Seeking the Sputnik of AI

**Hugo de Garis Interviews Ben Goertzel
on AGI, OpenCog, and the Future of Intelligence**

A couple months after I (Ben Goertzel) interviewed my good friend and sometime research collaborator Hugo de Garis on some of his wilder theoretical ideas [http://hplusmagazine.com/2011/01/18/is-god-an-alien-mathematician/], he suggested it would be interesting to play a role-reversal game and ask ME some interview questions – about my AGI research and my views on the future of humanity and intelligence.   His questions were good ones and so I happily obliged!

Hugo:
About 5 years ago, I was staying at a mutual friend's apartment in Washington DC, just before moving full time to China. At the time you took the view that it would NOT be necessary to have a full knowledge of the human brain to be able to create a human level artificial intelligence. You thought it could be done years earlier using a more humanly engineered approach rather than a "reverse engineering the brain" approach. What are your thoughts on that attitude now, 5 years down the road?

Ben:
Wow, was that really 5 years ago? Egads, time flies!!

But my view remains the same….

Neuroscience has advanced impressively since then, but more in terms of its understanding of the details than in its holistic vision of the brain. We still don't know exactly how neurons work, we still don't know how concepts are represented in the brain nor how reasoning works, etc. We still can't image the brain with simultaneously high spatial and temporal precision. Etc.

Artificial General Intelligence hasn't advanced as visibly as neuroscience since then, but I think it has advanced. The pursuit of AGI now exists as a well-defined field of research, which wasn't the case back then. And many advances have been made in specific areas of importance for AGI – deep learning models of perception, probabilistic logical inference, automated program learning, scalable graph knowledge stores, and so forth. We also have a vibrant open-source AGI project, OpenCog, which I hope will take off in the next few years the same way Linux did a while back.

Both approaches have a significant ways to go before yielding human-level AGI, but I'd say both have the same basic strengths and weaknesses they did 5 years ago, having advanced steadily but not dramatically.

Hugo:
So, which approach do you feel will build human level AI first, your symbolic engineered approach, or reverse engineering of the brain? Why?

Ben:
I wouldn't characterize my approach as "symbolic", I think that's a bit of a loaded and misleading term given the history of AI. My approach involves a system that learns from experience. It does include some probabilistic logic rules that are fairly described as "symbolic", but it also includes dynamics very similar to attractor neural nets, and we're now integrating a deep learning hierarchical perception system, etc.   It's an integrative experiential learning based approach, not a typical symbolic approach.

Anyway, quibbles over terminology aside, do I think an integrative computer science approach or a brain simulation approach will get there faster?

I think that an integrative computer science approach will get there faster UNLESS this approach is starved of funding and attention, while the brain simulation approach gets a lot of money and effort.

I think we basically know how to get there via the integrative comp sci approach NOW, whereas to follow the neuroscience approach, we'd first need to understand an awful lot more about the brain than we can do with current brain measurement technology. But still, even if one of the current AGI projects – like the OpenCog project I cofounded – is truly workable, it will take dozens of man-years of effort to get to human-level AGI by one of these routes. That's not much in the historical time-scale, but it's a nontrivial amount of human effort to pull together without serious backing from government or corporate sources.   Right now OpenCog is funded by a ragtag variety of different approaches, supplemented by the wonderful efforts of some unpaid volunteers – but if this situation continues (for OpenCog and other integrative CS based AGI projects), progress won't be all that fast, and it's not clear which approach will get there first.

What I'm hoping is that, once OpenCog or some other project makes a sufficiently impressive AGI demonstration, there will be a kind of "Sputnik moment" for AGI, and the world will suddenly wake up and see that powerful AGI is a real possibility. And then the excitement and the funding will pour in, and we'll see a massive acceleration of progress. If this AGI Sputnik moment happened in 2012 or 2013 or 2014, for example, then the integrative CS approach would leave the brain simulation approach in the dust – because by that time, we almost surely still won't be able to measure the brain with simultaneously high spatial and temporal precision, so we still won't be able to form an accurate and detailed understanding of how human thinking works.

Hugo:
As machines become increasingly intelligent, how do you see human politics unfolding? What are your most probable scenarios? Which do you feel is *the* most probable?

Ben:

I see human organizations like corporations and governments becoming gradually more and more dependent on machine intelligence, so that they no longer remember how they existed without it.

I see AI and allied technologies as leading to a lot of awfully wonderful things.

A gradual decrease in scarcity, meaning an end to poverty.

The curing of diseases, including the diseases comprising aging, leading ultimately to radical life extension.

Increased globalization, and eventually a world state in some form (maybe something vaguely like the European Union extended over the whole planet, and then beyond the planet).

