

## **Alternative Models for Managing Self-Replicating Nanotechnology, Including Analysis of Foresight Institute Guidelines Applied to Land, Sea & Space-Based Scenarios In Which Self-Replicating Nanotechnology Is Necessary**

Talk at 1st Annual Workshop on Geoethical Nanotechnology, July 20, 2005 by Dr. Martine Rothblatt

Dr. Martine Rothblatt: Well let me roll right into the next presentation: Alternative Models for Managing Self-Replicating Nanotechnology -- specifically an analysis of the Foresight Institute guidelines. Molecular nanotechnology, as all of us know -- but I'll be a little bit more explicative for the benefit of the people on the webcast -- is the ability to program matter with molecular precision. Not the same thing as nano-dimensional materials, and does not imply self-replication, which is itself a subset.

The point has been made by several of the speakers today that self-replication is unlikely for MNT manufacturing, at least in the near term, for various reasons. And I take no issue with those reasons; basic manufacturing self-replicating technology may either be unnecessary or just unnecessary in the near term, as Ray indicated. More is needed in the longer term.

The Foresight Institute's orientation is to educate and create shared understandings; to develop community controls; maximize research, development and commercialization, distribute MNT benefits to the third world and to help end poverty; respect ecological and public health principles and develop a means to restrict the misuse of MNT internationally. This is a distillation of several pages of MNT background and preamble.

The intent of their guidelines, they developed a set of guidelines to help achieve the orientation that they have, was that these guidelines could, over time and with subsequent iteration, become sufficiently specific that they could form the basis for a legally enforceable framework, and that this framework would consist of things such as lab certifications, random open inspections, professional society norms, insurance requirements, stiff legal and economic penalties, and other sanctions. So these are the intent of what the guidelines might ultimately become.

In a nutshell, or as you'll see I worry perhaps, in an eggshell, the actual guidelines themselves, we've talked about their orientation and their intent, are to be flexible -- the preamble emphasizes a need for flexibility. And that with regard to the principles, the main point is that because nanotechnology risks vary, nanotechnology consequences should vary -- very much using the kind of sentential if/they consequence thinking that Wyre Sententia mentioned to us earlier. So things such as legal liability, market costs, built-in safety measures -- all of this should be adjusted to the level of risk that the particular form of molecular nanotechnology entails. So they envision somehow scaling all of the various risks of molecular technology in the way that there could be scaled sanctions or guidelines or economic costs, deprivation of market access, etc., depending on the risks.

There's also certain development principles built into the Foresight Institute guidelines. The one which is of most interest to this presentation is that there be no uncontrolled replication. In fact, basically, self-replication outside of a controlled environment is completely proscribed under the Foresight Institute guidelines. And that finally, there are design guidelines which have been

developed that discourage mutation and discourage access. The real question is, are these really solid guidelines? Like a nutshell, you think of a walnut nutshell as hard, or are these guidelines really more like an eggshell, that as soon as you get to the real world and people start developing real technology, that they're just going to like crack and fracture and turns out that you just get messy yolk all over yourself?

The guidelines are also expressed as scorecards for professionals, for industry, and for government policy. Basically what the scorecards do is they allow people to score themselves on a scale of zero to five, along these three different types of guidelines. Exactly what's done with the score, it really isn't said in the guidelines. You could have a high score, a medium score, a low score; the guidelines give you no guidance in terms of what to do with your score.

Here's a flowcharting of the Foresight Institute guidelines that, I think pretty clearly identifies what they address and what they don't address. If we start with this Roman numeral block I here, this decision block, the first question to ask is whether or not MNT is self-replicative. If the answer is no, then under the Foresight Institute guidelines, in a nutshell, you could say anything which is ecologically, health, in terms of especially health for the workers, but public health generally, and national-security-friendly is okay. So as long as no self-replication is involved, anything is okay that meets these three criteria.

