

Notes on Robert McIntyre's Brain Preservation Talk at the Long Now Foundation



by **JohnGreer** 13 min read 28th Apr 2021 No comments

This is cross-posted from my blog and is written more for a general audience rather than LessWrong people who will be more familiar with some of the relevant concepts.

These are my notes on Robert McIntyre's talk at the Long Now Foundation:
Engram Preservation: Early Work Towards Mind Uploading | Robert McIntyre

https://youtu.be/FCK6Yrx_PSQ

I stumbled across the Long Now Foundation back in 2011 and heard about their 10,000 year clock, a project to design a clock to keep time for 10,000 years (and bring media attention to their project) and it's cool seeing they're still doing stuff. (Ten years ago isn't that far back even in normal time so I hope their foundation lasts longer than that.)

Robert McIntyre is the CEO of a company called Nectome. Nectome's goal is to try to better understand human memory and preserve brains. I know what you're thinking... the human memory part seems normal enough but preserving brains? For who, zombies? Stick with me.

I remember hearing something about brain preservation and a thing called the "Large Mammal Brain Preservation Prize" being won and a company called Nectome doing it but didn't look too much into it.

I re-stumbled across Nectome in reading the writings of fellow cryonics and life extension supporter, Mati Roy.

Let's dive in:

"I consider myself an archivist. And what I work on archiving are human memories."

Stages of Information Transmission in History

Robert talks about dividing human history into different stages. We advance from one stage to another by developing technologies that allow better transmission and preservation of information. Every time we invent tech that does this, it radically catapults

our society to new heights. Note: Regular people often think about technology as just gadgets like a TV or iPhone but technology (of course, depending on semantics) does include things like language and writing.

This is a similar paradigm to the transition from the hunter-gather stage to the agriculture stage to the industrial revolution. It cannot be overestimated how massive these changes were.

This video discusses this briefly in a nice way:

<https://www.youtube.com/watch?v=mrWPSP35gUk>

He calls the pre-language stage intuitive and talks about how weird this must have been. It is a trip to think about what our qualia would have been like in a pre-language era. Like what would our thoughts have been like?

Then we eventually developed language, which was a completely transformative change. The problem with oral communication is it's extremely low-bandwidth: not a lot of information can be transmitted and it takes a long time to do it. Elon Musk has made the same point about bandwidth for the importance of Neuralink.

Side note: Some of this oral history talk reminds me of Sam Jackson's speech in this scene in Unbreakable (spoilers).

Then we went from just having oral communication to having symbolic communication with writing. This again was massively game-changing.

Writing is amazing because it can, among other things, preserve and transmit more information than any one person can remember and for much longer periods of time. For instance, Shakespeare has been dead for over 400 years but we can still enjoy his plays.

Writing still has downsides, though. It takes a lot of time to parse it and we still lose a lot of information because it's hard to record things in writing.

Robert argues we lose most of the value/wisdom by only being able to store writing. He doesn't talk about video recordings or anything so I think his argument loses some of its strength which I go into more later on.

Information Theory

He then goes into some information theory.

Does information being preserved depend on the technology to read/extract/understand it? Was it always preserved or only preserved after you invent the technology?

(Robert says it was always preserved. Preservation does not equal the ability to read it.)

His very brief dive into information theory reminds me that I wish someone could point me to a post or video on information theory where they showcase all the distilled, useful parts.

He brings up the injective function and says that preservation means that different things remain different. On the one hand, this feels like an elegant way of capturing what preservation is, and on the other hand, it seems like it may not capture what it actually is or only be one part of preservation, but I don't know.

So he boils it down to:

Preservation = differences stay different

Preservation \neq Understanding

Example: He can't read Chinese but he could preserve Chinese books.

In practice, this happened where people started storing DNA back in the 70s before we had the ability to read it (which wouldn't come for a few decades). More on this later.

How do we know how memories are stored in the brain? The brain is full of electrical activity so how do we know that's not essential for memory storage?

People who fell into ice were sometimes able to be brought back and then often had their memories or most of their memories intact, so this suggests we don't need the electrical activity for storage.

