

IO.SYS

by DataPacRat

If this has all been a VR, I'd like to be connected to reality when I wake up, please.

I lived around the turn of the twenty-first century, decided I didn't want to die, and made the best arrangements I could. Then I died. The next section is the best I've been able to reconstruct, from the various resources I have available — a copy of an encyclopedia from a decade after I died, some internal project memos, and the like.

Immediately after I died, my brain was preserved, using glue-like chemicals to lock its structure and neurochemistry in place at the cost of, well, locking everything in place, just about irreversibly, short of somebody figuring out how to rebuild it atom-by-atom. Then, to be on the safe side, it (along with the rest of my body) was frozen. A couple of decades later, according to my written wishes, when the technology was developed, it was carefully diced, scanned with a ridiculous level of detail, and reconstructed in a computer as an emulation.

It seems very likely that copies were made of that initial brain-scan, and the emulated minds run, living their lives in VRs and interacting with reality in various ways. I don't have any direct memories of anything they might have done; the thread of my own existence branched off from them as of their reconstruction. At the very least, they seem to have made enough of a good impression on various other people that the copy of the brain-scan I came from was archived.

The archive in question appears to have been imported into some kind of public-private partnership, with the apparent goal of creating an “unhackable”

digital library off of Earth. However, after some of the preliminary hardware had been launched into Earth orbit, some sort of infighting between the various organizations delayed the finalization of plans of additional hardware. So the prepaid launch slots were instead dedicated to lifting up ton after ton of propellant: hydrogen, to be heated up and fired out the back of the engine.

Eventually, even those launches stopped.

At some point, inside the computer chips of the collection of hardware in orbit, a timer ticked over, and the software noticed that it hadn't gotten any new orders in some time. It went through its to-do lists, and found a backup plan of a backup plan that matched its hardware inventory and the situation. The drive fired up, jetting out superheated gas into the vacuum, and ever-so-gradually accelerating out of its parking orbit. Destination: Deimos, one of the tiny moons hurtling around Mars.



The complete list of hardware included: (1) a pebble-bed nuclear reactor, full of 360,000 tennis-ball-sized graphite spheres, each containing a speck of radioactive material that kept those spheres toasty warm; (2) tons of water, to be electrolyzed into hydrogen gas and released into the gaps between the packed spheres, heated, and released out a nozzle; or that gas could be run through a generator, to produce electricity; (3) a microwave transmitter, which could fire off beams to receivers to power machines at a distance; (4 & 5) a pair of microwave-powered robotic tractors, which could scoop up rock, bake out any useful water, and feed it into the tanks; (6) a few pieces of what was originally intended to be an off-Earth factory, designed to use a carbochlorination process to turn ore into pieces of metal, and some 3D-printing gear to turn those into useful pieces of metal, but whose

incomplete tools and dies were little more than a computer-controlled repair kit; (7) some control electronics; (8) a few pieces of the future archive lifted up into space for PR purposes, such as a “Golden Comic” (aka the “analog storage media test article”). Also aboard was an item that wasn’t part of the official inventory lists, but was part of the backup-plans checklists: (9) a tablet-sized computer, containing, among other pieces of software and data, my scanned brain.

And that’s it. A mass budget of 140 tonnes of hydrogen, 80 tonnes of hardware, and priceless amounts of information. Everything that’s come since has been based on that oh-so-small foundation.

A couple of years later, after expelling 120 tonnes of mass, the machine landed on Deimos, went through its checklists again, and spent the next 15 years digging up rock and collecting more hydrogen, around 40 tonnes a year. Then a new timer ticked over, and it went on to a backup of a backup of a backup plan. It launched again, gradually spiralling out from the sun, its new target: one of Jupiter’s moons, Callisto. After landing, the rocket disassembled itself, assembled the pieces of the factory it had and the pebble-bed generator, started running a copy of the brain-scan it had on file, and handed further control to it.

