# [Existential Risk / Opportunity] Singularity Management

**October 31, 2018**

## Contents:



---



---

## The Intellectual Challenge of Global Risk Reduction

by James Blodgett

Global risk reduction is a challenge because it is tremendously important, but many versions of risk reduction seem unlikely to work. It is important because the stakes could be trillions of human lives. Nick Bostrom estimates roughly that, if humans are able make extensive use of what current physics sees as the reachable universe, that might enable as many as $10^{58}$ human equivalent lives. [Bostrom, Superintelligence, location 2452 in the Kindle version.] That is much more than a trillion times a trillion. This achievement seems improbable, but there are existence proofs for technology that might make it possible. Even if our probability of effectuality in helping to make this happen is one in a billion, the expected value (probability times value) of our effort would still be more than a trillion times a trillion lives. Even if we can't settle the reachable universe, there is enough material in our solar system's asteroid belts alone to enable us to build O'Neill space habitats for trillions of people, and the technology that could enable this is more or less plausible. Even if this does not work, we still have 7.5 billion people on Earth. With good and plausible management we could continue to support at least a few billion at a time for millennia, summing to more than a trillion when adding all of the generations during that time. Unfortunately, the probability that we will go extinct before we are able to do any of this is also more or less plausible. Lord Martin Rees, among much else Astronomer Royal of Great Britain, estimates that the probability of human extinction within a century is 50 percent. [Rees, Our Final Hour, Basic Books, New York, 2003.]

The intellectual and motivational challenge is that, even though expected value is the gold standard for much of decision theory, it seems difficult to build a motivational argument based on expected value that convinces many people. Note the problems with expected value at the extremes in our trolley problem in the April 2018 EROSM. Much of the problem is what economists call "the tragedy of the commons." The English commons was a plot of land in a village that anyone could use to graze his or her cattle. The benefit to an individual farmer of feeding his cattle was more than the benefit to him of preserving the common land, so the commons were often overgrazed and ruined, even though that outcome was bad for the commonwealth. Many natural resources can be analyzed as being a commons. Without regulations, anyone can throw his garbage in the ocean, and anyone can pollute the atmosphere. The problem is that a benefit to others is often not considered when making a personal decision. This is especially true when the benefit to others has a low probability. Despite these difficulties, people can sometimes see the value of benefits to others. A large number of people do see the moral value of conservation, and also the moral value of being careful with technology that could be an existential risk. Likewise, a large number of people do see settling space as an investment for the future and as a backup for Earth. However, our potential for developing space requires investment, often very expensive investment that is difficult to recoup in the short term. This is a prospect that discourages investors who want to see short term returns on their investment, and discourages voters who have the same feeling about their tax money. For these reasons, the people who do see the future value of these things for the commonwealth are usually not a majority. However, sometimes creative intellectual and motivational arguments are able to assemble a majority, or a new technology shows a way that is more practical and is able to attract funding, for efforts in these areas. Our job is to try a lot of things in the hope that something works.

Great things have been done in the past. We live in the golden age of all golden ages, and the future could be exponentially better. However, risk also grows exponentially as technology advances. Artificial Intelligence (AI) theorists worry that an AI tasked with making as many paperclips as possible will discover magic technology that enables it to turn everything in the universe, including us, into paperclips. The problem is not only the AI, it is also the technology. If humans developed similar technology, the natural intelligence in our own heads would be adequate to ensure our demise. If anyone could push a button that would destroy the universe, some nut would surely do so. One hopeful consideration is that, while the effectuality of technology has grown exponentially recently, there are signs that that growth may be slowing. With luck, there is no magic technology that can destroy the universe with the press of a widely available button. However, there are technologies that go in that direction, so we do have a job on our hands.

# Happy Halloween

by James Blodgett

This issue of EROSM is being published on Halloween. Halloween is a time when we enjoy contemplating gruesome monsters. Most Halloween monsters are imaginary, but humans themselves have sometimes been gruesome monsters. The issues we work with in this SIG could be scarier than gruesome monsters, and at times solutions are problematic. These could be reasons for despair. "Right stuff" test pilots had a relevant aphorism: When the plane is about to crash, you have to be "afraid to panic" and instead of screaming in terror "fly the airplane." Sometimes that works. I am rarely discouraged, mainly I think because I have an airplane to fly. I am flying that metaphorical airplane by writing this. I have a quest, similar to but I think more realistic and more important than the crazy but glorious quest of Don Quixote. In "Man of La Mancha" he describes his quest in the song "The Quest," often called "The Impossible Dream" because of its first line. It is available on YouTube at https://www.youtube.com/watch?v=RfHnzYEHAow . I mentioned this before in the July 2017 issue of EROSM. It is worth a reprise. The motivational value of this song is demonstrated by the number of times its several versions have been played on YouTube and elsewhere and by the many singers who have covered the song, even Elvis Presley. Perhaps if we make our pitch well enough, we can motivate others with our quest, a quest that is difficult but not impossible and a quest that does have a high expected value for the future. Or perhaps we will come across and help develop and promote a technological fix that is more practical, albeit we had better be careful with anything that potent. With luck we will achieve something. At least our quest is an interesting hobby, and at least we are trying to be the good guys.