[Existential Risk/Opportunity] Singularity Management

October 28, 2017

Contents:

What I am Doing Here Today

by James Blodgett

I am mostly a nonfiction writer. But some interesting fiction is based on reality, and interesting "postmodern" writing is a mix of weird things. My quote marks around "postmodern" are because I am using it as a metaphor for recent literary weirdness, not officially defined postmodernism; indeed an official definition would be controversial. For example I was thinking partly of **Gravity's Rainbow**, which is considered postmodern by most, but especially of **Zen and the Art of Motorcycle Maintenance**, which may not meet someone's technical definition. The action of the latter book involves a man on a motorcycle on vacation with his young son on the back of his cycle, and his two buddies, an artist and his wife, on another cycle. The action is minimal. They ride through towns. They look at scenery. They stop at diners. Occasionally they exchange a few words. The real action is in the protagonist's head, as he spins an elaborate philosophy that has some small relation to the events of the trip. It also considers motorcycle maintenance, a minor concern on the trip, used as an example of our relation to the technological world, and the

relation of that to the artistic and philosophical world, and how different types of people and different philosophies relate to such things.

Philosophy and activism about management of singularities is a hard sell. It is beyond most people's immediate concerns, and it is often felt to be beyond their area of effectuality. **Zen and the Art of Motorcycle Maintenance** sold over 5 million copies. As an activist I would love to be able to emulate its marketing success, although while its philosophy is intriguing, I don't think the majority of readers were converted. The following is my attempt, as an advocate for activism towards improving human futures, to try some postmodernish writing to see if that works. I offer it here as a market test, to see if it is worth writing more. Part of my strategy is to try a lot of things in the hope that something works. Send me an email and tell me if this works for you. Start at http://www.global-risk-sig.org/contacts.htm .

Management of Positive and Negative Singularities as a "Postmodern" Book Sample

by James Blodgett

Introduction

There is much talk in some circles of "the singularity." Books have been written about it. A technological singularity is the postulated result of advancing technology if it increases exponentially, as it appears to be doing in some areas. (But note that exponential growth often hits limits.) In math, a singularity is the point where a function goes to infinity. Exponential growth does something like that. A technological singularity is the point when technology "goes to infinity" and becomes transcendently amazing. Amazing technology could allow us to do tremendously great things. However, a negative version might do tremendously bad things. Managing such things sounds like hubris, but we can at least try to tweak probabilities in favor of the good version. Because of the large number of human lives that would be affected, there is tremendous expected value (probability times value) in even a small tweak in probability, something we can plausibly accomplish. Some heroes of history have had the luck or skill to accomplish more than just a tweak. We might hesitate lest our actions disturb a balance, but we are constrained to do something, since doing nothing is also an action that has consequences.

It would be easy to become didactic about this and to lecture about what humanity must do. However, the subject is complex, potentially controversial, and can become boring when elaborating a premise with which the reader does not agree.

The Greeks had a solution for this: the Socratic dialog, a form of bull session featuring Socrates, the world champ at that form of discourse.

Greek drama featured three players. Freud postulates three components of human personality. I offer a dialog among three components of my own personality. You will see how Socrates crashes the party.

==========

James: Allow me to introduce myself. I am the didactic component. I am the professor giving a lecture. My objective is to build a system for thinking about this area. I also introduce my colleagues: Jim, our critic, and Puck, our humorist.

Puck: Ouch! I can't get rid of that silly attribution! Let all know that everything that appears under that name is not Puck! Right now I am preparing to channel Socrates, who is bored in Heaven and wants to get into this thing. Besides, if we get him into this, we can claim it as a Socratic dialog! Think of the marketing value of that!

Puck: This is neat. I find that I can control the strike-through. See: ~~strike-through~~. Socrates can use this to remind folks who is really talking.

~~Puck~~: **Socrates**: Thank you Puck. And thanks for that strike-through. Hello James. I hear that you profess to have ideas for the management of singularities.

James: Isn't that an anachronism? What does an ancient Greek know about singularities?

~~Puck~~: **Socrates**: We are following Earth tech up here in Heaven. You guys are doing an incredible job. We didn't even know how to build arches for what you now call **arch**itecture. Earth right now looks like a singularity to me.

