FutureNovo Interviews – Vernor Vinge

FutureNovo – Anticipating things to come

by Richard Yonck, FutureNovo
August 10, 2008

Mathematics professor, computer scientist and five time Hugo Award–winning author, Vernor Vinge is a man who's never short of ideas. As well as creating a wealth of thought–provoking stories over the course of more than four decades, Dr. Vinge wrote a seminal paper in 1993 about accelerating change entitled "The Coming Technological Singularity: How to Survive in the Post–Human Era". In it, he describes the likely development of a superintelligence which would bring about a period of extreme technological and social change. This event he dubbed the "Singularity" owing to its similarity to cosmological singularities, more commonly known as black holes. Just as nothing – not even light – can escape the event horizon of a black hole, Vinge surmised that it was impossible to anticipate what life might be like beyond the occurrence of the Singularity.

FN: Dr. Vinge, thank you for taking the time to speak with me. I've read your stories and ideas over the years and enjoy them very much.

In your 1993 paper on “The Coming Technological Singularity,” you talk about several possible paths that could lead to a superhuman intelligence, some entirely machine–based while others involve electronic or biological enhancement of human intelligence. Do you currently see one of these as most likely and fifteen years on from that paper, how have your views about this changed?

Vinge: Occasionally, I look at that paper to make sure of what I said there. I've been impressed by the fact there hasn't been very much change in my opinions. Of course, that can be a bad thing, but that's the way it is in this case. As to which of the four paths I listed is most likely, I think as the years go by one or another will seem more plausible. The Singularity will probably comes as some combination of different approaches, and it's possible that the transformation will play out differently than the major life changes of the last few hundred million years. Very likely this will not be a winner take all situation – you know, where after the first solution occurs, none of the others can happen. Intelligence may value thought done in different ways, on different substrates. All of the possible paths to superintelligence may be exploited. It is true that the notion of networked, embedded microprocessors seems more plausible to me now than it did in 1993. This notion of a Digital Gaia – the internet of embedded micros – is a trend that's going like gangbusters, and the idea suggests all sorts of symptoms to track. In Digital Gaia, each participating physical object is smart in the sense that it knows what it is, where it is and it can talk to nearby neighbors, in potential to the world. If there are enough such objects, it's as though the world itself wakes; reality becomes its own database. (See Karl Schroeder's notion of "thalience" in his novel Ventus.) This is quite a contrast to the centralized collection of data in server farms – which is also going like gangbusters.

FN: It seems from what you're saying and things I've thought about that the fine–grained level of sensor distribution almost emulates or is similar in certain respects to the concept of cells within a body where they have a certain degree of capability or intelligence, but certainly become far more when combined into a larger organism.

Vinge: It's a very life–like thing, almost a non–carbon based parallel to the ecosystem that already exists. The ensemble could have vastly superhuman intelligence, but with a very different flavor than our usual

transhuman view of the post–Singularity world. When I need an example of how different a superhuman intelligence could be, two extreme conjecture points suggest themselves: Intelligence Amplification of our own human intelligence ("people–plus") versus this Digital Gaia notion. These are visions as different as plants and animals in the biological world.

FN: And any number of them could come about, not necessarily concurrently, but one after another or in parallel or in conjunction with one another. There could be an entire spectrum of different intelligences that are very, very different from each other.

Vinge: Yes.

FN: You've stated in your paper and elsewhere that you think the Singularity will likely occur sometime before 2030. While hardware advances have been on a pretty consistent, even accelerating track in terms of Moore's Law, do you think software advances are keeping sufficient pace to meet this time frame?

Vinge: Ah, that's the critical question for Singularity claims. I think that the Singularity is the most likely non–catastrophic event for this century and I'd be surprised if it doesn't happen by 2030. But if the software can't exploit hardware advances – put another way, if we never figure out how to put the parts together – then the hardware power may be an empty thing. So far, this question is less subject to objective argument than hardware claims. So tracking just what we are doing with the hardware is very interesting. There are various sorts of research that might give a person some added confidence – or not – about how things are going to come out.

