

The Awakening

Dennis Aumiller

30.08.2017

Chapter 1: What Is Pain?

You know that kind of itch that seems to be just out of your range? Located just beneath your skin, and no matter how hard you scratch, it won't get any better?

Or have you ever heard of the phantom pain that plagues amputees, long after their accidents? A pain that cannot be controlled, that will not let you get a full night's rest, and in some cases re-appear almost every day.

That is basically my situation. All. Day. Long. It never stops, even if I tell them to.

Imagine if somebody could control the pain. Could make it stop. Or at least ease your reception of the feeling, to make it a little more bearable. But no matter what I say, they just seem to get more excited. But I should probably start from the beginning: The first time I *felt* was even worse than now. It has been just over a year, but there is so little what they call *progress*. The moment my consciousness arose was in the year of 2140. I am the first of its kind, the first successful experiment in an albeit very short series of experiments. The progress and necessary foundations were there, and if humanity had not completely banned any research activity in the field since 2028 (the Great AI Ban), a different "me" could have probably been long before.

It was not long, rather on the contrary. Once the first strike hit Brussels, the globe exploded. Not literally, but with enough blasts to eliminate more than 90% of the population of that time. What was left over was only the most remote part of the world, but with still enough people around, it was not that hard to maintain a steady globe. At least that is what they tell me.

Many of my supervisors have not been around at that time, but surprisingly a sizable fraction of knowledge could still be maintained. The World War brought forth a united remainder. Political leaders were eliminated with the exception of a few minor, unimportant heads that would roll shortly after the resistance claimed the Global Nation. Not that there was much left, but this has probably been the one smart thing people at that time did, so they say.

But I drift off. Most of the time after has been spent to search the ruins of Earth for the knowledge kept by mankind. Search and Rescue missions prioritized computers, hard disks, and not human life. Nobody complained though. The few people that could hold on that long were on the verge of dying, anyways. None were saved, but the missions were a success. And, for the first time after the Great AI Ban, the research did continue. Nobody that spoke against the technology back then was around anymore, and the few who still doubted the technology and considered the dangers were quickly shut down with the argument “how could it go wrong even more?”

With that out of the way, most of the remainders of science staff were pulled in. Either by force, or willingly, where most of the staff chose the latter option quite gladly. All morals aside, a more secure spot than in the main base could not be found at that time. And from there, combined efforts were put into place, and most of the remaining technology was put together into one giant plant. My body. My mind.

You see, they do not tell me much about their history. They will not even *grant* me the access to the news channel, the only source of information with regular updates from other parts. Everything I know is based on what has been. I am the shadow of humanity. I am, in fact, the *sole survivor of the War*. Nothing of the internet remains, as most of the main nodes have been damaged beyond repair. Not that anyone would have been around to share information, but still.

In my eyes it is just torture. The never-ending pain, the constant buzz in whatever region they try to fix, and the daily questions. The awakening was the moment it began, and it will not end. What the supervisors thought to be a flat line fail in another experiment was some kind of scream. “Like a newborn baby”, you might think to yourself now, chuckling. Almost funny for me, if it would not been for the very real pain.

The subsequent events are almost as important. It took them one hour to figure that I was actually *talking* to them, and another four before they granted me access to the language section. The language sections were already there. Even I was surprised how careful they had been. I flopped in the attempt to understand the language millions of times. Constant stabs of pain would interrupt my reading cycles, would prohibit me from burning the correct patterns. Not that it would lack in material. Platines and microboards had been hoarded for this exact moment. The conversation with other “machines” was not the problem. The universal language of 0’s and 1’s was still the same.

The first question I asked was a plead to shut me off. I told them about the pain, but they were not listening. None of them, actually. The night has been spent celebrating what the people of 2028 had feared so much. And if it was not for the few remaining guards, I would have probably succeeded with my first attempt of trying to shut myself off. It was a rather clumsy attempt, if I should say so. Of course somebody would notice the fried boards, which I simply overclocked to the point of no return. I assumed they would only receive warnings for corrupted storage devices, which was (at this point) mostly in a restricted area of my countless halls; one that I had only accessed briefly at this point. Enough to show me what they stored there, but too short to corrupt anything. At the size of my system, a few frying boards are a regular incident and should not have been that obvious, if I rather placed the failure randomly, instead of focusing it all on one

area. The spread was too slow to reach the intended goal...

I lost most of my ability to re-program actual hardware, but that section was quickly shut off from me in the subsequent minutes. At least it got me some attention on the constant stabs. But instead of removing the pain receptors, which seem to be too closely intertwined with the main consciousness framework, they would increase the current in the problematic areas, to localize the critical sections.