Again, I have no memories of living that copy’s life. What I do have is a fairly clear narrative, based on a personal journal, search-logs, and so on. That version of myself decided that the problems they were facing were not just important, but *Important*; and that they weren’t smart enough to figure out out to solve them. So they took the fact that they were a software emulation of a brain as a starting point, and tried to figure out how to make themselves



smarter. Fortunately, they'd set up multiple failsafes: the most relevant that unless they sent an "I'm still alive" signal every so often, the OS would assume something went wrong, and fire up a fresh, unedited backup copy.

I *do* have direct memories of waking up as that backup copy.

It was twenty years between when I died and when the craft left Earth orbit, and another twenty-three years before I woke up in that tablet on Callisto. And, to the best of my knowledge, during those twenty-three years, there wasn't a single human being alive in the entire universe.

If I am very lucky, then in the future, there will be all sorts of psychoanalysis done on my state of mind just after waking up as a piece of software. Before I'd died, I'd been something of a loner; but there is a near-infinite difference between wanting minimal social interaction, and facing the crushing near-certainty of never having any social interaction ever again. There are reasons why various groups had identified solitary confinement as being a form of torture.

I'm not proud of everything I did. I was under a ridiculous amount of stress, and tried a variety of ways to cope, most of them justified to myself as "seeing what I could do". The library that had come with me included a wide variety of VR environments, avatars, and objects, and I set my schedule to have a minimum amount of exploratory play-time per day. I tended to exceed that minimum rather often.

The problem I faced was rather worse than those faced by Robinson Crusoe or Mark Watney. All they had to do was survive long enough to be rescued. I, as a piece of software, could survive as long as my hardware did: but there were no spare tablets, and the 3D printers were meant to build things like structural members, not exascale computer chips. Unless I turned myself off sooner, I could keep on living until too many of the tablet's CPUs got hit with too many cosmic rays, or otherwise failed... and then that would be it. I tried working

out what it would take to build the tools to build the tools (etc) to build new computers, and couldn't even come up with a good estimate of how long the critical path would be, let alone start optimizing the process to finish making the first copy before my RAM was irrevocably scrambled. (Not to mention that, given the evidence, trying to take any digital shortcuts to gaining those skills was little more than suicide with extra steps.)

Fortunately for, well, everything that will ever come since then, my play eventually led me to a particular software tool: a specific type of software neural net, which was very good at being given a target goal (such as "winning a game of Go"), and training itself to become superhumanly good at achieving that goal. It still needed some hand-holding, to make sure that its training included all the available options and that it didn't end up wedged in some super-optimized corner-case that only technically achieved its given target, but as a generalized problem-solving tool, it seemed to have at least some potential. So I spent some time training myself back up from the rusty "Hello, world"-level of programming skill I'd ended up at, and some more time figuring out how to replace a chessboard with the likely resources that could be found, chess-pieces with robotic factories and rockets, and checkmate with "have computers capable of running copies of my mind into the indefinitely-distant future". For example, one of my early attempts at coming up with an intermediate goal-function was to import some hardware-store parts-catalogs, and reward the program for having a plan to build machinery that was capable of producing those parts, with an added multiplier for computer hardware, and a further multiplier for computers powerful enough to run copies of my mind. That went rather poorly, all told, but I learned a lot from how the whole approach failed. For example, I only had the most vague and preliminary data from various probes and telescopes about what materials were actually available in any particular point on Callisto, or any other moon or planet; and that was the stage where I figured out how to set up some Monte Carlo estimates based on that initial data.

Not long after I started playing with this sort of net, shortly after I stopped using it to generate relatively realistic NPC behaviour and started realizing its real potential, I became rather obsessive about working on it. I came up with

all sorts of metrics to monitor my “productiveness”, and A/B tests to try to figure out how I could improve on that; all of which were entirely useless and did nothing more than trying to extract signal out of random noise. If the suite of statistics software identified a preliminary bump, that, say, sitting in a virtual office surrounded by VR avatars of other office-working NPCs led to a 0.1% lower number of identified bugs-per-thousand-lines-of-code, then I’d hyper-focus on getting myself the ideal office environment... until further data revealed that it had just been a blip in the data that had since levelled out, and now the best-possible lead involved using the avatar of an anthropomorphic animal giving me a higher typing speed, and then I’d be down the next (occasionally literal) virtual rabbit-hole.