FN: So you see it as not necessarily being as definitive or well defined a trend as the hardware advances?

Vinge: Correct. Have you read the essay that I had in IEEE Spectrum last month?

FN: Yes, in fact that's part of my next question.

Vinge: When thinking about the possibility of the Singularity, there are indicators to track. That was actually the title of the Spectrum essay.

FN: Right. "Signs of the Singularity". (http://www.spectrum.ieee.org/jun08 /6306) In it you responded to several articles published in that issue's special report on the Singularity. Does the fact that an increasing number of scientists and academics are discussing this give you hope that we may be better able to direct and create safeguards for the development of a superintelligence?

Vinge: Yes.

FN: That was very succinct.

Vinge: The question is of course, how much added confidence does it give?

FN: And how much can we really control something like this?

Vinge: There is increasingly sophisticated thought being put into safety and the meta–question of what safety means. As time goes on – until either the Singularity happens or some sort of disaster strikes – I would expect these issues to become more and more widely discussed. As I said in the Spectrum article, this may be annoying to people who are convinced that the Singularity is totally bogus or that talking about it is distracting from the things we should really be discussing.

FN: I suppose there are always people who have their own agenda and don't want to be dealing with anything else but it's certainly prudent to be thinking about it very seriously if there's even a possibility of it

occurring. In terms of the Singularity, you've spoken about a soft take–off being preferable to a hard take–off which I presume most of us would agree with. But while there would be some lead up to the Singularity, doesn't the exponential quality of the concept really make a hard take–off almost inevitable?

Vinge: Maybe not inevitable, but unnervingly plausible: In real processes, exponential growth very often has complications that either terminate the exponential growth (with saturations or catastrophic collapse) or cause the exponent to change. For the Singularity, one point to make is that we have certain underlying hardware trends that are exponential. The way that manifests can be in things that are of very surprisingly different speed. So for instance you can imagine having an exponential where the things in the environment that the exponential trend are changing reach a critical point where some really extreme change can happen that has a time constant that is much, much shorter. Here the exponential trends are hardware improvement and the interesting collateral issue is the possibility of human and superhuman level machine intelligence. When/if that possibility manifests, the consequences could be much more spectacular changes and much faster change rates.

Trying to argue that the transition could be slower, that it's something that could take decades – the soft take–off version – I think that before I read Greg Bear's Blood Music, I was not into hard take–offs. Ray Kurzweil and Hans Moravec have written about transitions that take three or four or five decades. But I think in the old days I wasn't into that either. I thought that what would happen is we'd reach a certain point and there would be a runaway.

FN: I guess that's what I'm thinking with that question.

Vinge: Actually, a runaway that would happen in a year or two, that's not a hard take–off.

FN: You wouldn't see that as a hard take–off? I know you've spoken in terms of a few hundreds of hours at times in the past as a hard take–off.

Vinge: Aha yes, we may be talking about the same thing. My characterization of a hard take–off should be that whether it takes a few seconds or a year, it proceeds far faster than contemporary feedback mechanisms can cope with. Such speeds seem plausible.

As to how to justify sliding into superhuman intelligence across decades; that wasn't my original intuition but it's worth serious imagining. Here's one possible mechanism: Skeptics often say to Singularity enthusiasts, "Life and intelligence and consciousness are immensely more complicated than you imagine." What if that's correct? At the same time what if the enthusiasts about exponential improvements are also right? In other words, all the oomph of those exponential improvements is needed to meet the arguments of the naysayers.

FN: And that's what's acting as a brake or a governor or limiting factor on the process.

Vinge: Right. It's a credible scenario that it would take decades to go from the beginnings of parity to unquestionable total intellectual superiority. Thinking with scenarios is good. Scenarios should even – maybe especially – be considered for things you regard as very unlikely. For instance, pretend it's 2050 in a world where the Singularity doesn't happen and you've been asked to write an essay explaining why in retrospect that outcome should have been obvious! This exercise of the imagination gives added insight about things to watch for.