If there is self-replication involved, the next question to address is the one in diamond two, which is whether or not the self-replication occurs in a controlled environment. If the answer to that is no, the guidelines are quite clear, it's banned. If the answer is yes, you move onto decision block three, and decision block three says Is there replication audit, meaning like where have the nanobots been and what are they doing and what do they know, mutation, and access? Are all these things highly controllable even within the controlled environment? If the answer is no, again, it's banned, even though it's occurring inside a controlled environment. If it doesn't have these type of controls, that technology is banned under the Foresight Institute guidelines. If yes, it is okay so long as the party that is operating that technology is an ethical party. So it's a very highly restricted access criteria to who is accessing this technology, and exactly who is defined as an ethical party. It will probably be a party that had an almost maximum score on all of those charts that were mentioned earlier, the scorecards.

Here's a quick critique of the Foresight Institute guidelines, based on that flowchart parsing of them. The guidelines seek specificity in terms of exactly the level of consequence, whether it's a marketplace access consequence, and economic penalty consequence, or a legal sanction. They seek a lot of specificity based on the nature of the risk, but yet the overall philosophy is one that asks for minimal regulation. My experience, as a lawyer at least, is that those two goals can be very antithetical. The more you try to spell something out, and in fact the reason why the code of federal regulations is much longer than any of the law books, is because they try to explicate every different possibility in all of our different areas of our economy and our lives. So when you try to get specificity, it's going to be very quick that you lose minimality as your regulation.

Another criterion is while the Foresight Institute guidelines want minimal regulations, they actually include strong anti-misuse penalties and even criminal sanctions. They recommend criminal sanctions for misuse, so there's a little bit of inconsistency there as well. While the

Foresight Institute guidelines make rather clear, as well as some of their recent publications, that self-replication is unnecessary, unneeded, and even unlikely, in their point of view, it would be uneconomic and therefore unlikely. In fact, the guidelines are overwhelmingly consumed with the issue of self-replication to the point of 10 of the 27 guidelines deal explicitly only with self-replication, an additional six of them indirectly, so basically more than half of the Institute guidelines go either explicitly or implicitly with self-regulation.

So one might think that while they say it's unneeded, unlikely, unnecessary, maybe the authors of the guidelines scenario aren't very well equipped to deal with the "black goo" scenario. Instead, our fear of man-caused harm which we see every day overwhelmingly dominates our thinking, even though it's only one Terry Schiavo, or one person gets shot in L.A., we obsess about that. It's horrible that 52 people died in London. It's horrible that over 3000 died in the World Trade Center. It's also a fact that 150,000 died in the tidal waves or the tsunamis that followed in the wake of a volcanic eruption, 20,000 people die every single day of hunger. So, you know, how do we deal with these situations?

I think it's also possible that the Foresight Institute guideline's authors actually have conceived a clever tradeoff, which is: trade what we don't have, which is world-saving molecular nanotechnology, they all realize, we all realize that at this moment we can't open up any drawer and pull out the kind of Mulhall scenario-saving molecular nanotechnology. That stuff is not available today.

So why not tell all the people who are afraid of nanotechnology we don't want self-replicating nanotechnology? In the meantime being a safe harbor to develop non-self-replicating nanotechnology so we can build a big technology base and gain a certain level of benefit from MNT, and basically use that safe harbor ban on self-replicating MNT. This is nowhere said in the Foresight Institute guidelines, but one could read this plan, I think, into the guidelines.

To me, there's an alternative approach, which I call the "geoethical approach," and it's based upon a threshold analysis. The geoethical approach is appropriate for what I call "Your life or mine" type of questions. Saving my life may endanger your life. It applies to global impact technology and MNT seems to be a good candidate. The basic three principles of geoethics are: take technology forward if and only if it can meet three principles:

One that it helps relatively most those who are worse off, on the logic that those who are worst off will be the ones who suffer the worst from the unintended risks or consequences of the technology. That the risks are consented to by all those people who may be affected, either directly or through their representatives. And finally that the system has a way to self-finance or audit its compliance so that agreements are not made on paper, only to be violated shortly afterwards.