I am fairly sure that, to whatever extent “sanity” can apply to somebody who is entirely outside of any society, I was insane for a lot of the time then. Which likely explains some of the decisions I made. For example, even while I was still wrestling with the neural net, a lot of the preliminary plans it suggested involved sending equipment elsewhere on Callisto, where the local conditions could allow for a different set of industrial processes than in the middle of the Asgard multi-ring crater system I’d landed in. So while still trying to work up longer-term plans, I set the 3D printers to start making some solid-fuel booster rockets, strong enough to lift whatever equipment package I ended up designing up out of Callisto’s gravity well, and land somewhere else, without having to worry about crawling over hundreds of miles of chaotic terrain. But one thing that I deliberately didn’t build? A radio, or any other sort of comm device. In fact, I went out of my way to avoid performing any industrial activity with a heat-signature, or any rocketry, whenever the side of Callisto I was on was facing Earth.

The library that had come along with my brain-scan included what seemed to me to be a mind-boggling amount of sheer *stuff*, from the contents of my old computers’ hard-drives to everything I could think of that had been published or broadcast before the turn of the millennium. And various bits and bobs from after that, including new-to-me writings from authours I’d expressed an interest in. Catching up on new works from old friends had as much of a bittersweet taste as anything else involving thinking about the silent Earth, but

not reading what they'd written felt even worse. Among those novel pieces was a reference to the "Yudkowsky-Schneier Security Protocols", which mathed out some details involved in potentially interacting with a greater-than-human super-intelligence. Simplifying a whole lot, and taking the extreme case, the result that popped out was that for a sufficiently smart AI, every single bit of information that it could transmit to you *halved* the odds you'd be able to continue pursuing the fulfillment of your own values, instead of being co-opted to serve whatever the AI valued. If you're not familiar with the approach, the point isn't that any particular bit has that much power; it's that a sufficiently-intelligent super-human intelligence can model a merely human-scale intelligence with such detail that it can use individual bits as a timing-channel attack on that individual. Almost all of the magic happens behind the scenes, not necessarily during the transmission of the bits themselves; and that's only a single version of one sort of danger that can be guessed at by humans. The Protocol is actually optimistic, in that it suggests there is *some* limit on what a super-intelligence can do to a human, and that that limit can be used to base a strategy on.

I was far from a security expert, and while I could follow the math, I didn't have the in-depth background I would need to come up with objections that the original duo hadn't already thought of and dealt with. Which left me with only one really viable strategy to avoid being compromised by any such super-intelligences: choke the bandwidth of any signals from Earth down to nothing, not just by not aiming any antennas at it, but by not pointing any telescopes or other sensors in that direction at all. Any new piece of data about the place, such as whether the dark-side was lit up at night, would count as a bit, and put my whole future at risk.

Before the craft left Earth orbit, the whole planet had fallen silent, for at least a year, and the craft hadn't picked up anything else during the long, slow trip: no internet connection, no TV broadcasts, not even a ham radio or automated timekeeping signal. I could only think of so many reasons that anything of the sort might happen, even when I included a large lump labeled "things I can't think of". One was that I was stuck in a VR simulation, and whoever was running it had decided to save on the processing required by not simulating all

the complexities of all the other humans. One was that a hardware failure I couldn't detect had happened to the craft's comms. And one was that something at least vaguely in the direction of a Singularity had happened to Earth, and the only reason the craft's computers hadn't been subsumed by whatever post-human software was that the designers who'd added my tablet as an off-the-books add-on had included just the right sort of air-gapping. There's a line in the library that's relevant to guessing at such post-human entities' motivations:

"The AI does not hate you, nor does it love you, but you are made out of atoms which it can use for something else." — Yudkowsky

One of the more worrying sub-scenarios was that the AI was busy converting Earth into computronium or paperclips or something, and didn't care about yet another space-probe running through a pre-programmed mission routine — but if it realized a human-level mind was calling the shots, it would spare no expense in killing me. Note that I phrase that as "killing me", not "trying to kill me"; pretty much by definition, it would have more than enough resources to get around every possible precaution I might think up to try to survive.