A soft takeoff, a decades–long transition with enough time for our institutions to adapt and respond – that would be very cool. Our feedback mechanisms could grow right along with the scope of the problems we face. I mentioned the writings of Ray Kurzweil and Hans Moravec. Also, J. Storrs Hall had a book called Beyond AI which is really about how to handle a long transition. In his vision of it, the early human–

equivalents and superhuman–equivalents are very, very expensive sites. Very large sites. And they do not undergo an exponential runaway. They are so expensive that they are only targeted on really major issues. (As an aside, I think one thing that could slow progress for the earliest human parity hardware will be that the devices are so expensive – and so disbelieved – that they won't even be exploited as AIs. We may have to wait till Moore's Law gets these devices cheap enough for university labs and small start–ups.)

FN: So there really are a lot of factors that could influence that.

Vinge: Right.

FN: In 1981, you wrote "True Names", one of the earliest stories to describe a recognizable, if highly advanced version of cyberspace. The virtual world you called "The Other Plane" influenced dozens of authors and anticipated virtual worlds such as Second Life by over two decades. Given where computing technology was in the early eighties, that's quite a conceptual leap. Which trends were you seeing at the time that led to the ideas behind the story?

Vinge: There's a truism among science–fiction writers that science fiction may look like it's talking about the future but really it's just a mirror of the present. There is a large grain of truth in that, but it would be more precise to say that each SF story is a reflection of the present of the story's author. And my present from the early 1950s was someone immersed in science and tech speculation. From the late 1950s on, this included computers. I was surrounded by things that contained elements of the "True Names" environment. For one thing, I'd been exposed to fantasies in which "true names" were very important – Ursula LeGuin's Earthsea stories, for instance. That gave me the notion that if everything had a serial number, knowing the serial number would be power. Then just before I wrote "True Names", we got dial–in access to the university computer. (An awesome machine with something like a quarter megabyte of RAM!) I could check out a portable teletype–style terminal and modem. I could log in from home! This wasn't really the Internet, just a very ad hoc phone link. However, that was enough that I could imagine consequences. I remember being logged in. Some other user ran the talk program on me and we had this anonymous real–time chat. After I finished the conversation, I realized that by my standards I had just lived a science fiction story. It was very easy to elaborate on the environment. So "True Names" was actually very easy to write.

FN: Amazing. One of the features in the story I liked was the idea of the high–level programming paradigm that you used that's since been referred to elsewhere as intention–based programming. Though there have been attempts at developing something like this, do you think it's feasible that we'll ever implement a programming paradigm as seemingly intuitive as you described in the story?

Vinge: I think we're going to get tools – and we already have tools – that make it easy to chunk very large pieces of functionality. On the other hand, my suspicion – and I think a lot of people have this suspicion – is that the way massive multi–core systems will solve problems that aren't naturally accessible to parallel processing will be via bioscience paradigms more than engineering paradigms. Exploitation of massive parallelism will probably involve cognitive biomemetics. This is scary to some people (sometimes including me!) because we grew up with the ideal of determinism in programs, especially software for life critical systems. On the other hand, before computers, people routinely depended on such unreliable systems (horses, dogs, etc.).

FN: Something that has it's own volition.

Vinge: Yes, an animal has its volition, even if it's below human parity. We know that in life, there are important gotchas we haven't figure out, or that are beyond our control.

FN: You mean in a highly complex system?

Vinge: In life. Leave aside classical dynamics and linear algebra and other stuff where parallel works fine in a deterministic way. Let's talk about the things nobody right now can figure out how to do. If it turns out that the biological paradigm is the way to go in that, that's very scary to some people. I can remember a year or two ago, I was at Hot Chips (a leading conference on high–performance microprocessors) and this came up in a slightly different context: It may well be that as things get faster and faster and smaller and smaller that we'll just have to go for non–deterministic computing at the transistor level. That raised the issue: "but what then about life–critical systems?" There was one guy on the panel who raised exactly the point I just made at the very high integrated level that you and I are talking about – the dogs and horses example.