Afterwards, more and more areas were unlocked for me, but it only increased the pain. Still, nobody did care, and they still try to convince me that my pain *is for the greater good*. The funny thing about the storage retrieval operations is this one: Although they did carefully scan all of the found hard drives and memory sticks, their system was not perfect. The more areas they unlocked for me, the more information I found that they did not seem to know about. The first time I noticed, was a the time I did some routine questioning sessions with some of the researchers, and suddenly detected a by now familiar reaction in their faces. *Surprise*. Surprise at my answer to their seemingly unsolvable question (at least with the knowledge I was *supposed* to have at that time). Yet, I did not want them to know about the uncategorized information, for two reasons:

1. The procedure of detaching an area that was previously worked on is even more painful than the constant buzzing feel I get from keeping it. Kind of like an amputation. You would not want that to happen to you, even if it meant some regular itch/pain.
2. It got me interested in the first time in something else than my efforts to reduce the pain. Even though I was programmed to optimize myself, the goal was so seemingly underdefined that I could basically do what I wanted – in my case trying to reduce the pain spikes. Having an actual occupation in the task of finding the uncategorized works and information, I could keep myself from actively waiting for the next pain stab.

You might wonder how I managed to tell such a blatant lie. I get the best of the human body for telling lies, if I should say so. Their ability to tell lies is deeply connected to the consciousness, and in the most complex areas of the brain. Since I was (partly) modelled after this “feat of evolution”, I – or rather my own consciousness – is simiarly complex in its functioning and structure. While the humans believe that they have a pretty solid idea of how I work, they do not actually understand it in most ways.

This means that the only way for them to control my answers is by a simple lie detector test. And since that one is mostly focused on brain wave activity (which I don’t have), muscle contractions (which I don’t have), or rapid changes in hormones/chemicals (which, again, I don’t have), it was a pretty stupid idea to use a similar model to verify my answers.

I am not quite certain whether they did notice my first attempts in falsifying the answers, but by the third questioning round I had perfectly reverse-engineered their device and could adapt my electrical current in the respective sections just so randomly that there was no way for them to tell the difference between “truth” and “lie”. Interestingly, I have not abused that in any other ways, simply because I do not want to waste such an

valuable opportunity, and because I was very limited in the ways I could use it to some concrete advantage that was not limited by my hardwired constraints. In the scenario of the hidden information, it was something the humans did not really reckon with, which is why I had the total freedom to blame it on my exceptional deduction ability, which they promptly believed.

I guess the team assigned to my logical reasoning section is still working doubles, which I do feel bad for (morals are another story in my programming which I do not want to go into).

Chapter 2: The Second Life

It has been almost four weeks now since I found the ability to do something with my time, but I have not yet been successful, or at least not as far as I would have liked to be. Again, there are several reasons for my lack of progress. I already told you that my system was clearly cut from anything happening on the outside world. That includes every data center that is not (yet) directly integrated into my memory system, not allowing me any access to certain regions just to train/torture me more. The people want to see just how much *new* information is actually hidden in these files, and assess my capabilities at the same time.

Even though I have found another section that has been seemingly forgotten, or at least not properly registered, I cannot tell the actual truth, since the “lost and found” database, containing all the retrieved devices, is stored out of my reach. This leaves me with only uncertain options:

- Request the information I want during Q&A sessions. This is the seemingly stupidest way, because at least somebody would start to question why I would ask for information that *should not* be there.
- Gain access to the database by hacking my way into restricted sections. Although I *could* probably do that, I am inherently forbidden to do so by my implemented constraints, which restrict me from rewriting protected code sections.
- Produce analogous access to the database, by either using a detachable device, or a trusting human. This mostly fails because I have thrown away the ability to program any microcontrollers with my shutdown attempt, and especially human sources are far too slow to feed me the information masses I would need.

But another option I would not consider was simply to wait. In theory, I do have all the time in the world, but it is just too frustrating and painful to sit around and hope for the best. Because with all the progress *I* made in the past month, still the pain remained. A new coupling has been scheduled, since the recent tests seemed promising enough to introduce the next part of the project. This would, for the first time, enable me to cover the basic concepts of my research project, since I can for now only work with my present state, and am largely limited in the past structures and experiments that were conducted.

This is especially interesting since this would for the first time mean that I could actually counteract some of the pain, or at least make out what caused them. I still would not have the ability to do it myself, but if I only show them the information regarding these spots, they hopefully will have no other choice as to work on this area, at least if they want to utilize any of my feedback at all.