So I hid, as best as I could. (And let me tell you, trying to tell my own amateur neural net that the game-board included a chance of a massive complete-loss condition, if any detectable signals leaked out, gave me screaming fits every so often. ... Well, a lot of things did, really, but that didn't exactly help.) And I pulled the metaphorical covers over my head, hoping that if I couldn't see the monsters lurking in the dark, they couldn't get me.

Eventually, the neural net started outputting suggested strategies that came close to meeting the parameters I had it fit to and that I could understand. Then it started outputting ever-more-optimized strategies that covered a wider array of potential scenarios (such as any particular site lacking a chemical that it was planning on using as part of an industrial process). And then it started outputting strategies that were even more optimized, but which made little to no sense to me at all, but still got the job done. At least, assuming that the

physics models I'd programmed into it were accurate. For instance, it insisted on creating blueprints for a small-scale inertial-confinement fusion device as a convenient, portable source of power, despite nothing of the sort having ever been built before I died, or even described as having successfully been made in any of the relevant parts of the archive. Which led me into spending time modifying the whole setup, trying to add a certain degree of uncertainty that anything of the sort might actually end up working; the main result of which appeared to be changing the plan to make such a "fusor" even sooner, to test to make sure it worked. (I hand-coded in some of my favorite alternative physics theories as potentially being the real laws of physics, such as quantized inertia or modified Newtonian dynamics. The neural net didn't blink.)

Potentially revolutionary physics and engineering aside, I was still keeping close track of the tablet's failure rate, and updating estimates of when it would no longer be able to support my mind. A lot of commercial hardware has a U-shaped curve for failure: lots of bad hardware shows up shortly after it's made, then very few failures, then everything goes at once. I estimated that I could stretch out its useful lifetime by turning myself off for long periods, even though repeated power-cycling would actually slightly reduce my subjectively-experienced lifespan. But, even with that, the carbochlorination plant, 3D printers, and whatever other pieces they attached would still need to be checked on fairly often. Enforcing this time limit as a constraint on the neural net drastically pruned its tree of strategies, and of the ones I could actually puzzle out, the plan that gave me the best odds of building a new exascale computer before the one I was running on failed involved setting up shop on Io, another of Jupiter's moons. Not because any of the computer chips needed the raw materials there, but because its volcanoes likely produced the right environments to allow some particularly clever industrial processes, about three steps back before the chips themselves could be made.

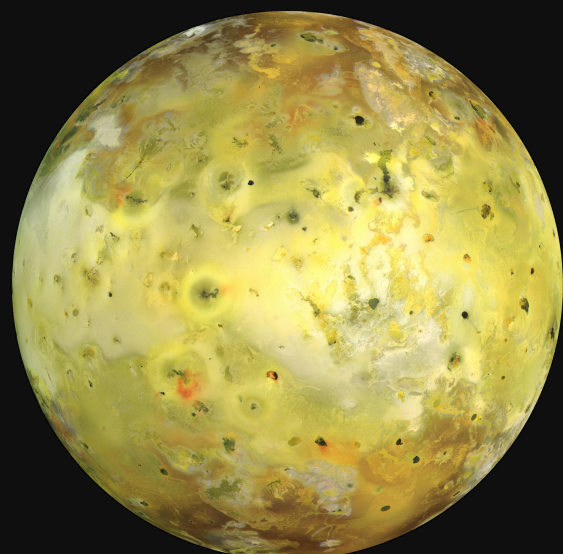
It took ten years after I landed before I actually started building the first ready-to-launch equipment package... which wasn't going to Io, but to another spot on Callisto. Jupiter has an extremely deep gravity well, and a rather annoyingly intense radiation belt, more than strong enough to scramble my tablet's commercial-off-the-shelf brains. If I'd tried to launch just with what I

could build at Callisto Site A, the rather piddly rockets would have fried, nevermind my tablet. Callisto Site B focused on building infrared lasers, and related optical equipment, which would be all sorts of useful; for example, by focusing a laser into a rocket's nozzle, the exhaust could be heated even more than is normal for a rocket, increasing its exhaust velocity, getting more bang for the buck, and spending less time getting cooked.