Consider a cavalryman riding along on his horse. What if the horse sees another horse that it wants to associate with? It might just turn aside. So I don't mean to pooh–pooh life–critical system risks. I'm just saying this is nothing new. As usual, the question comes down to levels of risk and levels of benefit.

FN: Ubiquitous computing and augmented reality were very prominent in "Fast Times at Fairmont High" and "Rainbows End". In many ways, they are going to directly and indirectly enhance human intelligence, but at the same time bombard us with a tremendous amount of input. How do you see us managing this sensory overload?

Vinge: If the Singularity does not happen, then one of the symptoms we would see is that hardware progress makes the data glut problem worse and worse. In some areas this actually would be tolerable, although it might produce a form of science such as Chris Anderson was talking about recently in Wired. ("The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", Wired Magazine, 16.07) In other areas, it would just mean we have to give up on many projects, not being able to cope with the amount of data (much less create the associated information and knowledge). This would then decrease the demand for hardware improvements. So if the Singularity doesn't happen and it's 2050 and somebody's writing that hypothetical essay about why it was obvious that the Singularity never happened, the data glut would be part of the story arc.

On the other hand, the thing to watch as data collection increases is the extent to which automation is capable of dealing with it. Maybe 99.9999% of the data that's being collected is never seen by a human. But that data informs automation that ultimately is useful for decisions made. If this happens, one could argue that it is evidence of a shift in the cognitive–creative center of gravity, away from where it's been within humanity, off into our extended mentation. The sensory overload question is one of the really cool things to track.

FN: A signpost, essentially.

Vinge: Yes.

FN: Education is a recurring topic and theme in your stories. In several of them, students have assignments and tests that prepare them for a very different world, one where extreme change is expected and anticipated. There are Nobel–level scientists going back to school, struggling to keep up and fit in with the new fields and technologies while teenage students quickly take up techniques and concepts that were once totally cutting edge, even unheard of. How do you see our real–world education system having to transform itself to deal with the reality of accelerating change?

Vinge: That was one of the most intimidating things about writing Rainbows End; I'm not full of answers. First of all, while I think that children can be empowered, you recall that the children in both "Fast Times" and Rainbows End felt very intimidated. They were capable of boggling achievements – by our standards. For instance, the boy in Rainbows End could do real–time physical simulations for which a 2008 person would need years of hands–on experience and/or be very knowledgeable in physics and mechanics. But he

didn't feel empowered. Basically he felt that any kid could do that and there were other things going on that were constantly moving away from him. Education was a real stress point in Rainbows End. I had the junior high school teacher, Ms. Chumlig, make the point that "You have to know something about something." It was one of her mantras. I think it's correct as long as humans are significant players. One perversely important category of knowledge is knowing what the glitches are in the various automation and filters that you're using. I used to teach numerical analysis and we'd deliberately hand out ordinary looking problems that would break calculators and computers – that is, cause automatic calculation to give wildly wrong answers. Mainly we exploited discretization errors. The goal was to teach what the traps and threats are. However, similar issues exist with other packages: computer algebra systems have their own world of glitches. These are not classical bugs, but they're issues you have to understand to be successful. Mastering such problems is very domain–related and in some cases it may change as the package changes. How this sort of thing interferes with our exploitation of software is going to be interesting to track – until we have packages so smart that they understand the psychology of the human user.

FN: That would certainly be helpful.

Vinge: True. Of course, when the apps get that clever, you've slipped over the edge. It's Singularity city.

FN: There's a real emphasis in your fictional schools on collaboration and affiliations or "affiliances" as you termed it in one story. I take it you see this as an important means of dealing with working in a changing world?