James: I know your methods, Socrates. You are here to expose the weakness in my argument. I welcome you. I agree that there are weaknesses. It will help to work them over. The Socratic dialog has what Puck calls "marketing value" because it was another great Greek invention.

~~Puck~~: **Socrates**: Do you think to flatter me that they have named this method after me? If you know my methods, then you know that I am just an old man with much to learn.

James (smiles)

~~Puck~~: **Socrates**: I know that sounds like my "just an old man" ploy, but this time it is true. I am usually the intellectual hero, with the script writer on my side. That is less true this time.

James: I have some kinship with the scriptwriter. I can tell you that he is aware of weak areas in my exposition and wants to explore them. I think that we are equally matched.

Types of Singularities

Jim: You have defined "The Singularity" as that time when technology "goes to infinity." Then things are supposed to become awesome, but also unpredictable. Most singularity writers go on to tell nice stories about what will happen. If the singularity is unpredictable, how can they tell stories about it?

James: Writing attracts an audience because of its persuasiveness, not its consistency. To give singularity proponents credit, I think they use unpredictability as rhetoric to evoke the wonders of the singularity, not as a precise description. Then they use stories to tout the enticements of the heaven they are promoting. I am not a "singularitarian," by which I mean I don't worship "the singularity." The "the" indicates that there is only one. I see many potential types. Some are wonderful; some are awful.

~~Puck~~: **Socrates**: How do you measure where they stand on the scale from wonderfulness to awfulness.?

James: For rhetorical simplicity, I like to denominate in human lives enabled (or lost). Some singularities could enable trillions of human lives, others could kill us all. Others work their wonders in other ways, without necessarily increasing or decreasing the population of humans. If there is no change in numbers, then we can consider something like QALYs, Quality Adjusted Life Years. These were developed by medical people to rate the value of medical intervention. Consider as an example a patient with a slowly developing brain tumor that will kill him in a few years. An operation could extend his life but reduce his ability to speak. Will he feel that the extension of life is worth its reduction in quality? Some singularities involve not more people, but longer life, or various changes in humans or substitutes for humans that may or may not enhance their quality of life. For example, if we extend a human mind by interfacing a human brain with a computer, does this improve the quality of life of the resulting transhuman?

Jim: I suppose we could ask him.

~~Puck~~: **Socrates**: Is there some basic principle, so we can know in advance?

James: It would be nice to have a philosophical principle for determining what is good in life. That is somewhat of a digression right now.

~~Puck~~: **Socrates**: I respect your strategy for developing your thesis, but you know that it will detract from that thesis if you cannot answer my questions.

James: Agreed. We will discuss that later.

~~Puck~~: **Socrates**: What types of lives might result from different types of singularities?

James: Space resources could enable trillions of human lives. "Uploading" could as well, if you consider an uploaded mind to be equivalent to a human mind. Uploading involves recording the brain's entire connectome, i.e. the state and interconnections of all of its trillions of synapses, and then running a simulation of that on a computer, so that (if this works) the mind is put into the computer. The computer could simulate an environment to give the mind something to work with, or the mind could control a robot body in our world. A mind running on a computer could live almost forever, as long as the computer could be kept running. If it lives in a simulated environment, that environment could have marvelous properties. Multitudes of minds could communicate and interact and merge in marvelous ways. They could improve themselves with few limits. One of the biggest limits is knowing whether a proposed improvement is truly an improvement. Uploading is one of the most popular of singularity predictions. I personally think that a high-resolution upload, one that I would consider to really be me, is probably impossible. It would seem necessary to scan the state of trillions of synapses and their interconnections. I doubt that this can be done remotely. We might invade the brain with trillions of nanobots, but I don't see how they could map out the interconnections. We might freeze, slice, and scan the brain. This might work, but requires slicing with a precision and a lack of trauma that may not be available. (Recent success at preserving brain structure when frozen suggest that it might.) Everything we can imagine is not automatically possible. But this is a digression. Another prediction is the creation of artificial, robot minds. How do we rate the QALYs of a robot mind?

Jim: I suppose that depends on its design. Robot servants would increase the quality of life of the humans that they serve.

~~Puck~~: **Socrates**: Does being served increase one's quality of life?

