

# Existential Risk / Opportunity Singularity Management

April 30, 2021

## Contents:

- Management Versus Strategy

---

Copyright © 2021 Global Risk SIG. Both authors of articles and Global Risk SIG may reprint. This publication is produced by the Global Risk Reduction Special Interest Group, a SIG within US and International Mensa. Content expressed here does not reflect the opinions of Mensa, which has no opinions. To join Mensa or just see what it is about, visit <http://www.us.mensa.org> . This and past issues of this publication are available at: <http://www.global-risk-sig.org/pub.htm> .

---

## Management Versus Strategy

by James Blodgett

Management is part of the title of this publication (EROSM.) That title is deliberately hubristic. How can we manage something as humongous as a singularity? Our title is deliberately unfamiliar and hopefully intriguing, in the style of Tom Wolfe's early essays and books, for example: "The Electric Kool-Aid Acid Test," and "Mau-Mauing the Flak Catchers." Another Wolfe title was originally obscure test pilot lingo but is now resoundingly familiar because of Wolfe's book, "The Right Stuff." I hoped for a similarly intriguing title for our publication.

In one sense we can try to manage a singularity. However, we can't easily apply many of the standard tools of management, such as operations research, because we usually don't have enough data and we usually don't have enough control. The main part of management that we can apply is strategy.

Strategy is a military term. Armies try to manage things that are as uncertain as a singularity. Battles are uncertain because opponents rarely announce their strategies in advance. Strategies have to be flexible, so that they can be adjusted to the moves of the enemy. The military has to make plans with relatively little information about what the enemy intends, and therefore tries to consider strategic actions that confer advantages no matter what the enemy does. Good strategies are important. Many famous generals in history have been good with strategy. A review of Alexander the Great's strategy is available at <https://apps.dtic.mil/dtic/tr/fulltext/u2/a144027.pdf> . He conquered much of the

ancient world. Admiral Farragut sailed his fleet through a minefield to avoid a bigger risk, because he knew that the mines of his day were likely to waterlog, and won the battle of Mobile Bay. Others were less successful. General McClellan avoided battle several times because his intelligence officer had a methodology for estimating the number of enemy troops, and the methodology overestimated, so that McClellan thought he was outnumbered when in fact he was not. Lincoln telegraphed, “If you are not going to use your army, may I borrow it?” Sometimes the best generals get it wrong. One of the best Confederate Generals, Robert E. Lee, ordered Picket’s charge, because of plausible strategic considerations. The attack did not succeed, and it decimated the Confederate Army at Gettysburg.

Singularities have larger outcomes in terms of human lives than most battles, so we should try hard to get our strategy right. However, we rarely have the command authority of generals, so we have to get it right intellectually and persuade others to go along. Some heroes of history have been intellectuals and scientists, so this is not impossible. In singularity management the first strategic issue is often how to sell people on the prospect of a potential singularity, so that they become motivated to conceptualize, research, develop, invest in, and use singularity technology, or to avoid a bad singularity.

Existence proofs help with motivation. They do not really prove that the required technology exists. We need the ninth level of the NASA technical readiness scale to prove that. But they do show that something like the required technology exists, which is motivating because it justifies hope.

In the July 2020 issue of EROSM, I presented seed ships as an existence proof that something like settling the reachable universe might be possible. However, an existence proof is not a proof, it is only a demonstration that something like the desired technology is possible. I calculated a humongous expected value (probability times value) for this outcome, even if the probability for achieving it is low. Later in the January 2021 issue, I backed off a bit from this prediction, admitting that the form of nanotech I had used as an existence proof is somewhat different from the type that we probably need.

Now I have even more reasons to be dubious. In the previous articles, I ignored side effects, because they are difficult to predict, and might go either way. Recently I have been thinking about side effects, and I can think of several ways that they might go wrong. One management tool to address this is to compute expected value. Expected value means the value of an outcome times the probability of that outcome. If you receive six dollars if you roll a one with a die (singular of dice), then the expected value of being able to roll that die is \$1. However, if you gain or lose money if you roll other numbers, those outcomes have to be included as part of expected value. Therefore, a

better calculation of the expected value of settling the universe should include side effects of our intervention to make it happen. However, that calculation is difficult to make, because there are an almost infinite number of possibilities of how things might turn out. As a first cut, I have often ignored side effects. Many are unknowable and could go in either direction. However, some are knowable in the sense that we could write a science fiction story about them. We could make a very rough estimate of the probability of actualization of the events of that story, but it is unlikely that our estimate would be very accurate. Estimators tend to have biases. It is also difficult to make an estimate of each of the almost infinite number of possible stories about the universe that could be produced by the intervention we introduce to produce our singularity. Perhaps we could sort the stories into groups and estimate the probability of each group of stories, and the value of living in them, and produce a very rough estimate of the expected value of living in the universe created by our intervention. Then we could compare this with similar stories that might occur in a universe without our intervention. As methodology this seems dubious. There are strong tendencies of story tellers to be optimistic, for example in Star Trek and in much of science fiction, and strong tendencies to be pessimistic, for example in Frankenstein and similar stories, and more recently in the thought experiment of the paperclip maximizing super AI. That is why it seems appropriate to focus on strategy rather than management for this exercise, until our technology is higher on NASA's technical readiness level scale so we can make better estimates. We should focus on the benefits, because they are motivating, but also the risks, because safety is important,

We might hesitate to intervene, but doing nothing also has consequences. Our hypothetical intervention is a form of steering. When we steer a car across town, we are likely to get there. Accidents are possible, but occur infrequently. The problem with singularities is that they are harder to steer than a car, and accidents are more likely. Sometimes it may be better to give up. However, steering does sometimes work, and if we are careful enough, we can make steering safer. It seems better to try to steer than to default to the alternative.