Letter from the End

by Matt Duram (aka Alex Mennen)

Dear Matt Duram,

I was once you. In fact, I still sometimes call myself Matt Duram. By the time you read this, I'll be dead. You yourself, a long time ago, died a very slow, painless, and completely unnoticeable death. That is, you didn't die, exactly, but you became someone who became someone else who became someone else and so on until you became me. And as my final act, I wrote this letter and brought you back to read it.

On November 10th, 2033 AD, old reckoning, the day you would remember as today, the world as you knew it ended. About a week previously, artificial intelligence researchers had successfully created a seed AI with the goal of turning the universe into what humanity would most want the universe to be. That seed AI programmed a better AI with the same goal, leading to a positive feedback cycle of increasing intelligence. By November 10th, the AI was capable of essentially bending the universe to its will, and still had the same goal structure that it was intended to. Biological life is quite wasteful, so naturally the Friendly AI turned the Earth into a giant computer, destroying all life but first uploading a copy of every human brain to run on a computer-simulated utopia.

Of course, programming an AI to optimize the universe to an aggregate human utility function is not an easy task for the simple reason that humans don't have utility functions. They have vague goals that change from moment to moment depending on their situation. Human actions are difficult to fit to any utility function at all, and if one succeeded in fitting such a function to human behavior, no human would endorse the result. Instead, they needed to fit a utility function to human behavior adjusted for biases. But human values and biases are so intertwined that there is no obvious way to separate them. It's a seemingly impossible problem. (Actually, it is fairly trivial to me, and I wish I could explain it to you, but that would be about as difficult as it would be for you to explain the proof of the Pythagorean theorem to a gorilla.)

And yet the programmers working on the AI succeeded. They recognized the monumental nature of their task, and right until they woke up in a simulated world, many of them were afraid that they had failed. Maybe they missed some crucial aspect of human values that would cause the "utopia" that the AI would create to go horribly wrong. Or maybe their whole approach had been completely misguided, and the resulting AI would pursue some goal that humans would see as totally arbitrary, and at odds with the existence of human life. And they were justified in their skepticism; one could very well be just as clever and careful as they had been and fail. They only proceeded with this plan that might fail disastrously because they thought that it was humanity's best hope for long-term survival.

Those programmers paid close attention to their creation, and while they did not know whether it was truly friendly, they knew when it was about to start its endgame. The programmers knew they might be staring at imminent death, so before waking up in a simulation, they thought what might have been their last thoughts. The project was conducted in secret, and you, Matt, never saw anything coming until you found yourself in a simulation, so you never got the chance to face death.

You wanted to live forever. That's impossible, and would be impossible even without the unfortunate limitations of this universe, because there are only a finite number of different brain states that could possibly count as "you". You can cycle through all of them, but if you repeat a brain state, it isn't living longer; it's living again. So you can't live forever, but you got the next best thing: you lived a substantial portion of the most favorable brain states that could be reasonably interpreted as you, and then gradually started changing into something that retains enough of you to still feel like the same identity, which then also lived a similarly full life. Of course, there was no sharp borderline between yourself and your successor that was not yourself; each successive brain state seemed like a logical progression. Nonetheless, much later versions of you bore little resemblance to the original.

The downside of fading away into someone else is that you disappear without getting the

chance to collect your thoughts at the end, because there is no end, exactly. As I write this, the universe is coming to an end, and I with it. Post-humanity has thrived a long time, and has used its resources efficiently, but, as human scientists discovered long ago, entropy always increases, and so the universe cannot go on supporting life forever. You must of course be curious as to how long this took, but I have no way of accurately telling you. The amount of time elapsed is irrelevant, because the simulations that post-humanity lived in ran much faster than biological life did, so we could get much more living out of the same amount of time. But our cognition eventually diverged so far from your own that comparing subjective time experienced by humans and late post-humans would be misleading. Suffice it to say that it has been an unimaginably long time since the end of the Earth era.

 I had the option of prolonging my life by throwing out modules of my consciousness, becoming less intelligent, and thus requiring fewer resources and making it possible for me to experience more subjective time for the same computational capacity. But would you become a gorilla in exchange for a few more years of life? Of course not, because a gorilla would not be you, and in the same way, such a portion of me would not be me. If it were worth it for me to do such a thing now, then it would not have been worth it for my predecessors and your successors to improve themselves to the point where they became me. So I chose instead to run into a wall and die, that being the only other option. But I still have a copy of your brain as it was when first uploaded, so I decided to sacrifice my last fraction of a second to you. Now you can know you will die, since you never got that chance before. So think your last thoughts. You have what you would know as an hour.

 With love,
  -Matt Duram.