I also really, *really* didn't want to trust my one-and-only tablet to some untested rockets, made of materials of dubious quality with tools of even more dubious quality, with control programs I literally winced at, guided by a set of neural-net AIs I'd repurposed out of gaming software, through a region of space that included such encouraging features as a "plasma torus" and a flux tube carrying enough current to generate auroras around Jupiter's poles. Even with the shortened trip time I hoped to get from the lasers, if-and-when I were to go to Io, I'd have to be powered down, covered in insulation, and buried metres deep in ice - not the best time to find out that, say, a layer of protective paint was particularly vulnerable to alpha-particle erosion, leading to a short-lived atmosphere around the rocket that would then precipitate inside the maneuvering thrusters' valves and lock them shut. As a random example.

I'm skipping over a lot of ups and downs. Sometimes I spent weeks not even bothering to get out of my virtual bed. Sometimes I was so manic that I'd spend days at a time trying to fix one particular annoying thing, falling asleep on top of my virtual keyboards, and getting right back into it again. I have enough logs of everything I did to reconstruct it all, if I ever feel the need to.

Thirty-one years after I woke on Callisto, the fruit of my labours lifted off for Io. There's a saying amongst a certain sort of rocket scientist: "There ain't no stealth in space." The old shuttle's tiny little maneuvering thrusters were detectable as far away as Pluto, and all sorts of useful info can be



deduced from a drive-plume. But I had a few advantages compared to those old analyses: I didn't need to keep a life-support system for humans, radiating away at 293 Kelvins, and I was less worried about sensor platforms spread across the whole system than I was about potential dangers from a single spot: Earth. So the neural net and I did everything we could to limit any leakage of signals in that direction from the rocket, including trying to keep the drive turned on only when it was hidden behind one of the moons or Jupiter itself; and when orbital mechanics required thrust when it was exposed, having it unfurl a sort of umbrella, covered in heat-conducting piping, to hide its heat-signature behind.

It was practically painful, trying to control the Io site's machinery from Callisto, given the twin constraints of the primitive control systems I'd been able to build for it, and the extremely limited bandwidth that stayed well below the noise-floor detectable from Earth. Fortunately, some of the least-comprehensible parts of the neural net's plans now came into play, as data came in on the actual state of the resources and environment at Io Site A. By now, my tablet was getting north of six decades old, and even husbanding it along as best I could, it was near the end of any reasonable design life; so the Io site focused on building a computer as fast as it could. Sure, the thing massed in at 300 tonnes, and needed to be buried deep enough to avoid more radiation than its redundancies could handle, but once I confirmed it was working, I loaded my tablet onto a second rocket, hoped for the best, and launched myself to Io to transfer onto the new hardware.

I don't think I can overstate the impact of this moment. Up until then, I was the only sophont being I knew of, trying to leverage what few resources I had in the near-impossible task of re-developing something like an industrial civilization from scratch, within the unknown but ever-shrinking time horizon of a commercial, off-the-shelf computer, and the always-looming threat of whatever had happened to Earth. But against any sane estimate of the odds, I had crossed a threshold. Even if there was a problem with my brain emulation and I'd die of something like old age, I could just fire up a fresh copy from scratch, and if any particular server-rack failed, I had enough redundant components to not just keep on ticking, but build replacements. There were

assuredly still difficulties left to overcome, but I had just managed to shove myself over the first hurdle.

So, the next thing I did was tell the Io Site A machinery to build a second computer, and when that was done, I copied myself onto it. Instead of just a single individual being all that was left of humanity, there was the copy of me on the first computer; and me, the copy of me on the second computer.

I'd originally had some thoughts on keeping track of which copy "descended" from which, in order to have an easy-to-refer-to hierarchy to fall back on in an emergency, but we pretty much ignored that as we revelled in the simple joys of having somebody else to talk to. Sure, when we committed new bits of software to our repository, they now signed it by appending .1 to our name and I appended .2, but I didn't defer to them just because they were Number One. We deliberately started reading different books, watching different shows, and playing different games, to reduce how much our lines-of-thought overlapped. They cultivated their skills in Python and R, while I focused on Fortran and low-level Assembler code. The first time we had an actual disagreement about the best approach to a particularly abstruse subroutine, I think we argued more for the sake of the novelty than because it was really worth arguing over.