Because it'd be way questionable to conduct experiments like this, it's good we get some value out of accidents like this.

This finding about people falling into ice and being semi-stable/able to be brought back led to some cool developments.

Memories are physical/DHCA

Deep hypothermic circulatory arrest or DHCA is a type of technique where you put someone in a hypothermic state so you can do surgeries on them where you cut off circulation and brain function so it minimizes damage to the brain. Pretty cool stuff.

Because of these things we know that memories are physical. Separately we know that memories are physical because of all the reasons why we know everything including consciousness is physical. (This is a longer discussion but it's mostly basic to people who aren't religious. He also very briefly gets into this at the end of the talk.)

Side note: this is about the continuously updating version of death. Back in the day, "death" would have been defined as not breathing or no heartbeat. Now we can "bring people back" with CPR, but were they actually dead? Now we have brain death which is quite complicated. My partner is a neurology resident and has participated in declaring brain death which is quite a complicated and rigorous process (although I would argue that it's mostly not useful). This point becomes important especially in light of cryonics where if we could bring people back many years after their bodies/brains were preserved, it's analogous to them not being dead in the same way that someone brought back with CPR isn't dead.

This whole line of thinking has led to the creation of the very useful concept: information-theoretic death. You think of what makes up someone's consciousness -- their memories and personality and whatnot -- and information-theoretic death is about the loss of that information. So if someone is "dead" or "more dead", the less information you can recover.

Memory Consolidation

He then briefly goes into how memory consolidation works. It is roughly divided into three parts based on how long the memory lasts and how much, if it all, is encoded.

1-30 seconds: electrical signals

(memorizing a phone number or the words you just heard lingering)

<2 hours: unstable changes to synapses

(intermediate process) some changes to synapses but changes are liable, by default will revert back to original stage