The emergence of a sort of "global brain", a distributed emergent intelligence fusing AIs and people and the Net into a new form of mind never before seen on Earth.

Increased openness and transparency, which will make government and business run a lot more smoothly. And will also trigger big changes in individual and collective human psychology. David Brin's writings on sousveillance are quite relevant here, by the way, e.g. the Transparent Society. Also you can look at Wikileaks and the current Mideast revolutions as related to this.

But exactly how all this will play out is hard to say right now, because so much depends on the relative timings of various events. There will be advances in "artificial experts", AI systems that lack humanlike autonomy and human-level general intelligence, but still help solve very important and difficult problems. And then there will be advances in true, autonomous, self-understanding AGI. Depending on which of these advances faster, we'll see different sorts of future scenarios unfold.

If we get super-powerful AGI first, then if all goes well the AGI will be able to solve a lot of social problems in one fell swoop. If we get a lot of artificial experts first, then we'll see problems gradually get solved and society gradually reorganized, and then finally a true AGI will come into this reorganized society.

Hugo:
In a recent email to me you said "I don't think it's productive to cast the issue as *species dominance*". Why do you feel that?

Ben:
A species dominance war – a battle between humans and AI machines – is one way that the mid-term future could pan out, but we have no reason to think it's the most likely way. And it's possible that focusing on this sort of outcome too much (as many of our science fiction movies have, just because it makes good theater) may even increase the

odds of it happening. Sometimes life follows fiction, because the movies someone sees and the books they read help shape their mind.

I find Ray Kurzweil a bit overoptimistic in his view on the future, but maybe his overoptimism is performing a valuable service: by placing the optimistic vision of a "kinder, gentler Singularity" in peoples' minds, maybe he'll help that kind of future to come about. I'd imagine he has thought about it this way, alongside other perspectives….

Another possibility, for example, is that humans may gradually fuse with machines, and let the machine component gradually get more and more intelligent, so that first we have cyborgs with a fairly equal mix of human and machine, and then gradually the machine takes over and becomes the dominant portion. In this case we could feel ourselves become superhuman god-minds, rather than having a (losing) war with superhuman god-minds that are external to ourselves. There would be no species dominance debate, but rather a continuous transition from one "species" into another. And quite possibly the superhuman cyborgs and god-mind AIs would allow legacy humans to continue to exist alongside themselves, just as we allow ants to keep crawling around in the national park, and bacteria to course around inside those ants.

Of course, you could point out that some human beings and some political organizations would be made very mad by the preceding few paragraphs, and would argue to wipe out all the nasty risky techno-geeks who entertain crazy ideas like gradually becoming superhuman god-mind cyborg AIs. So, could there be conflicts between people who like this sort of wild ambitious futurist vision, and those who think it's too dangerous to play with? Of course there could. But focusing on the potential consequences of such conflict seems pointless to me, because they're so unknown at this point, and there are so many other possibilities as well. Maybe this sort of conflict of opinion will someday, somewhere, unfold into a violent conflict or maybe it won't. Maybe Ray Kurzweil is right that the advocates of gradual cyborgization will have vastly more advanced capabilities of defense, offense and organization than their opponents, so that the practical possibility of a really violent conflict between the Cosmists and the Terrans (to use your terminology) won't be there.

After all, right now there is a conflict between people who want to roll back to medieval technology and attitudes (Al Qaeda) and modern technological society – and who's winning? They knocked down the World Trade Center, probably aided in many ways by their connections with the Saudis, who are wealthy because of selling oil to technological nations, and are shielded somewhat by their close connections with the US power elite (e.g. the Bush family). But they're coming nowhere close to winning their war on technological progress and cultural modernization. Our weapons are better – and our memes are stickier. When their kids find out about modern culture and technology, a lot of them are co-opted to our side. When our kids find out about the more violent and anti-technology strains of fundamentalist Islam, relatively few are tempted. My guess is this sort of pattern will continue.

Hugo:
Are you mystified by the nature of consciousness?

Ben:
Not at all. Consciousness is the basic ground of the universe. It's everywhere and everywhen (and beyond time and space, in fact). It manifests differently in different sorts of systems, so human consciousness is different from rock consciousness or dog consciousness, and AI consciousness will be yet different. A human-like AI will have consciousness somewhat similar to that of a human being, whereas a radically superhumanly intelligent AI will surely have a very different sort of conscious experience.

To me, experience comes first, science and engineering second. How do I know about atoms, molecules, AI and computers, and Hugo de Garis, and the English language? I know because these are certain patterns of arrangement of my experience, because these are certain patterns that have arisen as explanations of some of my observations, and so forth. The experiential observations and feelings come first, and then the idea and model of the physical world comes after that, built out of observations and feelings. So the idea that there's this objective world out there independent of experience, and we need to be puzzled about how experience fits into it, seems rather absurd to me. Experience is where it all starts out, and everything else is just patterns of arrangement of experience (these patterns of course being part of experience too)….

