Existential Risk / Opportunity Singularity Management

October 31, 2019

Contents:

Management For The Benefit Of Humanity

By James Blodgett

We are concerned here with management of productive resources for the benefit of humanity as a whole. There are reasons why this is difficult, but there are arrangements that can mitigate those difficulties.

Management of productive resources is easiest when the beneficiary of that management is a single individual, because in this case there are no conflicts of interest. When more than one person is involved there arise questions about what is produced and who gets what part of that production. Outputs from productive processes serve different purposes and have different values, sometimes negative values in the case of wastes. A classic problem with production is when goods accrue to the producer, and wastes are dumped into a common resource such as the ocean or the atmosphere. In economics this is known as "the tragedy of the commons." A "commons" was originally a plot of land where everyone could graze their cattle. The original tragedy of the commons was the tendency to overgraze such land to the point that it could no longer be used for grazing. More modern tragedies of the commons are pollution, and in general use of scarce productive resources in ways that benefits some people and exclude others, especially when this exclusion has existential results. A solution is to enact laws that forbid inappropriate grazing, inappropriate disposal of waste, or inappropriate use of resources.

A problem with development of appropriate laws and regulations is that their development requires a legislative body and an enforcement mechanism that has adequate scope. Scope can be inadequate 1) when a commons is not part of the jurisdiction of the legislative body, 2) when the population that is affected by a commons tragedy does not have a right to contribute to determination of the composition of the legislative body, 3) when enough members of the electorate do not care about other members and are willing and able to exclude them, or 4) when technology that can reliably sense things such as pollution is not available.

Historically, the passage of time brought expansions of scope that solved some of these problems. Political scope expanded as tribes became states, states became nations, and some nations made treaties with other nations and formed transnational organizations. Representation expanded with the enlightenment that championed representative government. Technical scope expanded with the advancement of science. However, we do not yet have a world government that can do the heavy lifting required for the more expensive versions of planetary management, and the more expensive versions of expansion into space. Such a government or such a project might be tragic if done wrong.

The most representative government would include representation of future humans. The laws of physics seem to prevent a real version of this, since the future appears to be indeterminate, and even if it is determinate, we can't communicate with it. From our point of view, future humans are potential, and there are more potentials than can ever be actualized. There have been serious proposals to include representatives for future humans in various planning bodies and legislative bodies, but these representatives can only estimate future needs.

Planning and spending for the future is of less personal interest when that future extends beyond our personal projected lifetimes, except as a form of charity. Life extension might mitigate this, since we would then have more of a personal stake in the future. Another form of mitigation is to expand the group with which we identify and for which we are planning. That expansion can, and often does, include future people. Good politicians work for the benefit of the whole polity, including the future polity. Our evolutionary psychology supports this since future people include our descendants. Even people without descendants have a stake in future people since we are all relatives and we all share many genes. Direct descendents have less of our directly inherited genes with each generation. Some random future people share more of our personal genes than direct descendents.

We may want to expand the idea of what is ethically "human" to include people who are not technically our same species, people such as uploads, transhumans, space aliens, artificial intelligence that thinks like us, etc. A test is whether they have an

experience and an appreciation of life that is similar to ours, whether they have a "soul." Some say that only God can make a soul, but people who think that God is omnipotent cannot say that He is unable to provide a soul in these cases, and it would seem that a good God might do so. Suppose that God designed the universe so that something like us would evolve, and so that evolution would evolve ways to accelerate evolution, ways like 1) when the storage of genetic information transitioned from RNA to the more stable DNA, (i.e. the theory of the RNA world) and 2) when evolution developed a way to split and combine the traits of two organisms so the complete sequence did not have to happen by chance combinations of molecules, but rather by the much more efficient combination of traits already successful in separate organisms. (I.e. sex, but also incorporation of symbiotic genetic material such as mitochondria.) God might have planned human science as another way of accelerating evolution, so that any genetic modification that we introduce is not only our design, but also part of His design. (On the other hand, there are human and scientific reasons to be careful about genetic modification. It is a form of technology with which we should be careful, because it is a form of technology that could get out of hand.)

The problem with the question of whether intelligences with a substrate that is different from ours have a similar experience of the world is that it is not easy to sense another being's experience of the world. It would be tragic to give the universe to robots based on the idea that they are as human as we are us when they in fact have no more consciousness than a refrigerator, but it would be a tragedy not to make this use of the universe if robots are the only way to make use of it and if they do have souls. "I think therefore I am" shows us our own soul, but doesn't work to show us the existence of other souls. Other people could all be part of a simulation. It would be helpful in sensing other's souls if we could read and share minds, and something like this is vaguely plausible by moving electrical signals between brains. It seems more plausible that we will have to rely on something like a Turing test focused in this direction, and on shared experiences. Usually we postulate the existence and the consciousness of others based on their similarities with us, both the similarities of their apparatus for thinking, and the similarities of their reports of their feelings, their experiences, and their poetry. If this is the best we can do, this will have to be the basis for this piece of piece of our philosophy.