James: Later, guys. Then there is life extension, giving us longer life. That automatically increases QALYs since they are denominated in years, as long as the quality of life remains consistent. There is also the potential of genetic or artificial enhancements to improve current humans.

Jim: How do we know that it is an improvement?

James: We will have to develop standards.

~~Puck~~: **Socrates**: How do you do that?

James: I hope that standards can be developed by diverse consensus, by lots of people reacting to developing possibilities. But right away I propose a simple standard, that of utilitarianism, the greatest good for the greatest number. That "greatest number" automatically endorses the value of large numbers of humans.

Jim: Some people would not like that, fearing overpopulation.

James: Overpopulation is a problem in a limited world. A singularity removes most of the limits.

Jim: It does not remove all of the limits. The universe is ultimately limited. I saw someplace that exponential growth of population would be limited at least when it resulted in a sphere of humanity expanding at the speed of light. It couldn't go faster.

James: Perhaps, but that is way beyond even most of the singularities we intend to consider here. Utilitarianism also has limits when it conflicts with deontology, with prescriptions like "thou shalt not kill." Utilitarianism might approve killing someone if it made many other's lives better. Most people wouldn't agree with that in most cases. We are using utilitarianism as a rhetorical simplification here, not as the ultimate philosophy. We will explore relaxing simplifications later.

Recruitment Pitch:

James: This dialog is partly a recruiting pitch. We are setting out logic that justifies the moral value and the importance of trying to steer humanity towards positive singularities and away from negative ones. Someone has to do the steering. In our discussions, the "someone" we imply can mean a generalized someone. Or that "someone" can be authority figures or politicians. Since in a republic we get to select politicians, we can contribute by selecting those who want to steer where we think we should go. Or that "someone" can be something on the scale of a nation or even on the scale of all of humanity. However, that "someone" can also be you, the reader. Part of our thesis is that actions of ordinary people can make a difference. Working on this issue is a way to contribute.

Jim: Ordinary people make a difference because their actions might tweak probabilities, at least by a tiny bit.

James: Right. Our experience is that this is an esoteric pitch. Most people don't get it, or don't believe it, or want to leave it to others. In general, that is a good thing. We

don’t need everybody in the world working on these issues, although that might help when it comes to voting. We especially don’t need people trying to steer when they have a poor sense of direction, and might steer in the wrong direction.

Jim: Okay, but even if most people don’t answer the call, a small percentage of the world population will. There must be thousands who are working in this area already. Isn’t it a problem when many people try to steer? What if they all steer in different directions?

James: A consensus of a group of people usually results in better decisions that those of a randomly selected member of the group. If done well the result can be somewhat like a congress or a parliament. Designing the averaging function makes a difference. If everyone has a rudder, the boat will move in the average of the directions to which the various rudders point. It could be a simple average, an all-or-nothing function like a vote, or some other form of consensus. Mathematical economists have considerations that can help with designing the averaging function. It also helps if individual members have ways to address the group, so that those with persuasive arguments are able to convince others who may start with no strong preference. Even without a specific forum for deliberation and decision making, the climate of opinion has some aspects of this, and the climate of opinion often informs decisions.

Jim: Wouldn't it be best to select people who are best qualified to make decisions?

James: That is the point of representative democracy, and it is the reason that qualified people are appointed to administrative positions. But selecting the best qualified is not easy. The best solution might be to have a philosopher king. The problem is that there is no sure way to assure that a candidate for king is the best philosopher.

Puck: But that is no problem. I am available.

James: Puck, your availability illustrates the problem.

Puck: Hey, I make the jokes around here!

James: People who are interested in this area, who can see things that might be done or can see the value of other’s ideas, already have an important qualification. They are already likely to be better than average at making decisions in this area. However, we have seen some raucous debates among people interested in this area. Perhaps raucous debates are a good way to sort things out, but some ideas really are better

than others. We should suggest practical guidelines and ethical principles for work in this area. Things like “first, do no harm," although “first do no harm” is not exactly right. Drug side effects cause harm, so drugs are approved despite that potential harm when the side effects are better than the disease cured by the drug. We will need to spend some time with guidelines and ethics, and hash them out with dialogs. We need to discuss the questions I have ignored today. (I am sorry about brushing off some issues, guys, but I hope we can have fun with them later.)

To be continued?