You could call this Buddhistic or panpsychistic or whatever, but to me it's just the most basic sort of common sense.

So, while I recognize their entertainment value, and their possible value in terms of providing the mind's muscles a cognitive workout -- I basically see all the academic and philosophical arguments about consciousness as irrelevancies. The fact that consciousness is a conundrum within some common construals of the modern scientific world view, tells us very little about consciousness, and a lot about the inadequacies of this world view…

Hugo:
Do you think humanity will be able to create conscious machines?

Ben:
Absolutely, yes.

Hugo:
If someone holds a gun to your head and forces you to choose between a god like artilect coming into existence but humanity gets destroyed as a result, OR the artilect is never created, and hence humanity survives, which would you choose and why? Remember the gun at your head.

Ben:

Well, I guess none of us knows what we'd really do in that sort of situation until we're in it. Like in the book "Sophie's Choice."   But my gut reaction is: I'd choose humanity. As I type these words, the youngest of my three kids, my 13 year old daughter Scheherazade, is sitting a few feet away from me doing her geometry homework and listening to Scriabin Op. Fantasy 28 on her new MacBook Air that my parents got her for Hanukah. I'm not going to will her death to create a superhuman artilect. Gut feeling: I'd probably sacrifice myself to create a superhuman artilect, but not my kids….   I do have huge ambitions and interests going way beyond the human race – but I'm still a human.

How about you? What do you reckon you'd choose?

Hugo:
I vacillate. When I look at the happy people in the park, I feel Terran. When I stare at astronomy books where each little dot is a galaxy in the famous Hubble "Deep Field" photo, I feel Cosmist. But if I REALLY had to choose, I think I would choose Cosmist. I think it would be a cosmic tragedy to freeze evolution at our puny human level. This is the biggest and toughest decision humanity will ever have to make. "Do we build gods, or do we build our potential exterminators?"

Ben:
Well, let's hope we don't have to make that choice. I see no reason why it's impossible to create vastly superhuman minds – and even merge with them – while still leaving a corner of the cosmos for legacy humans to continue to exist in all their flawed monkey-like beauty! …

Hugo:
How do you see humanity's next 100 years?

Ben:
I guess I largely answered this already, right? I see the creation of superhuman AGI during this century as highly likely. Following that, I see a massive and probably irreducible uncertainty.   But I think there's a reasonably high chance that what will happen is:

… some superhuman AGIs, seeded by our creations, will leave our boring little corner of the universe

… some humans will gradually cyborgify themselves into superhuman AGI god-minds, and probably bid this corner of the Cosmos adieu as well

… some humans will opt to stay legacy humans, and others will opt to be cyborgs of various forms, with various combinations of human and engineered traits

… the legacy humans and "weak cyborgs" will find their activities regulated by some sort of mildly superhuman "Nanny AI" that prevents too much havoc or destruction from happening

That's my best guess, and I think it would be a pretty nice outcome. But I freely admit I have no strong scientific basis for asserting this is the most probable outcome. There's a hell of a lot of uncertainty about.

Hugo:
Do you think friendly AI is possible? Can you justify your answer.

Ben:
Do I think it's possible to create AGI systems with vastly superhuman intelligence, that are kind and beneficial to human beings? Absolutely, yes.

Do I think it's possible for humans to create vastly superhuman AGI systems that are somehow provably, guarantee-ably going to be kind and beneficial to human beings? Absolutely not.

It's going to be a matter of biasing the odds.

And the better an AGI theory we have, the more intelligently we'll be able to bias the odds. But I doubt we'll be able to get a good AGI theory via pure armchair theorizing. I think we'll get there via an evolving combination of theory and experiment – experiment meaning, building and interacting with early-stage proto-AGI systems of various sorts.

Hugo:
Do you see the US or China being the dominant AI researcher nation in the coming decades?

Ben:
Hmmm, I think I'll have to answer that question from two perspectives: a general one, setting aside any considerations related to my own AGI work in particular; and a personal one, in terms of the outlook for my own AGI project.

Generally speaking, my view is that the US has a humongous lead over anywhere else in terms of AGI research. It's the only country with a moderate-sized community of serious researchers who are building serious, practical AGI architectures aimed at the grand goal of human-level intelligence (and beyond). Second place is Europe, not China or India, not even Korea or Japan…. The AGI conference series that I co-founded operates every alternate year in the US, and every alternate year elsewhere. The AAAI, the strongest narrow-AI professional organization in the world, is international in scope but US-founded and to a significant extent still US-focused.