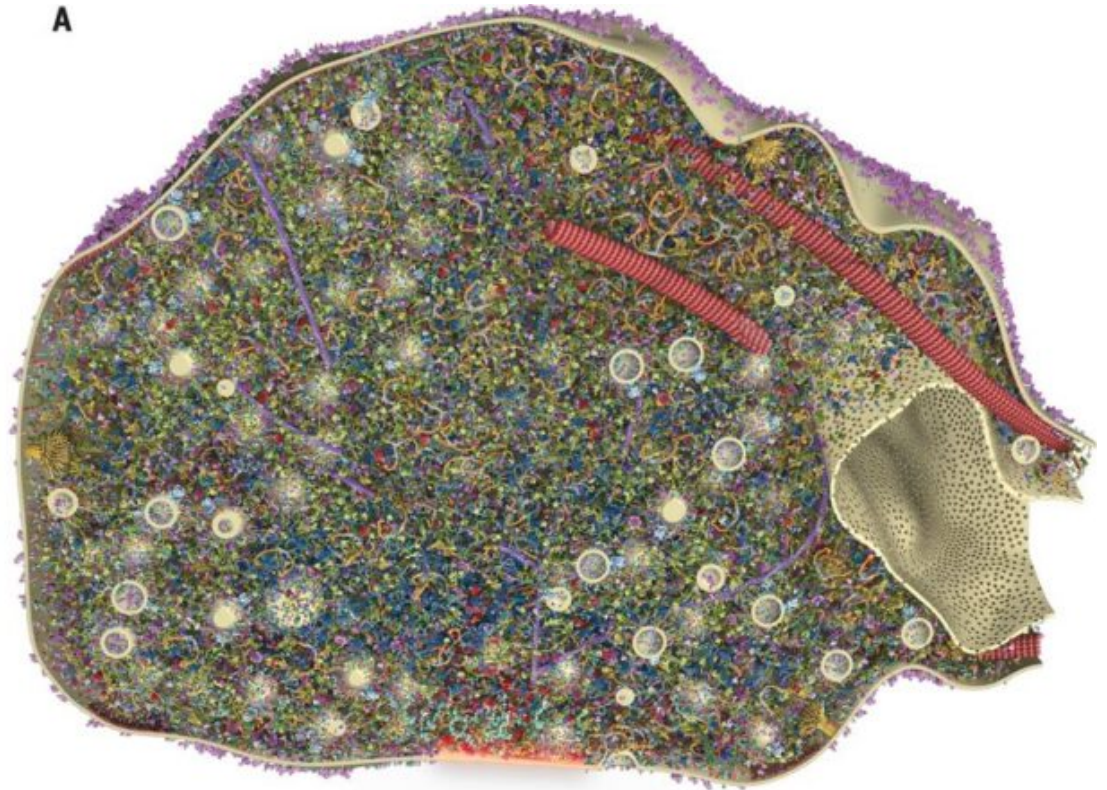
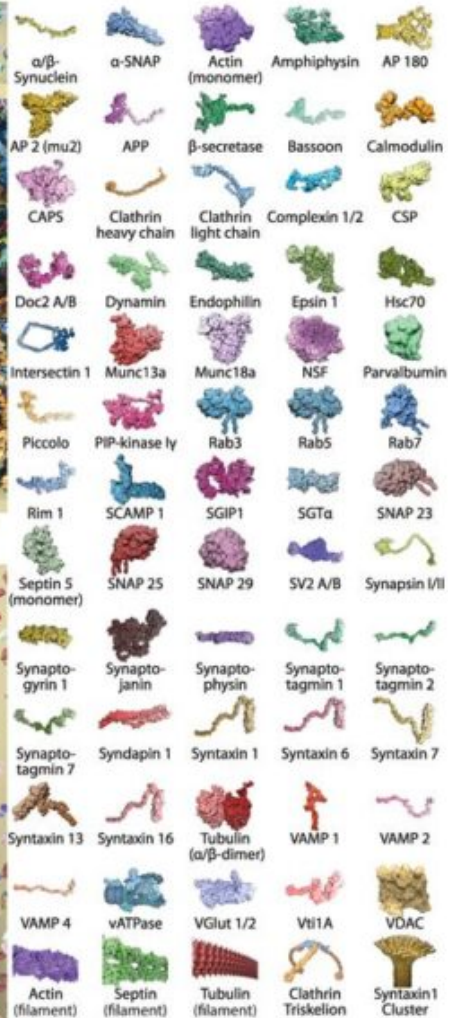
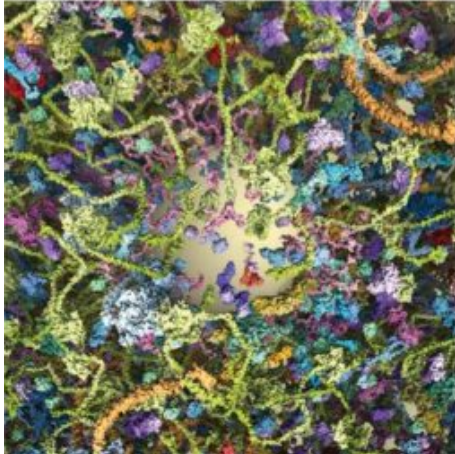
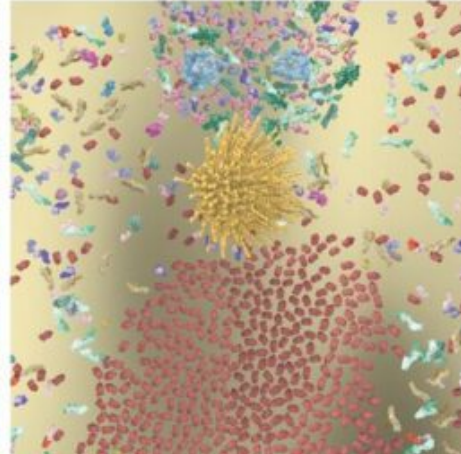
(He doesn't go into what makes them not revert to their original stage but it'd be nice to know more about that.)

>2 hours: structural changes to synapses

(generally encoded as long-term memory, could potentially stay with you your entire life)

Synapses can increase and decrease in size, temporarily or permanently.

He shows this cool, weird image of a synapse. I looked it up and it's the first scientifically accurate 3D model of one.