These principles can be distilled from principles that have been laid out by several modern philosophers: John Rawls, Jurgen Habermas, individuals like that. So an example might be: Should we permit open-system disaster-mitigation self-replicators? Benefit: who would it help those most at risk? Yes, the whole purpose of developing self-replicating system for disaster mitigation is to help those people who are most at risk, those who are suffering from the disaster.

Consented to by all who may be affected? You would need representative organization, and I'll get to that in just a moment. But the answer is, you would need to have the representatives of these horribly affected people consent to it. And finally, who would ensure compliance? You would need a compliance-monitoring organization.

So what I would propose is a geoethical implementation, would be to form via treaty an organization that would call IntelRep, International Self-Replicating Technology Organization. And via treaty, to make this organization be exclusive, give it exclusive worldwide rights to self-replicator production rights. After all, the mainstream view is just to ban the technology at all, so it seems if the mainstream is just to ban it, it would be better to give to an organization to develop it rather than to just totally ban it.

Every country and corporation can be a member, and this organization would finance itself by selling to its members the products and the benefits of self-replicating technology for resale. It would also, in turn, as part of the treaty, if you want to be part of IntelRep, you have to agree that IntelRep can monitor companies within your country that are doing molecular nanotechnology, which is very much in line with what the Foresight Institute recommends.

The IntelRep-level functionality, in other words the type of functionality that would occur on the level of an international organization, IntelRep, it would be a treaty-based nongovernmental organization. Every country would be open to have an ownership stake in it. Some countries could decide that they would appoint a private company, like we used to have Comsat here in the United States to be its representative in IntelSat, so it would always have to be a government entity.

Labs, I believe, should be set up in both Asian, American, and Euro-African land masses so there could actually be competitive juices flowing at the different labs within IntelRep would be set up to compete with each other, so that we could move the technology forward as much as possible, all under the IntelRep umbrella. The competences of this international organization would be to develop best NMT practices, which is something Foresight guidelines is mostly about. Global NMT inspections, which is another thing that the Foresight Institute guidelines calls for.

Self-replicating technology development, that's something that the Foresight guidelines are not opposed to, although they want it to be done only within a controlled environment and with a responsible party. This would obviously be a responsible, ethical party, having been formed by a treaty of all these countries. Fail/safe controls, they would develop that, and sales applications; develop ways to sell the services to their countries, much like today the Intelsat organization sells its communications satellite services to satellite organizations in each county, which then resell them to users.

The member-level functionality, so this is each country or a company or set of companies or group of companies within each country, would receive a nominal share just by signing up to the treaty organizations, but they can buy more. So a country like the U.S. would form an organization, it could buy maybe 20, 25, 30% of IntelRep because it would see that there would be huge revenues and profits coming from this organization, so they could buy more. But every country would get at least a nominal share. The members would enforce IntelRep domestically,

providing training opportunities. This is a great way that every country in the world could send people here to IntelRep to get training and begin to diffuse molecular nanotechnology worldwide, such as the Women's Nanotechnology Initiative, and receive a turn on investment from self-replication product sales.

Getting it started, I would recommend that the leading MNT firms and NGOs like CRN and the Center for Cognitive Liberty and WTA and Extropy Institute, leading firms could co-draft a straw man treaty, and it sounds like some work was already being done that way at Arizona State University. Get it blessed by their state ministries, open to ratification; enter into a force with a few signatures. As soon as some countries see that the U.S., Britain, you know, China are signing up to this treaty, I think you'll see every other country signing up to it right away. The Intelsat treaty, which again, is somewhat similar, has virtually every country in the world signed up for it and buying their share of that organization. Commence work right away and continually court all states with the goal of having every country in the world a member of it.