"The best we can do" is the best we can do. When Admiral Farragut sailed his ships into Mobile Bay, one ship was destroyed by mines (called torpedoes in those days). Other ships hesitated. Farragut made his choice, indicated by his paraphrased order: "Damn the torpedoes, full speed ahead," and won the battle. He might have lost, but he had reasons to think what he did was the best approach. We want to be careful, but we also have to act. Let's be careful, and let's make decisions that are as much as possible a group process and a benefit to all, but let's also be careful to do the best we can do, where "we" includes all of us. That involves trying things.

Permissible? Cargo Cult Thinking

By James Blodgett

Space development as an investment for humanity can be seen as cargo cult thinking.

Cargo cults happened when South Sea Island people were exposed to the paraphernalia of modern life. They wanted some of that too, so they set up mock airports, and appealed to their ancestors to land and bring them cargo. Their thinking resembles ours. They had an existence proof. They could watch foreigners receiving the cargo they coveted.

Some islanders understood the foreigners better, and actually did receive cargo. They set up enterprises that earned money so that they were able to order things.

Much of our hope for space industrialization and settlement resembles cargo cult thinking. We want it so badly that we misunderstand the difficulties. Real scientists speculate about things like worm holes, warp drive, and hyperspace, and science fiction writers enthusiastically adopt them to move their characters around the universe. Other scientists then warn us about their improbability. Both Michel Mayor who discovered an exoplanet and recently won a Nobel prize, and Neil deGrasse Tyson during a recent lecture in Dubai, have said that traveling to the stars is not going to happen because they are so far away.

Settling the galaxy could enable trillions upon trillions of human or human equivalent lives. It is not going to be easy, and may never happen. However, there are existence proofs that something like that is possible. Assuming that human life is valuable, it is worth trying because of expected value, which means probability times value. Even if the probability is low, the expected value can be high if the value is high enough. If the expected value is high enough, it is worth trying lots of things.

I keep talking about existence proofs for interstellar settlement. I should remind readers of what I am referencing. Some of the idea comes from Stuart Armstrong and Ander Sandberg, Eternity in six hours: Intergalactic spreading of intelligent life and sharpening the Fermi paradox, Acta Astronautica, Vol 89, Aug–Sept 2013, Pgs 1-13. http://www.fhi.ox.ac.uk/wp-content/uploads/intergalactic-spreading.pdf . (Stuart Armstrong commented on our Trolley problem in the July 2018 issue of ERSOM.) I mix this with a concept I call "seed ships" which has some precedent in science fiction. Seed ships have a payload of arbitrary size, perhaps the size of an actual seed. They contain artificial intelligence (AI), nanotech, cell templates, and the DNA of many species. The DNA is perhaps recorded, perhaps actual DNA which stores data compactly. We can already synthesize DNA from a record of its contents. A record might be a better way of preserving data since we could use error correcting code. We might correct errors in actual DNA by comparing several

copies. Given adequate technology and adequate scale, we might be able to send seed ships to billions of targets, of which some might be appropriate, the strategy of plant seeds. Upon arrival at a source of material and energy in another solar system, the AI directs the nanotech to reproduce and to build macro infrastructure like macro machines, robots, labs, incubators, green houses, O'Neill habitats, etc. Then the DNA is inserted into cell templates and grown to produce plants, animals, and finally humans. Humans are raised and educated by robots. Voilà, we are there! This may be less than satisfying to folks who want to go in person, but with enough altruism we optimize the greatest good for the greatest number. If this works that number can be really large. If we can make AIs that have souls, that might require less material and less infrastructure per individual, making that number even larger. The problem is making sure that they really have souls.

An existence proof for AI is the intelligence in our own heads. Google and Watson are getting close. An existence proof for nanotech, which means molecules that are machines, are the many molecular machines in our own cells, machines that would be called nanotech if they had been designed rather than having evolved. We can already sequence DNA and reproduce it from data. An existence proof for our artificial seeds are the seeds in nature that grow to become plants. Our postulated seed grows to become an entire ecosystem. My wife, a serious Catholic, thinks that seed ships are immoral because they involve unnatural reproduction. I don't think that Teilhard, a Jesuit priest who thought that God wants the universe to come alive, would agree.

Many people used to think that a mechanical airplane was impossible, despite Leonardo da Vinci's designs. This is surprising because birds were an existence proof that something like an airplane is possible. However, after watching the following video, I can see why people were skeptical about early aviation. https://www.youtube.com/watch?v=AXT4pgW_UGk . A while ago, someone posted a similar video of robot fails, making the point that robots are over hyped and improbable. I posted a link to the airplane video to make the point that fails don't prove impossibility. On the other hand, an existence proof doesn't prove the possibility of something that is only similar. Seed ships will require vast improvements of our current technology. However, similar vast improvements have become the routine results of Moore's law, which of course is not exactly a law and may be slowing in its area of application.

Even if we never reach the stars, there are amazing things we might be able to do in our own solar system. If they are amazing enough, they may give us the resources to reach the stars by brute force. for example, if we have a million O'Neill habitats, we might be able to nudge a few of them into a million year trip to other stars.

I contend that a carefully grounded version of cargo cult thinking is appropriate in areas with real possibilities and tremendous expected values, even if our ideas may not work, and keeping in mind the limits of utilitarian ethics at expected value extremes, as per our trolley problem in EROSM Apr 2018.