A**B****C****D**

Changes happen in thousands of synapses throughout the brain even for the most trivial of memories if you're going to remember it longer than two hours. One principle of the brain is it's quite distributed. You can destroy any one synapse and it doesn't affect anything. Pretty cool.

Encoding memory requires protein synthesis. If you inhibit protein synthesis you can't encode long-term memories. This manifests in things like being blackout drunk where you can't remember what happened before despite being conscious at the time.

This reminds me that an important point to appreciate generally is the brain is super duper extremely complicated. Learned this in my brain class in college. It's arguably the most complicated known thing in the universe.

Can you preserve synapses?

Yes, a chemical called glutaraldehyde can and we've been able to preserve them since the 60s.

Synapse preservation is analogous to DNA preservation.

We stored DNA before we could do anything with it and we could only scan it decades after we first started storing it and it was super duper expensive.

First genome scanning was in 2003 and was real expensive. (2.7 billion in 1991 dollars.)

Now you can do it for about \$1000 and it's only getting cheaper. Not bad, huh? (This is one of my main gripes with people who think technology only helps the rich. It's dumb.)

Minor point he made but with a big implication: We could preserve DNA way before but didn't know enough about it until the structure was discovered to confidently say we could store it. So this is a strong argument that it's often best to take action before you know something for sure.

His Background, Q and A, and Random Good Points

Robert says he was always interested in brain preservation but doesn't go into the actual reasons why, which would have been nice although it's possible he doesn't remember.

He volunteered for The Brain Preservation Foundation. He was going to make an explainer video à la Minute Physics but then thought he could just win the Brain Preservation prize the foundation offered.

He briefly mentions the fork in the road of: Do you attack a problem with the tools available now or build better tools? Side note: Better tools are often the big changers of innovation. Think telescope, microscope, transistor, etc.

He won the prize by combining techniques from two different labs. This is a good example of having different disciplines talk to each other and collaboration in general. Reminds me of how Terence Tao is famous for his collaboration even on fields outside of pure maths.

He says preservation is relatively inexpensive but the storage costs are very expensive because it has to be stored at a specific cold but not too cold temperature that's apparently only currently achievable with explosive gases. It seems like the real next step is figuring out how to preserve it in a way that it can be stored at room temperature or one of the easier cold temperatures. And seems like there's maybe a startup idea: coming up with a way to create that Goldilocks temperature safer and cheaper, but I don't know how big the market is for that.

In all this discussion, he knows the landscape of current tech and the costs. How he speaks, like knowing the amounts and costs of things, is very engineer/startup person which is always nice to see.

He thinks it'll be about 70 years until we can access information in preserved brains. I doubt it'll be that long if AGI goes right.

He talks about how the San Diego Frozen Zoo had the foresight and bravery to start preserving DNA of species in 1972. They could have been criticized by (stupid) people saying we'll never have a gigabyte of storage. Even if you could it would be astoundingly expensive. These imaginary critics could have called up Gordon Moore at Intel at the time.

He talks about how someone who had wanted to do a proof of concept for recording DNA might have started with a single base-pair. His team is trying an analogous thing with *C. elegans* (a common model organism) and showing how they have memory of their environment being shaken and you can see the changes inside them. So hopefully they'll be able to preserve and then show that the memory change is still there. Really cool, but don't know how it would work as a startup.

Fun fact: Spices like vanilla and cinnamon have aldehydes in them.

The host asks about Egyptian preservation and luckily he knows, but it's sort of a related fun fact vs something he'd need to know for his work. It's like a doc getting asked a fun fact about the heart but it doesn't really have anything to do with their profession. This reminds me that it's probably worth it for people to memorize trivia related to things in their work just to make these kinds of conversations flow better, e.g. how much a brain weighs, or when dogs were domesticated, or the etymology of certain words or whatever.

He brings up a great point that humans are continuously being born into a world with more and more powerful technology but aren't necessarily being born with more wisdom. This is dangerous. His point is basically that x-risk is going up because of this.