An Offer

by James Blodgett

We have done a few things to help our cause in this SIG, but frankly not much. I recommend that everybody try to help at least a little bit.

One way to help is to spread the word by discussing the issue. It is an interesting topic for conversations. Word of mouth can be effective advertising. This publication, EROSM, is becoming an interesting body of work. See the index of back issues, below. That index with links to individual issues is on our website, at http://www.global-risk-sig.org/pub.htm . One way to spread the word is to forward or post that webpage URL.

Another way to help is to delegate the work to others by donating a few dollars (or more) to help with their work. One possibility is the Lifeboat Foundation. I like the direction and the diversity of that group. I hang around there (online) as if it were an explorers' club between expeditions. Unfortunately, despite some efforts, they aren't doing much at present, but hopefully that will change, and there are not many organizations that are doing much better. They accept donations at: https://lifeboat.com/ex/donate#form .

Another good group is the Future of Humanity Institute at Oxford. They accept donations at: https://www.fhi.ox.ac.uk/support-fhi .

If one really cares about these things, it is helpful to get some skin in the game and try to make a personal or a collaborative project work, and try to make it really be helpful. The first part of the job is to think about what will help and what will work. The second part of the job is to try to make that work in the real world.

My offer is that we can try to do some of that here. I would be glad to discuss reader's efforts in this area. Send me an email. See our SIG contact page at http://www.global-risk-sig.org/contacts.htm. Perhaps I could make a suggestion or

two, or just be encouraging. Perhaps we could organize a group to work on an issue. Also, EROSM could be used as a test market, or as the beginning of an effort to publicize an issue. Note that we do have editorial standards. Also, until our impact factor increases, it would be good to target other publications and other venues as well. My general strategy is to try a lot of things in the hope that something works.

Index of Current and Previous Issues of [Existential Risk/Opportunity] Singularity Management (EROSM)

These issues are available at http://www.global-risk-sig.org/pub.htm . Introductions and unattributed articles are by James Blodgett, Editor.

October 2017 Issue

- What I am Doing Here Today
- Management of Positive and Negative Singularities as a "Postmodern" Book Sample
- An Offer
- Index of Current and Previous Issues of This Publication

July 2017 Issue

- Introduction to Keith Lofstrom's article
- Lunar Resources by Keith Lofstrom
- Introduction to Leon Neihouse's article
- Fail Safe Institute: Concept of Operations by Leon Neihouse

April 2017 Issue

- Introduction
- Dealing With Bastards
- Why This Story Is Relevant
- Other Fiction Related To Our Quest
- Friday The Thirteenth Drill

January 2017 Issue

- CRISPR: A Job For Us?
- Upcoming Topics: Space Industrialization & Runaway Global Warming

October 2016 Issue

- Alexei Turchin's Charts of Existential Risk/Opportunity Topics
- Interview with Alexei Turchin (containing an article by Turchin)

April 2016 Issue

- The "Uploading" Branch of Singularity Thought
- Interview With Jeff Bone

January 2016 Issue

- Existential Risk, Willard Wells, and Me
- Interview With Willard Wells
- Blodgett's Post On Lifeboat Foundation Discussion Section On Yahoo Groups, Quoted (In Part) In Wells' Book
- Proposed Annual Gathering Presentation by David L. Minger

November 2015 Special Issue

-Why We are Publishing A Special Edition
- A New Risk Attribution by Otto Rössler
- Reflecting on China's Ambition to Build the World's Most Powerful Supercollider by 2020 by Tom Kerwick and James Blodgett
- Errata

October 2015 Issue

- Review of Eric Johnson, "The Black Hole Case: The Injunction Against the End of the World," Tennessee Law Review, 2009. Reviewed by Justine Jones
- Struggling With The Ethical Limits of Expected Value Utilitarianism as Applied To Positive and Negative Singularities

July 2015 Issue

- Global Risk Reduction Heros
- How an Individual Can Improve Humanity's Odds
- The Wright Brothers as Role Models

April 2015 Issue

- Pep Talk
- Book Reviews as Things We Can Do
- Book Review: Nick Bostrom, SuperIntelligence, Paths, Dangers, Strategies. Reviewed by James Tankersley Jr.

January 2015 issue

- An Opportunity To Industrialize Space
- What This Publication is About