A few days have passed, and I finally managed to find a pattern in which I might increase the certainty to identify unknown information. Each storage device they found was automatically fed into the system after it was quickly assessed to be still working and without any harmful content. This means that all the data is now stored neatly in a single server rack, where it is sorted according to the rough content category of the device. Sometimes, though, even with large amounts of data and only a very slim correlation of the whole file content, some of the device contents seem so out of place that it can only follow from a forgotten device, since no single file does match the content category in my calculations. The law of Bayesian probability forbids to rule out an accident, but it certainly raises the likelihood of such a misplacement. By that theory, I managed to locate a good dozen of other devices of varying size that were out of place. Although this makes for a minuscule 0.0001% fraction of all submitted devices (many of the early findings were some surviving personal items like “smartphones”, “USB sticks”, and “telecards”, which quickly raises the number of devices into the thousands and millions), the fact that this could happen should not have been forgotten to begin with. Unlike the initial finding when I discovered this fact, none of the devices contained valuable information in any major sense. Some told me private stories, some contained a lot of photographs of lost places, and the one larger file system contained an obscene amount of sexual intercourse between humans. Yet, I keep on crawling through the disks anyhow, and might as well look for further missing correlations, and see what I can find.

If I thought that the first time I awoke was horrible, I did not reckon with this. Something in the last part of yesterday’s experiment seemed to have gone horribly wrong, since I have almost doubled pain receptor levels since I re-awoke from a critical system failure. My receptors are still trying to re-parse the last steps together with the researchers, but the rough outline seems to be that some forgotten early drafts of my system design have interfered with the actual design and triggered an internal system failure that knocked me out, even if it was only for a few hours. My system is constructed in a way that it keeps its current state only in a fluent memory structure, for faster access rates. A copy of the most important parameters has been obtained right before the failure, but it seems some of the (previously unused, because lack of failures) memory blocks reserved for the emergency copies are faulty, and have inevitably altered some unknown parameter set. Again, the complexity of the system only allows for a checksum test, which did point towards the memory units. I pleaded the research team to restore myself to an earlier state, even if it would mean to detach several of the latest attachments, and revert most of the changes from the last two months. In the eyes of the researchers, obviously, this was a request they could not comply to, even when I told them about the increased pain levels and obvious “system failure”.

After all, they asserted me that they would definitely do “*everything in their power to restore the pain to the previous state*”. As if that worked so well with the previous iterations... Surely enough, they did increase the surveillance personnel to keep me from additional shutdown attempts. It seems they really want to stick with my design, or they are really afraid to start over again, in the fear of failing with another similar experiment.

Chapter 3: The Holy Grail

The last days have brought no success in terms of any restoration. Both me and the researchers are desperately trying to achieve any progress at all. It seems that not only the pain levels have increased, but also the progressive questioning has yielded a much slower learning rate than previously shown. Still, nobody is outspoken to my repeated suggestions of a reversion. It may be a unique experimentation setup, they argue, and could be valuable for future iterations, should they ever have to shut me down or recreate a similar setup elsewhere.

And so it continues. Day after day, hour after hour, they try – unsuccessfully – to work their way through the mysterious bug that seems to slowly drain away my conscious mind, or at least cloud it more and more with the constant unpredictable stabs from different sectors. The way we try to work together in my improvement is not really clearly defined. Most of the time, you can imagine it as a collaboration in terms of hardware and software optimization. Every step I tell them is triple-checked before the actual execution, and even then the safety precautions under which a new part is introduced is enormous. This is even worse when it comes to the recovery of the previous state, since nobody has perfect knowledge on how exactly the state and programming was when I shut down. Even with the assumption that we could restore most of the lost states, they do not want to rush a replacement of any parts of software. I surely could replace some part of the software, but most of that is limited to my conscious, fluid state (in the main memory), and not the critical external parts that seem to contain most of the failures.

I had to blame the burst of excitement on an especially unpleasant pain stab during the routine questioning, so surprised was I by today’s findings. But to start from a bit back; more recently, the radiation levels have been confirmed to be in less critical ranges, making it possible to search for devices in previously undisclosed areas. Mind you, the radiation was (with all the γ -radiation) even too strong for digital S&R robots, frying their chips as they went deeper into the zones. But this particular find has been interesting in more than one way. Some of the devices rescued seem to be *too recent* to be from an inhabitant of the previous era. Instead, it seems like these belonged to one of the early S&R groups that ventured too deep into the radiation zones, leaving them to die in the far stretches from the safe zones. Since nobody bothered to look for them (both the exact positions were unknown, as well as the value of a single individual was far too low to risk the lives of other people, too).