The US also has by far the world's best framework for technology transfer – for taking technology out of the lab and into the real world. That's important, because once AGI development reaches a certain point, tech transfer will allow its further development to be funded by the business sector, which has a lot of money. And this kind of thing is hard for other countries to replicate, because it involves a complex ecosystem of interactions between companies of various sizes, universities, and investors of various sorts. It's even hard for cities in the US, outside a certain number of tech hubs, to pull off effectively.

Also, most probably the first powerful AGIs will require a massive server farm, and who's best at doing that? US companies like Google and Amazon and IBM, right? China may have built the world's fastest supercomputer recently, but that's sort of an irrelevancy, because the world doesn't really need supercomputers anymore – what we really need are massive distributed server farms like the ones operated with such stunningly low cost and high efficiency by America's huge Internet companies.

And culturally, the US has more of a culture of innovation and creativity than anywhere else. I know you lived for a while in Utah, which has its pluses but is a very unusual corner of the US – but if you go to any of the major cities or tech hubs, or even a lot of out-of-the-way college towns, you'll see a spirit of enthusiastic new-idea-generation among young adults that is just unmatched anywhere else on the planet. Also a spirit of teamwork, that leads a group of friends just out of college to start a software company together, cooperating informally outside the scope of any institution or bureaucracy.

Look at any list of the most exciting tech companies or the biggest scientific breakthroughs in the last few years, and while it will look plenty international, you'll see a lot of US there. Many of the US scientists and technologists will have non-Anglo-Saxon-sounding names – including many that are Chinese or Indian -- but that's part of the US's power. Many of the best students and scientists from around the world come to America to study, or teach, or do research, or start companies, etc. That's how the US rose to science and engineering prominence in the first place – not through descendants of the Puritans, but through much more recent immigrants. My great-grandparents were Eastern European Jews who immigrated to the US in the first couple decades of the last century. They were farmers and shopkeepers in Europe, now their descendants are scientists and professors, executives and teachers, etc. This is same sort of story that's now bringing so many brilliant Asians to America to push science and technology forward.

So, hey, God bless America! What more can I say….?

Not many people know that I live near Washington DC – a lot of people assume I'm from California for some reason. I've lived a lot of places (Brazil, Oregon, New Jersey, Philly, four of the five boroughs of New York City, Australia, New Zealand, New Mexico) but never California…. Not yet, at any rate. Though (my companies) Novamente and Biomind have had plenty of customers there, and I've become painfully accustomed to the red-eye flights from DC to San Fran and LA. As you know I live in Maryland just north of DC, a few miles from the National Institutes of Health, for which I've done a lot

of bioinformatics work; and I've also done some AI consulting for various companies working with other government agencies. I've become a bit of a "Beltway bandit" since I moved here in 2003. DC has its pluses and minuses, and I wouldn't say I fit into the culture here too naturally; but there's a lot more interesting R&D going on here than most people realize, because the culture here isn't publicity-oriented. And in some ways there's a longer-term focus here than one finds in Silicon Valley, where there's so much obsession with moving super-fast and getting profits or cash flow or eyeballs or whatever as quickly as possible…. The Chinese government thinks 30 years ahead (one of its major advantages compared to the US, I might add), Wall Street thinks a quarter ahead, Silicon Valley thinks maybe 3 years ahead (Bay area VCs typically only want to invest in startups that have some kind of exit strategy within 3 years or so; and they usually push you pretty hard to launch your product within 6 months of funding – a default mode of operation which is an awkward fit for a project like building AGI), and DC is somewhere between Silicon Valley and China….

But still … having said all that … there's always another side to the coin, right? On the other hand, if – IF IF IF – the US manages to squander these huge advantages during the next few decades, via pushing all its funding and focus on other stuff besides AGI and closely allied technologies … then who knows what will happen. Economically, China and India are gradually catching up to the US and Europe and Korea and Japan … they're gradually urbanizing and educating and modernizing their huge rural populations. And eventually China will probably adopt some sort of Western style democracy, with free press and all that good stuff, and that will probably help Chinese culture to move further in the direction of free expression and informal team work and encouragement of individual creativity – things that I think are extremely important for fostering progress in frontier areas like AGI. And eventually India will overcome its patterns of corruption and confusion and become a First World country as well. And when these advances happen in Asia, then maybe we'll see a more balanced pattern of emigration, where as many smart students move from the US to Asia as vice versa. If the advent of AGI is delayed till that point – we're talking maybe 2040 or so I would reckon – then maybe China or India is where the great breakthrough will happen.