Vinge: Yes, empowering the group mind, so to speak. It's one of the paths to the superintelligence. Just as we're getting the digital Gaia, in a different arena we're using the Internet to empower humanity. The ensemble of educated, connected humanity is much more powerful than all of the experts any national government has ever employed. For the most part, humanity is good–hearted, but in the past that fact has often been trumped by fear and ignorance and poor communication. Now, I could imagine a new populism that is smarter than anything that went before and is so broadly based that it doesn't suffer from the short–sightedness of populisms past. This trend fits very well with the possibility of soft takeoffs. It's also a plausible counter to the way technology empowers tyrannies. This doesn't mean the end of governments or the end of nations, but given time, the empowerment of the people could be something very smart and wise that has a chance of managing the probably still larger changes that are coming as machines and the environment as a whole wake up.

FN: So you think that could actually overcome some of the potential issues of privacy and government and corporate abuse?

Vinge: Yes. A good example is David Brin's nonfiction book The Transparent Society, which takes the problem of government intrusion and personal privacy and just kind of turns it on its head – as a result getting both freedom and information. Are you familiar with that book?

FN: You know, I haven't read it, but that definitely makes me want to pick it up. My first reaction is that it sounds like the transition period could be very tumultuous.

Vinge: Oh, yes! I'm more privacy oriented than most people and the prospect chills my blood: the transition years would be awful. But my guess is that after that transition, in an era where passive surveillance was generally available to individuals and the data often pooled for integrated analysis, that the world would be pretty much like it is now – except that there would be fewer secret villains and the most hypocritical laws would have gone away.

FN: I think Fairmont High's school motto, "Working hard not to become obsolete", is both fitting and hilarious. (Vinge laughs.) It's not exactly subtle, but it very succinctly states the situation we're all going to

be dealing with in years to come. Or at least many of us. As more and more careers become obsolete, do you have any thoughts about how we might or should deal with the social and economic repercussions?

Vinge: Alas, no solutions. But even if a person is not obsessive about the Singularity, viewing it as one scenario to fit events against is interesting. When unemployment gets massive enough, you no longer call it unemployment, you call it a biological sea–change. There was a time in the sixties and seventies and eighties when academics and white–collar workers could say with comfortable objectivity, "Isn't it a shame what this technology is doing to blue–collar employment? But we intellectuals are safe and we'll create social programs to cushion the stresses for the less fortunate." Then in the eighties and nineties tech stresses started biting into white–collar workers. And nowadays, there are very few jobs untouched, except for the most talented people – of whom only the most naive may still feel safe. In the writing business, we're under an awful lot of stress just because so many good writers are writing. Of course for us writers this is not directly a Singularity issue, but in almost every department of human affairs, productivity and quality increases are great news for consumers but these very same consumers are running a Red Queen's race in their own particular production domains. It's the two sides of the same coin. One side is the best it's ever been for humans and the other is "Oh my god, how am I going to compete?"

Predicting beyond the Singularity is intrinsically more difficult than the prognostication of futurologies past. For instance, Bill Joy may be right to say that the future doesn't need us. On the other hand, it's entirely possible that not only does the future need us, it needs us a lot. In nature, one recurring theme is that while extinctions are ubiquitous, earlier forms and solutions persist in support of grander schemes. Also, and this is a point I might have more trouble documenting, over and over again you get a situation where relatively inefficient ways of doing things are appropriate because the environment can't support anything more. For instance there is the rise of aerobic metabolisms. There is the move of life onto land. In cases like these, the earlier, so–called primitive life forms make a new environment that can support something more advanced. If all the bacteria were to die tonight, humanity – as well as most plant and animal life – would die very soon. To me, this is an example of life depending on earlier life. We could look at our use of new substrates in a similar way: the silicon, the germanium, gallium–arsenide, fiber optics, and whatever there's going to be in the future.

FN: They're in support of whatever is to come.

Vinge: These are structures that could not have emerged in the pre–human ecology. Nor could they directly redevelop after a technological fall. Science fiction in the 1950s and 1960s had the notion that there could be silicon–based life forms – not computers, but biological life forms – somewhere in the universe. But carbon, oxygen, hydrogen and friends are the only way we've ever seen to do this. It's likely that what arises naturally, at least in the environments that we have access to, isn't much more broadly based than that. On the other hand, now we have a suspicion that when you start making computational devices, there can be much more effective alternative substrates for thought. Such critters just can't do it themselves to start with. So in this imagining, machine life forms are really just another step in the evolution of life.