I want to stop and mention that some people's solution to this is to be Ludditeesque and ban things they're uncomfortable with like genetic engineering. This overweighs the cost of action without thinking about inaction, much the way institutions like the FDA only take into account the risks of letting potentially harmful drugs into the market vs the costs of keeping helpful drugs from people. This is dumb.

How do we know that using glutaraldehyde in preserving the brain doesn't harm memory?

He does a good job of breaking down the problem: either some structure of memory preservation is so fragile that glutaraldehyde fixation destroys it but the structure survives all these other things like seizures, depolarization, etc. or it does preserve it.

He talks about encrypted information that you couldn't unlock would still be preserved because different things are still different. This is a semantic argument to me that isn't that strong. If you can never get the information it's not preserved to me. Sure, some super-advanced tech may be able to decrypt it but if different things are different but never readable, that's not preserved to me.

And sticking with the book example, only preserving the literal words does lose information which you may or may not care about e.g. the paper type, the font, etc.

Someone asks about recording all the activity in the brain, which is currently not possible, and he mentions an interesting idea of programming DNA to self-report what's going on.

The host tries to mention a movie and Robert keeps talking about the thing and the host immediately turns towards what Robert is saying. Good on the host.

The guy asking a question on 52:24 has a great voice. Can I hire this guy for narration and voice acting?

“Could you retrieve wisdom and experience independent of language?”

He brings up the analogy of a black box of a neural network that does something like telling dogs and cats apart. You can look at it and try to figure it out, but it would be difficult without running it. We may be able to unblack box things in the future too.

He says another thing that seemed to be minor but I think is a big deal. The easiest way to glean the wisdom from a preserved brain would be to simulate the brain and ask them what it was like to be there. I guess, but that’s kind of dumb. I could record them in 4k with my phone and boom, no need for brain preservation.

<https://twitter.com/JohnCGreer/status/1362562525678735360>

I get running a brain simulation would be way better but that’s like a trillion times more costly and difficult than just recording the person with the phone. It also reminds me that his concept of transmitting wisdom is really weird. If he’s implying we’ll all be able to upload our minds and meld with other people’s experiences then maybe, but that’s a huge leap that could have been addressed. Otherwise running a brain simulation and asking the person what an event was like is no different from the oral stage of history. And we can preserve that now with video.

Someone asks about ethical concerns and how he thinks about them. Ugh. I usually hate these questions because usually the person is on a moral high horse worried about those less fortunate but has an extremely poor set of ethics where they usually *act* like they care but don’t understand what actually leads to less suffering for people.

That said, he mentions they’re working with Anders Sandberg of the Future of Humanity Institute which is rad.

In response to ethical questions, he says a good way to think about it is to ask the question, “What’s going to enable human flourishing well?” which I think is great and helps clear up how to proceed ethically sometimes.

My translation of the rest of his ethics answer is that a lot of the ethical concerns are analogous to existing things like medical data and HIPAA, so privacy of information is important, and so are safeguards to ensure control and autonomy and having the information being destroyed if and when the person wanted.

What about the body? He argues quadriplegic people retain memory and personality implying the brain is where it's at. This reminds me of the argument against souls because when the brain gets messed up, their consciousness does too. One could argue the body is like an antenna receiving the soul that when damaged produces a messed up signal. Still, that's all bullshit.

Embodied cognition: he argues you have to have a body to learn how the world works.

He points out the brain is the hardest thing to preserve so if we can preserve that we can preserve the rest barring "a few stupid things that aren't worth going into". No, go into them! What and why, Robert?

He says it's a moot point because we can preserve the body as well. I say it's not moot if you can but aren't. It's maybe a moot point technically but it's not a moot point if you're not doing it. I'm not saying they should as it'd probably increase the storage cost by 10x or something but still.

At 58:00 he goes into a minefield of topics that have a long history of bullshit mixed with people making real attempts at solving them. Things like free will, souls, consciousness, and such.

He does give the caveat that he's trying a new argument so it may be less persuasive.

He's saying the hard problems aren't that hard or are mostly made up. He presents a simulation of a simple pendulum that exhibits a harmonic motion. He then adds a small pendulum to it and it creates a whole new chaotic motion.