I do think China is probably going to advance beyond the US in several areas in the next couple decades. They're far, far better at cheaply making massive infrastructure improvements than we are. And they're putting way more effort and brilliance into energy innovations than we are. To name just two examples. And then there's stem cell research, where the US still has more sophistication, but China has fewer regulatory slowdowns; and other areas of biomedical research where they excel. But these areas are largely to do with building big stuff or doing a lot of experimentation. I think the Chinese can move ahead in this sort of area more easily than in something like AGI research. I think AGI research depends mostly on the closely coordinated activity of small informal or semi-formal groups of people pursuing oddball ideas, and I don't think this is what Chinese culture and institutions are currently best at fostering.

Another factor acting against the USA, is that the US AI research community (along with its research funding agencies) is largely mired in some unproductive ideas, the result of

the long legacy of US AI research.   And it's true that the Chinese research community and research funders aren't similarly conceptually constricted – they have fewer unproductive conceptual biases than US AI researchers, on the whole. But if you look at the details, what most Chinese academics seem to care most about these days is publishing papers in SCI-indexed journals and getting their citation counts higher – and the way to do this is definitely NOT to pursue long-term oddball speculative AGI research….

You might be able to frame an interesting argument in favor of India as a future AGI research center, on this basis. They seem a bit less obsessed with citation counts than the Chinese, and they have a long history of creative thinking about mind and consciousness, even longer than the Chinese! Modern consciousness studies could learn a lot from some of the medieval Indian Buddhist logicians. Plus a lot of Silicon Valley's hi-tech expertise is getting outsourced to Bangalore. And the IITs are more analogous to top-flight US technical universities than anything in China – though Chinese universities also have their strengths. But anyway, this is just wild speculation, right? For now there's no doubt that the practical nexus of AGI research remains in America (in spite of lots of great work being done in Germany and other places). AGI leadership is America's to lose … and it may well lose it, time will tell…. Or America-based AGI research may advance sufficiently fast that nobody else has time to catch up….

Hugo:
OK, that was your general answer … now what about your personal answer? I know you've been spending a lot of time in China lately, and you're working with students at Xiamen University in the lab I ran there before I retired, as well as with a team in Hong Kong….

Ben:
Yeah, that was my general answer. Now I'll give my personal answer – that is, my answer based on my faith in my own AGI project.

I think that the OpenCog project, which I co-founded, is on an R&D path that has a fairly high probability of leading to human-level general intelligence (and then beyond).   The basic ideas are already laid out in some fairly careful (and voluminous) writing, and we have a codebase that already functions and implements some core parts of the design, and a great team of brilliant AGI enthusiasts who understand the vision and the details…. So, if my faith in OpenCog is correct, then the "US versus China" question becomes partly a question of whether OpenCog gets developed in the US or China.

Interestingly, it seems the answer is probably going to be: both! … and other places too. It's an open source project with contributors from all over the place.

My company Novamente LLC is driving part (though by no means all) of OpenCog development, and we have some programmers in the US contributing to OpenCog based on US government contracts (which are for narrow-AI projects that use OpenCog, rather than for AGI per se), as well as a key AGI researcher in Bulgaria, and some great AI

programmers in Belo Horizonte, Brazil, whom I've been working with since 1998. There's also a project at Hong Kong Polytechnic University, co-sponsored by the Hong Kong government's Innovation in Technology Fund and Novamente LLC, which is applying OpenCog to create intelligent game characters. And there's a handful of students at Xiamen University in China working on making a computer vision front end for OpenCog, based on Itamar Arel's DeSTIN system (note that Itamar is from Israel, but currently working in the US, as a prof at the University of Tennessee Knoxville, as well as CTO of a Silicon Valley software company, Binatix). Now, the AI programmers on the Hong Kong project consist of two guys from New Zealand (including Dr. Joel Pitt, the technical lead on the project) and also three exchange students from Xiamen University. In April I'll be spending a few weeks in Hong Kong with the team there, along with Dr. Joscha Bach from Germany.

My point in recounting all those boring details about people and places is – maybe your question is just too 20th century. Maybe AGI won't be developed in any particular place, but rather on the interwebs, making use of the strengths of the US as well as the strengths of China, Europe, Brazil, New Zealand and so on and so forth.

Or maybe the US or Chinese government will decide OpenCog is the golden path to AGI and throw massive funding at us, and we'll end up relocating the team in one location – it's certainly possible. We're open to all offers that will allow us to keep our code open source!