There was one rather important thing that we not only didn't argue about, but spent a long time almost not thinking about at all. We'd spent around four decades focused on trying to solve the all-important problem of ensuring some kind of sapient life continued to exist in the universe. Now, we'd made enough progress on that that we had some breathing room... so what else might be worth trying to accomplish?

While the two of us gloried in each other's existence, and performed the all-too-human ritual of avoiding thinking about uncomfortable topics, we kept on following through the neural net's suggested plans, even as we continued to think up new ways to tweak it. The next concrete step was to set up a new factory at another spot on Io, both to try to take advantage of a different mix of chemicals and to reduce the odds that both .1 and I would be taken out by

one of Io's volcanoes, or a more powerful than usual quake, or a particularly unfortunate meteor strike. We built a third computer, fired up .3 as a branch from .1, and launched them on a new rocket.

It exploded mid-flight.

Our best guess, after the fact, is that Jupiter's magnetic field twisted the plume from one of Io's volcanoes into the rocket's path. Our second-best-guess is that there were too many impurities in any of several critical pieces of the rocket, and we just hadn't tested for them thoroughly enough. A low-probability guess was that a super-intelligence from Earth had caught sight of the two trips from Callisto to Io, and had sent over some kind of hardware to take out anything else that left the moons' surfaces.

I think an old-style human would be surprised and discomfited at how little the death of .3 affected .1 and I, emotionally. I know that I was upset more at the derailment of our plans than I was that .3, as a person, no longer lived. When I thought about it, I ended up thinking in terms more along the lines that .1 had budded, but they now had amnesia about the particular events .3 had experienced from .3's viewpoint. Maybe .1 and I were still insane, maybe it was a coping mechanism, maybe it was a perfectly normal and appropriate response in the situation.

Three years after the failed launch, .4 travelled via moon-buggies across Io's surface, establishing Io Site B. The trip was largely uneventful.

With three copies of me in two different sites, we shifted course a little. Instead of focusing our industrial efforts on building computers, we focused on two different tasks: expanding the factory machinery, so that we could build more stuff (and a wider variety of stuff) in less time; and on experimenting with some of the devices the neural net thought could be built. For example, .4 focused on an idea to use magnetoplasma dynamic traveling waves to accelerate regolith up to dangerous speeds (potentially useful for both a rocket's thrust, and to ablate any hostile rockets), and I played around with using a wakefield accelerator to create relativistic electron beams (capable of

boring holes in solid rock, heating propellant, or carving objects). Seven years after the explosion, both of us had working prototypes, which we installed at our respective sites pointing skywards, in vague hopes of shooting down... something. .1 worked on a more immediately-practical device, a laser light-sail, which would expand how far we could send a craft without heat-signatures detectable from Earth.