I think I either don't understand the bullshit arguments he's refuting that people use to support the existence of free will or a soul or something. But I don't think he's really addressing things like the explanatory gap. I certainly believe it's all physical and there's nothing supernatural, but he's not solving or getting rid of the hard problem of consciousness.

This isn't my area of expertise and I don't know the history of its controversies or solutions or arguing it isn't really a thing. I'm curious as to what the QRI guys think.

His arguments sound like something Daniel Dennett would say, by which I mean they're kind of confusing. It could be confusing because I'm too dumb or it's not intuitive but it also may be there's some Eulerian going on. Mind you I only know DD as one of the Four

Horsemen and him arguing with Sam Harris about free will, so maybe he has tons of good ideas.

Wish they filmed after wrapping up. They could even have little go pros of people going up and talking to him, haha.

End Thoughts

I also just realized that he doesn't explain the title of his talk: "Engram Preservation: Early Work Towards Mind Uploading". An engram is a term for a physical unit of memory in the brain and mind uploading is the idea where we'll be able to copy the information in the brain and upload it to a computer.

This talk also makes me miss college and taking great classes.

Love people like this and engineers and such. Reminds me of how founding a startup and actually trying to build something is one of the best ways to learn the nitty-gritty details about a field. Vinay Gupta was like this.

Ultimately, really interesting, but I care less because it doesn't help me or my loved ones. Even if you could perfectly extract the information from the brain and upload it to a computer, I still wouldn't consider this "me". Why I wouldn't is a long discussion but is the same reason why I wouldn't take the transporter in the teletransportation paradox. For more on the messy subject of personal identity (which I haven't found a satisfying conclusion to, see Tim Urban's article: <https://waitbutwhy.com/2014/12/what-makes-you-you.html> and you can also play with some of the scenarios at <https://www.philosophyexperiments.com/>)

Alcor, the cryonics institution, discusses why his technique wouldn't help in reviving tissue: <https://www.alcor.org/2018/03/http-www-alcor-org-blog-alcor-position-statement-on-large-brain-preservation-foundation-prize/>

I really love and am sometimes manic about preservation so it scratches that itch, but this is even further away from someone being brought back from cryonics. By the time we get that information, AI will probably have either destroyed us or created a utopia for us where these types of things will matter much less. I care most about my loved ones and myself not dying and this doesn't affect that.

We seem to be biased to care more about losing an amount than acquiring an equivalent amount. So I would be more upset at losing \$100 than gaining it. Or more applicably, I

would be more upset at losing my current friends than gaining new ones.

This is related to status quo bias too, where we irrationally favor current circumstances because we'd be more upset by losing what we have than by gaining something else. Whereas if the situation were switched, we'd be worried about losing the other thing now.

Let's say I live in Jupiter, Florida with my wife Ann and my dog Bobo. I don't want to change anything in the past because it wouldn't lead to my current circumstances. Because if I hadn't gone bankrupt, I never would have moved to Florida living as a pool boy. But if I hadn't gone bankrupt and was married to Margaret and lived in Long Island with my cat Hazy, then I would feel the same bias in not wanting to change anything because I would lose what I had.

Outside of preventing the loss of people to death, I think there are good arguments that other things are more valuable, like putting effort into creating new experiences versus preserving old. As I've said, this is far from my natural inclination, I like doing both, but I do think it's important to address.

Still, it's a way cooler project and more important than what most people are working on.

It'd be nice to connect groups like r/DataHoarder to his work. I wonder what they'd think.

I don't see how it could sustain itself commercially because not enough people are forward-thinking enough to want to preserve their loved ones or their own brains. And those that are are probably more interested in cryonics. I would certainly consider paying for the service if I couldn't get, say, my dad to sign up for cryonics but could set someone up to preserve his brain instead. So I wonder how funding works. If it's just sustained on rich people who think it's interesting and are willing to throw some bucks at it.

Edit: Since writing the finalized draft of this I stumbled across this story written by Sam Hughes about brain uploading:

<https://qntm.org/mmacevedo>

I haven't read Robin Hanson's Ages of Em yet but if you find this area interesting you'll probably find that interesting as well.