So far I have found the Chinese research funding establishment, and the Chinese university system, to be much more open to radical new approaches to AGI research than their American analogues. In part this is just because they have a lot less experience with AI in general (whether narrow AI or AGI). They don't have any preconceived notions about what might work, and they don't have such an elaborate "AI brain trust" of respected older professors at famous universities with strong opinions about which AI approaches are worthwhile and which are not. I've gotten to know the leaders of the Chinese AI research community, and they're much much more receptive to radical AGI thinking than their American analogues. Zhongzhi Shi from the Chinese Academy of Sciences is going to come speak about Chinese AGI efforts at the AGI-11 conference in California in August – and I've also had some great conversations with your friend Yixin Zhang, who's the head of the Chinese AI Association. I went to their conference last year in Beijing, and as you'll recall our joint paper on our work with intelligent robots in Xiamen won the Best Paper prize! At the moment their efforts are reasonably well funded, but not to the level of Chinese work on semiconductors or supercomputers or wind power or stem cell research, etc. etc. But certainly I can see a possible future where some higher-ups in the Chinese government decide to put a massive amount of money into intelligent robotics or some other AGI application, enough to tempt a critical mass of Western AGI researchers as well as attract a lot of top Chinese students…. If that does happen, we could well see the world's "AGI Sputnik" occur in China. And if this happens, it will be interesting to see how the US government responds – will it choose to fund AGI research in a more innovation-friendly way than it's done in the past,

or will it respond by more and more aggressively funding the same handful of universities and research paradigms it's been funding since the 1970s?

So overall, putting my general and personal answers together – I feel like in the broad scope, the AGI R&D community is much stronger in the US than anywhere else, and definitely much much stronger than in China. On the other hand, AGI is the sort of thing where one small team with the right idea can make the big breakthrough. So it's entirely possible this big breakthrough could occur outside the US, either via natively-grown ideas, or via some other country like China offering a favorable home to some American-originated AGI project like OpenCog that's too radical in its conceptual foundations to fully win the heart of the US AI research funding establishment.

But ultimately I see the development of AGI in an international context as providing higher odds of a beneficial outcome, than if it's exclusively owned and developed in any one nation. So as well as being an effective way to get work done, I think the international open-source modality we're using for OpenCog is ultimately the most ethically beneficial way to do AGI development….

Well, what do you think? You live in China … I've spent a lot of time there in recent years (and plan to spend a few months there this year), but not as much as you. And you speak the language; I don't. Do you think I'm missing any significant factors in my analysis?

Hugo:
I put more emphasis on Chinese economics and national energy. Americans have become fat and complacent, and are not growing economically at anywhere near the Chinese rate. The historical average US economic growth rate is 3%, whereas China's is 10% (and has been sustained pretty much for 30 years). Doing the math, if this incredible energy of the Chinese can be sustained for a few more decades, it will put the rich eastern Chinese cities at a living standard well above that of the US, in which case they can afford to attract the best and most creative human brains in the world to come to China. The US will then see a reverse brain drain, as its best talent moves to "where its at", namely China. With a million talented Westerners in China within a decade, they will bring their "top world" minds with them and shake up China profoundly, modernizing it, legalizing it, democratizing it and civilizing it. Once China finally goes democratic and with its rich salaries, it doesn't matter whether the Chinese can be creative or not. The presence of the best non Chinese brains in China will ensure an explosion of creativity in that part of the world.

Ben:
Hmmm…. Chinese economic growth is indeed impressive – but of course, it's easier to grow when you're urbanizing and modernizing a huge rural population. To an extent, the US and Europe and Japan are growing more slowly simply because they've already urbanized and modernized. I guess once China and India have finished modernizing their growth rates may look like those in the rest of the world, right? So your projection that Chinese growth will make Chinese cities richer than US cities may be off-base, because

most of Chinese growth has to do with bringing more and more poor people up to the level of the international middle class. But I guess that's a side point, really….

About creativity … actually, I know many fantastically creative Chinese people (including some working on OpenCog!) and I guess you do too – what seems more lacking in China is a mature ecosystem for turning wacky creative ideas into novel, functional, practical realizations. I'm sure that will come to China eventually, but it requires more than just importing foreigners, it may require some cultural shifts as well – and it's hard to estimate the pace at which those may happen. But China does have the capability to "turn on a dime" when it wants to, so who knows!

About Americans being fat and complacent – hmmm, well I'm a little heavier than I was 20 years ago, but I haven't become a big fat capitalist pig yet … and I don't consider myself all that complacent! Generalizations are dangerous, I guess. Silicon Valley, New York, DC, Boston, Seattle, LA – there's a lot of energy in a lot of US cities; a lot of diversity and a lot of striving and energy. But yeah, I see what you mean – America does sort of take for granted that it's on top, whereas China has more of an edge these days, as if people are pushing extra hard because they know they're coming from behind….