The sales pitch would be that membership is free, you get a chance to help shape MNT replication policies, you get high-level training opportunities for people, perhaps competitive lab located in your jurisdiction, access to the benefits for your country, and a revenue opportunity for a return on the investment.

Responsibilities if you sign up: you've got to enforce the IntelRep self-replication exclusive nationally. So if people try to break into the self-replicating things, you've got to enforce this treaty obligation, sue those people in your country, or whatever. Screen potential customers within your country for responsibility level or signs of misuse. These are part of the standards that the Foresight guidelines look to, and disseminate best practices nationally. Again, another Foresight Institute objective.

So in conclusion, IntelRep meets Foresight very nicely, just like these two little cats. It provides international control against misuse, minimal red tape, third-world benefits, sales restricted to responsible actors, self-replication in open systems is possible (that's where it departs from Foresight and goes beyond Foresight, and I think it helps save the world there), misuse is actionable, and ongoing best practices study and modeling.

I've talked about how the financing works. It's interesting to see why I think countries would jump at an opportunity to invest in this. For example, if we say, for example, that the net present value of self-replicating technology is one trillion dollars twenty years from now -- so I'm not trying to say that if you had a worldwide exclusive right on self-replicating technology, you know, I'm not saying that you could get a trillion dollars for today, but I would bet in 20 years, you could easily get a trillion dollars, and probably far more, for an exclusive worldwide right to self-replicating technology.

Well, if that's true in 20 years, which I think is a very conservative assessment, at a very high discount rate, extremely high discount rate, 20%, so that's saying really self-replicating nanotechnology is never going to occur in 20 years, that's such a high discount rate. That's equivalent to saying that it's got a 20 billion dollar net present value. That exclusive is worth 20 billion dollars today, 200 billion dollars at a much more realistic 12% discount rate, so you could

go to the national, to the world capital markets, you could do an IPO for a substantial portion of that and easily finance all of the obligations described here, get three super-well-financed self-replicating labs set up around the world, put them in competition with each other under management, and get self-replicating nanotechnology on the fast track.

Other geoethical benchmarks: it would be helping third-world countries; we'd meet the risk principle, because by signing up to be a member of IntelRep, you're consulting to self-replicating technology, so we've met the risk principle. We've met the benefit principle. And, because it self-finances itself, we've met the assurance principle.

Why does the geoethical approach work so nice? Barry, thank you for that graphic there, you'll recognize that, these are astrobiology graphics there. The geoethical principle works because it sequesters the worse risks in a few highly visible places. That's why this approach always works with these kind of risks. It gives everyone an ownership of both those risks and the rewards. It gives risk managers a strong incentive for pursuing rewards so you avoid this bureaucrat mentality when the government is not in business to make money, I agree totally with Max, that the government just says no, no, no to everything. They're very, you know, risk-averse.

But here you put a quasi-government organization on a profit-making track to make rewards, and then they are incentivised to, in fact, take risks. And finally it balances incentives with controls by agents of the risked population. So, because all the world's countries are members of this, there's a balance of the risk and the reward.

In summary, I think the Foresight Guidelines are actually very clever, more clever than they spell out in black and white. But I believe that they can be improved upon and subsumed. Geoethics resolves the open system hold, which is a huge hole. IntelRep should be organized now, it's time to pass the baton. Onwards!

Dr. Martine Rothblatt: Chris and Ray.

Chris Phoenix: I want to respond a little bit of length. Is this going to be picked up through the transcription?

Dr. Martine Rothblatt: Yes. We're online. Right now we're online.