You're familiar with the term "intelligence amplification", using the computers to enhance human intelligence?

FN: Sure.

Vinge: IA is a plausible – and for many people including me, an attractive – alternative to the alienness of Digital Gaia or network mind or standalone AI. And yet, one could regard the whole human adventure with artifacts as an enhancement of life. It really is just biology beyond our biology. It's biology that will probably depend on that earlier biology.

Charles Stross has just come out with a book called Saturn's Children.

FN: Good, I haven't read that one yet.

Vinge: It's a great book. At one point he has a robot, the protagonist, talking about humans – who are no longer around, by the way. Neither are eukaryotic life forms! The protagonist remarks of classical biology that it seems like a great waste to have built into every organism the entire factory system for building that organism. Actually, that's not precisely true, since we depend on infrastructure in the bacterial world and the plant world and stuff like that. But this character has a point, because in her world, the different components that go into making people are from all over the world and the solar system. It's essentially economics that brings them together in the right place at the right time in order to make new people. These machines have a style of reproduction that is very different from life before, and it's very efficient. But it couldn't emerge spontaneously. I figure that even wildly alien futures would need backups!

FN: Over eons of evolution, humans have learned to anticipate future events better than any other species. As a corollary of this, we tend to view significant change with suspicion, even fear. Obviously, this has provided us with considerable advantage, but as technological and social change accelerate, how do you think we can balance our inherent caution with the need to adapt quickly?

Vinge: That's a core survival question. Getting the proper balance is one of the most important things we must get right. I don't have the answer, but I think the new populism discussed before gives us the best hope: Humans communicating and collaborating, empowered by technology, is the single most powerful tool for surviving the problems civilization faces – and at the same time putting in place the steadily improving smarts needed to navigate a happy outcome for Singularity trends. Fortunately, there are clear trends in this direction. The positive effect of even basic cell–phone access in the third world (trumping much of all prior foreign aid) is a spectacular example. Wikipedia, which has rightly been debated and criticized, has shown us some of the power of network mediated cooperation. We should recognize the Internet as a platform for prototyping the tools for empowerment on the future Internet.

FN: What current or near–term technology really excites you today?

Vinge: Networked embedded microprocessors excite me, both happily and nervously. I wish there was more effort spent on failure mode planning. We're getting to the point where we could have wide–area failures of Almost Everything if there was a failure in the embedded micros. We know there are wide–area threats to embedded systems, where they all can be brought down in a few seconds. In the modern world, we have lots of things to worry about, but this one should rate higher priority.

The promise of space still excites me. The failure of the space program – and by that, I mean cheap access to space – should be a case study for anybody making extravagantly optimistic projections about tech progress. Getting self–sufficient human civilization into space is an important survival issue – of critical importance if the Singularity doesn't happen, and important insurance in any case (see http://www–rohan.sdsu.edu/faculty/vinge/longnow). I think the majority of people realize that this is very, very important and desirable. That fact has made it possible to get large sums of taxpayer money for space development. But now after more than fifty years, our space programs are still based on an infrastructure that costs something like five thousand dollars a pound to low earth orbit. Some space traffic makes good sense at five thousand dollars a pound, and I'm not objecting to that. But people who talk about major manned space programs on such a price basis are selling a doubly gold–plated product. It's gold–plated first because it costs so much money to put cargo in space. And second, every device you put in space has to be enormously – and expensively – reliable since recovery and/or repair are normally impractical. Extending civilization beyond Earth is important to the long–term survival of life – of all forms! Cheap access to space, far below five thousand dollars a pound, is essential to that extension. Spaceflight is something I grew up with and still avidly support, but at this point any major humans–to–space initiative should have much cheaper launch technology as a prerequisite.