Look at the San Fran Bay area as an example. Sometimes the Silicon Valley tech scene seems a bit tired lately, churning out one cookie-cutter Web 2.0 startup after another. But then the Shanghai startup scene is largely founded on churning out Chinese imitations of Silicon Valley companies. And then you have some really innovative stuff going on in San Fran alongside the Web 2.0 copycat companies, like Halcyon Molecular (aiming at super-cheap DNA sequencing) or Binatix or Vicarious Systems (deep learning based perception processing, aiming toward general intelligence). You don't have a lot of startups in China with that level of "mad science" going on in them, at least not at this point. So there's a lot of complexity in both the US and China, and it's far from clear how it will all pan out. Which is one reason I'm happy OpenCog isn't tied to any one city or country, of course….

Hugo:
Can you list the dominant few ideas in your new book "Building Better Minds".

Ben:
Uh oh, a hard question! It was more fun blathering about politics….

That book – which is almost done now, but still needs some editing and fine-tuning – is sort of a large and unruly beast. It's almost 900 pages and divided into two parts. The first part outlines my general approach to the problem of building advanced AGI, and the second part reviews the OpenCog AGI design – not at the software code level, but at the level of algorithms and knowledge representations and data structures and high level software design.

Part I briefly reviews the "patternist" theory of mind I outlined in a series of books earlier in my career, and summarized in *The Hidden Pattern* in 2006. Basically, a mind is a system of patterns that's organized into a configuration that allows it to effectively recognize patterns in itself and its world. It has certain goals and is particularly oriented to recognize patterns of the form "If I carry out this process, in this context, I'm reasonably likely to achieve this goal or subgoal." The various patterns in the mind are internally organized into certain large-scale networks, like a hierarchical network, and an associative hierarchy, and a reflexive self. The problem of AGI design then comes down to: how do you represent the patterns, and via what patterned processes does the pattern system recognize new patterns? This is a pretty high-level philosophical view but it's important to start with the right general perspective or you'll never get anywhere on the AGI problem, no matter how brilliant your technical work nor how big your budget.

Another key conceptual point is that AGI is all about resource limitations. If you don't have limited spacetime resources then you can create a super-powerful AGI using a very short and simple computer program. I pointed this out in my 1993 book The Structure of Intelligence (and others probably saw it much earlier, such as Ray Solomonoff), and Marcus Hutter rigorously proved it in his work on Universal AI a few years ago. So real-world AGI is all about: how do you make a system that displays reasonably general intelligence, biased toward a certain set of goals and environments, and operates within feasible spacetime resources. The AGIs we build don't need to be biased toward the same set of goals and environments that humans are, but there's got to be some overlap or we won't be able to recognize the system as intelligent, given our own biases and limitations.

One concept I spend a fair bit of time on in Part I is cognitive synergy: the idea that a mind, to be intelligent in the human everyday world using feasible computing resources, has got to have multiple somewhat distinct memory stores corresponding to different kinds of knowledge (declarative, procedural, episodic, attentional, intentional (goal-oriented)) … and has got to have somewhat different learning processes corresponding to these different memory stores … and then, these learning processes have got to synergize with each other so as to prevent each other from falling into unproductive, general intelligence killing combinatorial explosions.

In the last couple months, my friend and long-time collaborator (since 1993!) Matt Ikle' and I put some effort into formalizing the notion of cognitive synergy using information geometry and related ideas. This will go into Building Better Minds too – one of my jobs this month is to integrate that material into the manuscript. We take our cue from general relativity theory, and look at each type of memory in the mind as a kind of curved mindspace, and then look at the combination of memory types as a kind of composite curved mindspace. Then we look at cognition as a matter of trying to follow short paths toward goals in mindspace, and model cognitive synergy as cases where there's a shorter path through the composite mindspace than through any of the memory type specific mindspaces. I'm sort of hoping this geometric view can serve as a unifying theoretical framework for practical work on AGI, something it's lacked so far.

Then at the end of Part I, I talk about the practical roadmap to AGI – which I think should start via making AGI children that learn in virtual-world and robotic preschools. Following that we can integrate these toddler AGIs with our narrow-AI programs that do things like biological data analysis and natural language processing, and build proto-AGI artificial experts with a combination of commonsense intuition and specialized capability. If I have my way, the first artificial expert may be an artificial biologist working on the science of life extension, following up the work I'm doing now with narrow AI in biology with Biomind LLC and Genescient Corp. And then finally, we can move from these artificial experts to real human-level AGIs. This developmental approach gets tied in with ideas from developmental psychology, including Piaget plus more modern ideas. And we also talk about developmental ethics – how you teach an AGI to be ethical, and to carry out ethical judgments using a combination of logical reason and empathic intuition. I've always felt that just as an AGI can ultimately be more intelligent than any human, it can also be more ethical – even according to human standards of ethics. Though I have no doubt that advanced AGIs will also advance beyond humans in their concept of what it is to be ethical.