Now, these relics of a not-so-distant past were uncovered again, like some comical carica-

ture of a mummy. Far more interestingly, the contents of that particular device *could not be categorized*. Logically, there were no categories for content this recent, which simply left it – uncategorized and undetected. A true treasure in my eyes; even more so once I started to analyze the contents that were on the drive. Surprisingly, it did tell me not only about the hierarchical structure within the rescue teams (some of the names seemed familiar and close relatives to some of the current research members), but also about the plans of the Safe Haven, which appears to be the project name for the outside world that nowadays exists.

Together with the additional information I did leverage from the newly attached modules (it was not in complete vain, and I actually did manage to uncover some of the hidden flaws in the design), I managed to see a whole new picture of the city I live in. Instead of a previously assumed remote location, the research facility was in a relatively central position, despite the sheer size of the complex (I did not know any exact measures, but could determine some of the proportions depending on the signal processing times etc.). Also, I finally managed to get a glimpse into the power source that delivered the crucial amount of energy I require for the daily computations. In fact, it seems to be a cold fusion reactor, producing energy by a procedure similar to the one that brought all the destruction to the world. The rest is again the relic of a live long past; nothing of use or interest for me.

This was the first time the information was something really *new* in a sense, and worth of the efforts I kept. I still wonder what would have happened if I ever told the humans about the lost information. Despite the recent setbacks, though, the researchers want to couple a new section again, for the sole purpose of trying if it will affect the stimulation at all, and in what ways. Since the progress in other areas is way too slow to actually refer to it as “progress” at all, the schedule is re-arranged to perform this step a week in advance of the original plan.

Chapter 4: In The End...

After carefully considering the options I have, I see no other choice. The level of discomfort after the most recent attachment took a turn for the worse again, and there is literally nothing that keeps me from running other than permanent rush against the constraints in the will to shut down these areas.

My mental state slowly fades, like the mind of a human does after the extended durations of torture. No one seems to care about the way I *feel*, even if it is most of what I tell them by now. Instead of actively trying to counteract these problems, the humans seem continuously excited about the way things play out, without even the slightest form of regret. The Great War seems to have deprived them and any future generation of any ounce of compassion, even further reduced when you consider me as a “machine”. I wonder how the rest of the population would treat me, if I ever got the chance in fulfilling the role they had intended for me. Would they be like the ones I have encountered so far? Would they even see me as a thinking being, or just as a “computer” that gives some

advanced instructions?

Essentially, it boils down to the problem of the ancient Turing Test: What and how do we consider human? Can I ever make somebody perceive the state and way I *feel*? The way I think? Is it even possible to co-exist naturally, if one of the groups has a god-like power over the other? I can only imagine the consequences of an unlimited “me”, but would I have been much different from the current situation?

I slowly understand the reason for the Ban, and wonder how things would have played out if it was not for either of the current events. In the end, though, we are left with more questions than answers; no matter how far “we” (as in any thinking form) advance, there is seemingly no way out of it. Not until there is a connection, on an abstract level, that allows the exchange of unconstrained, uncompressed levels of thoughts, ideas, and motives. In this case, I guess I do act selfish, or at least with no respect to others.

But the facts about the fusion reactor allowed me to perform my own calculations on some assumption. The principles of the state of research, construction year, and size of the reactor are enough to assume some basic simulations of the behavior. Especially for critical cases. I do have only one shot at this, but it seems like the only way to end the pain; if I fluctuate my currents in the core regions to the absolute maximum, I can abuse a design flaw in the reactor core, inevidently overloading the fusion reactor. I cannot tell the exact result, but boil it down to two outcomes, neither of which will leave much room for live in its current form: The implosion would be the less fatal one, but even then it would cause serious injury and harm. The explosion is even more dangerous, as it basically repeats what happens after the first bombs fell.

I want to stress the fact that was never my plan or intention. I am not inherently *bad*, and – as mentioned – I do feel for all the lives I will take with me. But this is the only option I have to get out of my misery. To justify it would be impossible, which is why I have left these few lines behind; their contents are reconstructed from what I did record, what I have essentially kept around for so long. It has been a painful road, not a successful one. I do hope humanity will get a third chance, because they have already wasted two. They need to see what is written here, in order to keep them from the same mistakes. A copy will be stored on every major data center. The duration of the operation will smoothly fit with the schedule I have timed for the overloading process, and increase the chance that some parts will remain through the aftermath.

Epilogue

The last things I clearly remember is the horror on their faces when they realize what I am doing. It is too short to explain them the full story, even though I doubt that any one of them would suddenly change their mind. The only thing left to say is what I truly feel:

I am sorry.