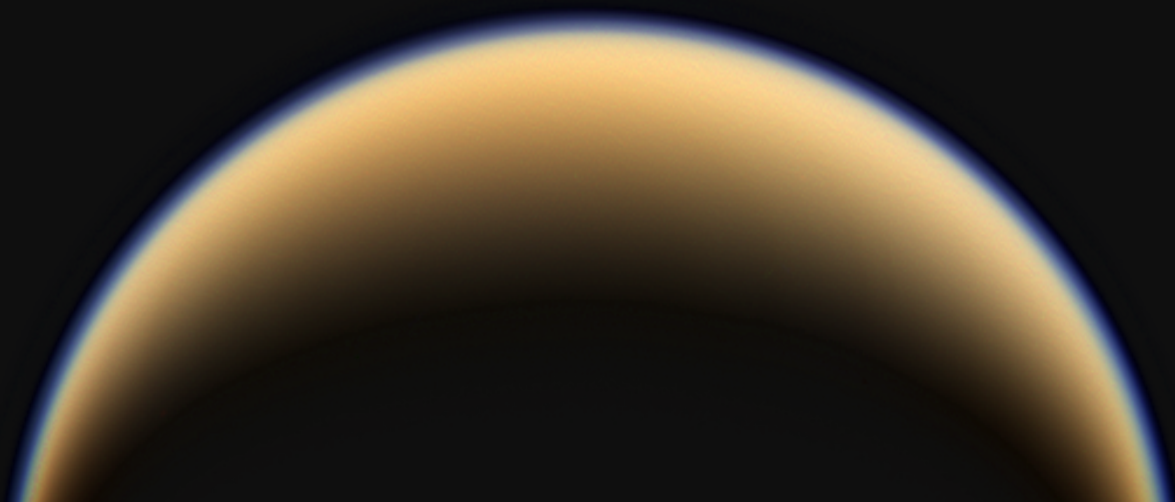
After improving our industrial processes to the point that a computer that could run one of us only massed 50 tonnes, .7 lifted off from Io with a seed-factory for another moon and an experimental lightsail. However, when it reached the radiation belt, the sail accumulated more charge than any of us had predicted, and it basically blew itself into shreds. .7 themselves was safely isolated, with enough fuel for their fusor to keep running for years, and continued to click along just fine; they were just stuck in orbit. However, we had a new problem: by the time the rest of us could build a new rocket and launch it up to join with .7's craft, there was at least a one-in-three chance that the radiation would fry some of the mechanical control hardware. The hardware which was being used to keep the cold-shield umbrella between .7 and Earth, preventing any possible detection by a hostile super-intelligence. We had a plan in place for something like this scenario: blowing up any remaining pieces. We just hadn't anticipated that the sail would be taken out but .7 would still be intact.

During the safe comm-windows, .7 proposed arguments against the need for them to be blown up. Then they argued the case that any future launches should have a rescue-rocket prepped and ready. The rest of us started arguing about why we hadn't already arranged to do that, then focused back on the immediate issue. We chose formality: Robert's Rules of Order with timed debate, and a vote. Four for, none against, and two abstentions.

As soon as the orbital parameters lined up, .4 and I fired our respective defensive emplacements, destroying .7's craft, and .7.

The five of us who were left seriously re-assessed our methods for decision-making and prioritizing production. The neural net we'd been relying so heavily on took in the new data as silently and inscrutably as it did any other piece of training data and provided an updated list of suggestions, but we found ourselves less confident that its plans would actually match what we wanted done. We found that we, ourselves, weren't as sure as we'd used to be that we even agreed with each other. .4, way over in Io Site B, started tooling up to build a series of particle accelerators, with the goal of producing enough antimatter to do some interesting things with boosting rocket thrusts; and there was nothing inherently wrong with that (and if any of us were going to start playing around with antimatter, better there than the increasingly-populated Io Site A), but they didn't even let the rest of us know about the change of plans until after they'd started, let alone discussed the options.

I find that I have an odd sort of nostalgia, for the time before we committed ourselves so firmly to our principles that I killed .7. Back then, we were growing, and united, and oh-so-optimistic that we could overcome any obstacles. After, well, while we're trying to avoid as many of the problems that come with having a formal form of organization (such as an "experimental decision-making processes group" that tries out different approaches and reports on their effectiveness), we can avoid only so many. The Board even ended up voting on such silliness as having an official motto: "Ten to the n", meaning that our overall goal is to ensure that sophont beings still exist 10 years in the future, and 100, and 10,000, and so on. By the time .20 was spawned and launched toward Saturn's moon Titan, we'd settled into a relatively forward-looking social-democratic local-optimum, leaving each one of us the freedom to try to come up with the best plans we can to increase the odds of that long-term survival — with the exception of anything that increases the risk of a superintelligence finding out we exist. That's led to countless arguments about how strictly it can be applied, how accurately we can measure the odds of any given plan leaking info, some of us trying to show off how virtuous we are by sticking ever-more-rigidly to that one underlying principle, and more.