That's Part I, which is the shorter part. Part II then goes over the OpenCog design and some related technical ideas, explaining a concrete path to achieving the broad concepts sketched in Part I. I explain practical ways of representing each of the kinds of knowledge described in Part I – probabilistic logic relations for declarative knowledge, programs in a simple LISP-like language for procedural knowledge, attractor neural net like activation spreading for attentional knowledge, "movies" runnable in a simulation engine for episodic knowledge, and so forth. And then I explain practical algorithms for dealing with each type of knowledge – probabilistic logical inference and concept formation and some other methods for declarative knowledge; probabilistic evolutionary program learning (MOSES) for procedural knowledge economic attention networks for attentional knowledge; hierarchical deep learning (using Itamar Arel's DeSTIN algorithm) for perception; etc. And I explain how all these different algorithms can work together effectively, helping each other out when they get stuck – and finally, how due to the interoperation of these algorithms in the context of controlling an agent embodied in a world, the mind of the agent will build up the right internal structures, like hierarchical and heterarchical and self networks.

I'm saying "I" here because the book represents my overall vision, but actually I have two co-authors on the book – Nil Geisweiller and Cassio Pennachin – and they're being extremely helpful too. I've been working with Cassio on AI since 1998 and he has awesomely uncommon common sense and a deep understanding of both AI, cog sci and software design issues. And Nil also thoroughly understands the AGI design, and is very helpful at double-checking and improving formal mathematics (I understand math very well, that was my PhD area way back when, but I have an unfortunate tendency to make careless mistakes…). The two of them have written some parts and edited many others; and there are also co-authors for many of the chapters, who have contributed significant thinking. So the book is really a group effort, orchestrated by me but produced together with a lot of the great collaborators I've been luck to have in the last decade or so.

Now, so far our practical work with OpenCog hasn't gotten too far through the grand cosmic theory in the Building Better Minds book. We've got a basic software framework that handles multiple memory types and learning processes, and we have initial versions of most of the learning processes in place, and the whole thing is built pretty well in C++ in a manner that's designed to be scalable (the code now has some scalability limitations, but it's designed so we can make it extremely scalable by replacing certain specific software objects, without changing the overall system). But we've done only very limited experimentation so far with synergetic interaction between the different cognitive processes. Right now the most activity on the project is happening in Hong Kong, where there's a team working on applying OpenCog to make a smart video game character. We're going to get some interesting cognitive synergy going in that context, during the next couple years….

The argument some people have made against this approach is that it's too big, complex and messy. My response is always: OK, and where exactly is your evidence that the brain is not big, complex and messy? The OpenCog design is a hell of a lot simpler and more elegant than the human brain appears to be. I know a fair bit of neuroscience, and I've done some consulting projects where I've gotten to interact with some of the world's greatest neuroscientists – and everything I learn about neuroscience tells me that the brain consists of a lot of neuron types, a lot of neurotransmitter types, a lot of complex networks and cell assemblies spanning different brain regions which have different architectures and dynamics and evolved at different times to meet different constraints. The simplicity and elegance that some people demand in an AGI design, seems utterly absent from the human brain. Of course, it's possible that once we find the true and correct theory of the human brain, the startling simplicity will be apparent. But I doubt it. That's not how biology seems to work.

I think we will ultimately have a simple elegant theory of the overall emergent dynamics of intelligent systems. That's what I'm trying to work toward with the ideas on curved mindspace that I mentioned above. Whether or not those exact ideas are right, I'm sure some theory of that general nature is eventually going to happen. But the particulars of achieving intelligence in complex environments using feasible computational resources – I feel that's always likely to be a bit messy and heterogeneous, involving integration of different kinds of memory stores with different kinds of learning processes associated with them. Just like the theory of evolution is rather simple and elegant, and so is the operation of DNA and RNA -- but the particulars of specific biological systems are always kind of complex and involved.

I'm sure we didn't get every detail right in Building Better Minds – but we're gradually pulling together a bigger and bigger community of really smart, passionate people working on building the OpenCog system, largely inspired by the ideas in that book (plus whatever other related ideas team members bring in, based on their own experience and imagination!). The idea is to be practical and use Building Better Minds and other design ideas to create a real system that does stuff like control video game characters and robots and biology data analysis systems, and then improve the details of the design as we go along. And improve our theories as we go along, based on studying the behaviors of our

actual systems. And once we get sufficiently exciting behaviors, trumpet them really loud to the world, and try to create an "AGI Sputnik Moment", after which progress will really accelerate. Singularity, here we come!