Chris Phoenix: Yeah, that's good. The Foresight guidelines, I wasn't involved in any of the revisions, but I think they can best be understood as something that was done over time and self-rep focus of them was from the earlier versions, so that the apparent contradiction, I think, was less intentional or clever than simply historical. As to whether self-rep is necessary for the disaster scenarios, I honestly do not believe it is. I believe that a non-self-rep solution is not only possible, but better than a self-rep solution for at least the three disasters you listed: asteroids, super volcanoes and tsunami cleanup. The InterRep organization that you described had lots and lots of good ideas for administering nanofactories, and we'll be studying that very closely. I think if you took self-rep out of the organization, it wouldn't hurt a thing and would make it more palatable and it would still be a very valuable approach. So basically, I think the focus on why is

self-rep in the guidelines probably is a distraction from how we can deal with very, very powerful technology, but your ideas work for that just as well.

Dr. Martine Rothblatt: Okay. Good. Well, Ray?

Ray Kurzweil: A comment on that: I think ultimately we will need self-replication for an ironic reason, which is to guard against pathological self-replication. We've debated this, and I think it's actually the crucial issue and I agree that in the early stages, probably a non-self-replicating nanotech immune system will work, but ultimately I think you will need self-replication for the same reason biology discovered that through evolution: that you need self-replication, you're not going to be able to get defenders everywhere they need to be in time, ultimately whether they're clever enough. Because that's ultimately possible, we're going to have to develop a nanotech immune system and that's going to have to have self-replication. And that's a disaster scenario that needs to be prepared for.

And then there's a further irony in it: from a software pathogen, you could turn this nanotech immune system into a self-destroyer. All the more reason to have a geoethical organization that can deal with this subtlety because whether the whole thing is, whether your immune system is friendly or not, and you can develop an autoimmune disorder is of profound importance. Really, a subtle distinction with catastrophic implications.

So I like this proposal. We need to develop some kind of organization that has the wisdom to develop these solutions, because there's going to be subtle differences between protection, defense, and offense. We're going to want the development of an immune system to be in responsible hands.

Dr. Martine Rothblatt: Okay, great, thank you Ray. I think I saw your hand up and Chris, so why don't we do Doug and Chris and then we'll talk all night offline, but remember, first stop at the porch for the group photo. So Doug, Chris, and then we'll go offline after that.

Doug Mulhall: Online, I want to say that this, in the years that I've been studying, this is the first rational approach that I've see that is possible workable to regulate MNT. I really want to congratulate you on putting this forward. This is a big step in putting this forward, and I'm very serious about that. The other thing is, I completely agree with Ray, we've got to deal with replication right up front, and I have another reason, and that is, I think that with emulating DNA processes, whether we like it or not, we're going to be faced with self-replication very quickly.

We've got DNA transistors out there already. I mean, we're going to have DNA emulation by self-replicating nanotech, I think, a lot faster than most people realize and we have to deal with it upfront, and I believe that Foresight has responded somewhat politically, has pulled back from it because of all the attacks, and I think we should be more upfront and say we're going to have this, we need to deal with it, and not be so timid about it.

Dr. Martine Rothblatt: Thank you Doug. Chris you get the last word in the geoethical nanotechnology workshop.

Chris Phoenix: Okay, three really quick points. First, the use of the DNA and DNA transistors has nothing to do with self-rep. It does not raise any risks in that direction.

Second, at some point we may have to deal with smart self-rep, but first we have to deal with stupid self-rep, which does not need self-rep to deal with it. You can deal with stupid self-rep with non-self-rep technologies. By the time smart self-rep arrives, a) we'll have some warning, and b) we'll have at least augmented humans to help us design better solutions than we can design here.

Third, on the topic of AIs, as I heard a lot of people assuming that if you can make something that's intelligent and reasonably well designed, we're home free. You have to design something that is better, has more common sense than an autistic human and that narrows the target quite a lot. If you can't do something that's better than an autistic human, then it's not really safely useable.

Dr. Martine Rothblatt: All right. Well, I thank you, thank you for your contributions during the day, Chris, I was really glad to have you here. And thank all of you for being here. We all deserve to give ourselves a big round of applause. [Applause] Thank you, Amara and Virtual Studio for bringing us online to the world. Thank you very much. [Applause]