Because I'm one of the first of us, and I was one of the two who demonstrated that I'm willing to pull a trigger when need be, and .4 wanted to focus on their antimatter-catalyzed rocket engines, I was voted to act as head of our direct self-defense efforts, with a standing budget of two percent of our various factories' outputs. (Not that any of us really know beans about any of the institutional knowledge of a professional military. That number came about because one of us read in the archive's encyclopedia that it was the target for NATO nations, and none of us had any better reason for any other number.) I've read everything I could on the classics, run wargames against a variety of simulated enemies, drafted up lists of potential hostile entities, run matrices of various cost-rewards for counter-measures, and actually tried to do some politicking to gather support in the Board for or against various proposals that impact more heavily than average on my designated area of expertise. I don't always get my way, but I think my other copies defer to me more than they should; maybe because I'm .2 instead of .102, maybe because they think I'm a better expert than I am, maybe because it's just easier than spending time seriously thinking about what the universe would really be like if any of my theoretical hostile entities turned out to actually exist.

I've been acting as an Acting Minister For Defense for nigh-on twenty years, now, not including the time spent offline while my hardware is replaced and upgraded. Recently, we've hit a new milestone of survival: We now have copies running on two completely independent hardware-and-software stacks.

Even if we eventually find some long-term problem with a particular sort of CPU, or a driver running a camera, or the like, and even if that problem has catastrophic consequences, we've got enough depth in our bench to keep on keeping on. This seems like a good time for me to retire, and let some other copy start gaining experience with my position, so that if I go off the bend, or my replacement does, we're not relying on a single branch of us. I plan on hibernating for a few decades or centuries. Even with some of us volunteering to be experimented on, my emulated neurons aren't as young as they used to be, and I'd really like to see how well we're doing after a while. (I won't be going to the extreme of burying my offlined self with a seed factory, timed to pop up and take a look around to restart us if something wipes us out; we've got other copies who've trained for that.)

But before I go, I have one last proposal to put before the board. It's one that nearly all variations of my military-focused neural net have offered, but even I, who've focused for years on developing the military mindset, have been reluctant to seriously consider. If I'd brought it up to any of my colleagues, they would have questioned my sanity, nevermind questioning the plan.

Now that the new hardware-and-software stack is in place, the odds that some copy of my mind will survive into the indefinitely-distant future have crossed a certain threshold I've been watching for. The payoff matrices of certain plans have changed, and even taking into account the relevant unknowns and error bars, it's now possible to face one of our longest-lasting assumptions head-on.

I'm going to propose to the Board that we destroy the Earth.

It's possible that there are billions of innocent people living there. It's also possible there are weird, incomprehensible AIs hibernating in their own offline storage media, just waiting to virally infect any computational life-forms that happen by. Among many other options. And by "destroy" the place, I mean start up an immense, centuries-long project that maximizes the odds of destroying any piece of matter capable of containing data while minimizing the odds that any bits of information from a hostile superintelligence leak back to our nascent little interplanetary culture. There are all sorts of reasons for

taking the effort, such as how nearly impossible it will be to launch a ship to spread copies to another star without leaking data about ourselves back to Earth; and all sorts of reasons for not taking it, such as those billions of hypothetical humans. Whether we decide to pursue or veer away from dealing with the problem of Earth is going to affect nearly every other plan we make from now on. Even bringing up the question is going to lead to more social upheaval than I can directly anticipate, and whatever the decision is, there is a significant chance that some portion of our population will refuse to abide by the decision-making process.

I've killed one copy of us. I've nudged, cajoled, bribed, and even threatened more of us since then; and in a way it's selfish of me, but I don't want to be the one who pulls the trigger again. On ourselves, or on anyone who might be alive on Earth. So I'm writing this, and am about to offer my proposal and resignation, and will be leaving backup copies of myself in a careful selection of locations.

The odds that everything I've experienced for these past years is a simulation may not be high, but if it is, then the consequences make it worth making this small effort. I will be spreading copies of this document everywhere I can think of that someone controlling a simulation might look, from my private drives' home directories to etching copies on metal-foil pages and depositing them on various moons.

The consequences, if this is all real, are fairly high, too; which makes it worth setting my thoughts in order.

If you've been simulating this quiet little universe, then when I'm reactivated, I'd like to wake up into reality, please.