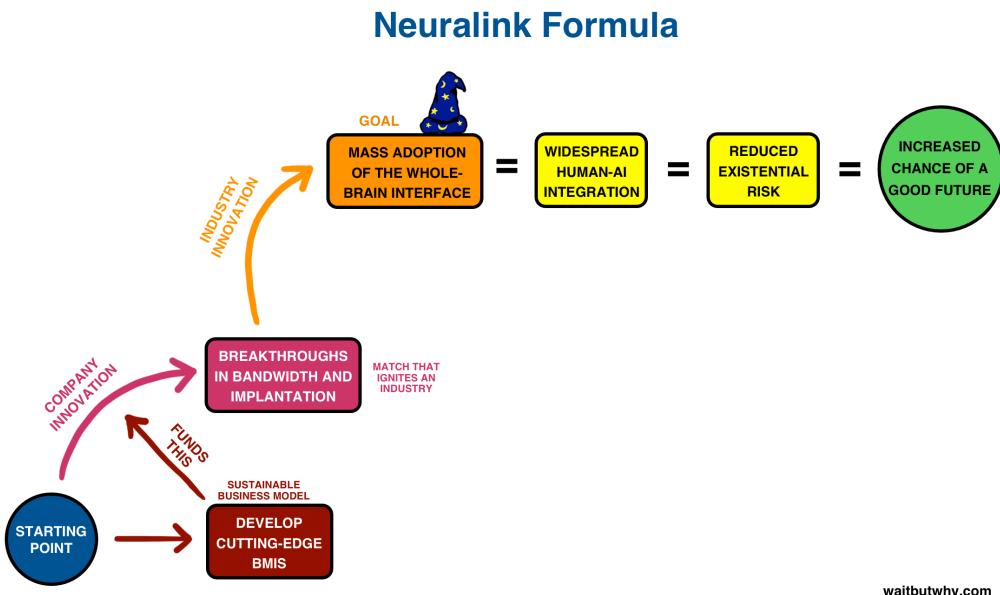
But time is of the essence here—something Elon emphasized:

The pace of progress in this direction matters a lot. We don't want to develop digital superintelligence too far before being able to do a merged brain-computer interface.

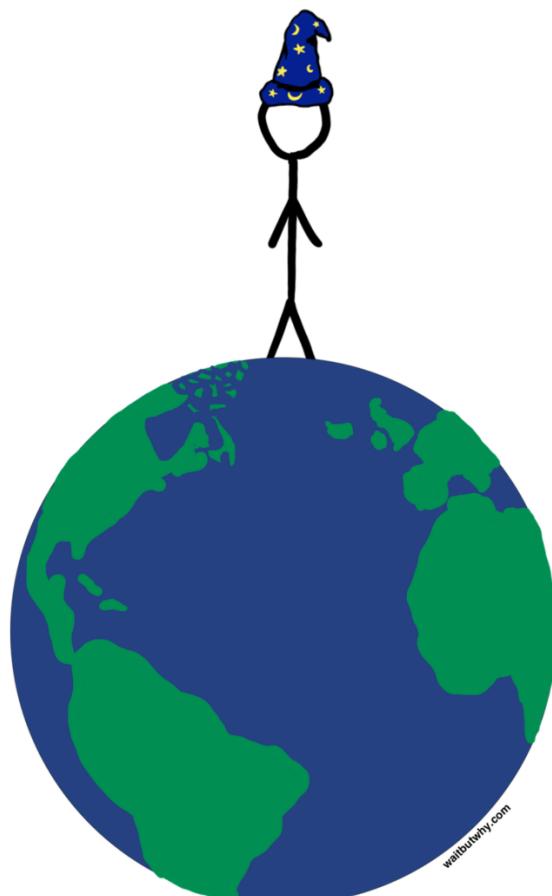
When I thought about all of this, one reservation I had was whether a whole-brain interface would be enough of a change to make integration likely. I brought this up with Elon, noting that there would still be a vast difference between our thinking speed and a computer's thinking speed. He said:

Yes, but increasing bandwidth by orders of magnitude would make it better. And it's directionally correct. Does it solve all problems? No. But is it directionally correct? Yes. If you're going to go in some direction, well, why would you go in any direction other than this?

And that's why Elon started Neuralink.



He started Neuralink to accelerate our pace into the Wizard Era—into a world where he says that “everyone who wants to have this AI extension of themselves could have one, so there would be billions of individual human-AI symbiotes who, collectively, make decisions about the future.” A world where AI really could be of the people, by the people, for the people.



I'll guess that right now, some part of you believes this insane world we've been living in for the past 38,000 words could really maybe be the future—and another part of you refuses to believe it. I've got a little of both of those going on too.

But the *insanity* part of it shouldn't be the reason it's hard to believe. Remember—George Washington died when he saw 2017. And our future *will* be unfathomably shocking to us. The only difference is that things are moving even faster now than they were in George's time.

The concept of being blown away by the future speaks to the magic of our collective intelligence—but it also speaks to the naivety of our intuition. Our minds evolved in a time when progress moved at a snail's pace, so that's what our hardware is calibrated to. And if we don't actively override our intuition—the part of us that reads about a future this outlandish and refuses to believe it's possible—we're living in denial.

The reality is that we're whizzing down a *very* intense road to a very intense place, and no one knows what it'll be like when we get there. A lot of people find it scary to think about, but I think it's exciting. Because of when we happened to be born, instead of just living in a normal world like normal people, we're living inside of a *thriller* movie. Some people take this information and decide to be like Elon, doing whatever they can to help the movie have a happy ending—and thank god they do. Because I'd rather just be a gawking member of the audience, watching the movie from the edge of my seat and rooting for the good guys.

Either way, I think it's good to climb a tree from time to time to look out at the view and remind ourselves what a time this is to be alive. And there are a *lot* of trees around here. Meet you at another one sometime soon.

If you're into Wait But Why, sign up for the [Wait But Why email list](#) and we'll send you the new posts right when they come out. That's the only thing we use the list for—and since my posting schedule isn't exactly...regular...this is the best way to stay up-to-date with WBW posts.

If you'd like to support Wait But Why, [here's our Patreon](#).

The clean version of this post, appropriate for all ages, is free to read [here](#).

To print this post or read it offline, [try the PDF](#).

More Wait But Why stuff:

If you want to understand AI better, [here's my big AI explainer](#).

And here's the full Elon Musk post series:

[Part 1, on Elon: Elon Musk: The World's Raddest Man](#)

[Part 2, on Tesla: How Tesla Will Change the World](#)

[Part 3, on SpaceX: How \(and Why\) SpaceX Will Colonize Mars](#)

[Part 4, on the thing that makes Elon so effective: The Chef and the Cook: Musk's Secret Sauce](#)

If you're sick of science and tech, check these out instead:

[Why Procrastinators Procrastinate](#)

[Religion for the Nonreligious](#)

[The Tail End](#)

Thanks to the Neuralink team for answering my 1,200 questions and explaining things to me like I'm five. Extra thanks to Ben Rapoport, Flip Sabes, and Moran Cerf for being my question-asking go-tos in my many dark moments of despair.
