

Interview: Dr. Ben Goertzel on Artificial General Intelligence, Transhumanism and Open Source (Part 1/2)

10 June 2011, by Stuart Mason Dambrot



Dr. Ben Goertzel. Photo courtesy: Neural Imprints (<http://www.neuralimprints.com/>)

(PhysOrg.com) -- Dr. Ben Goertzel is Chairman of [Humanity+](#); CEO of AI software company [Novamente LLC](#) and bioinformatics company [Biomind LLC](#); leader of the open-source [OpenCog](#) Artificial General Intelligence (AGI) software project; Chief Technology Officer of biopharma firm [Genescient Corp.](#); Director of Engineering of digital media firm [Vzillion Inc.](#); Advisor to the [Singularity University](#) and [Singularity Institute](#); Research Professor in the Fujian Key Lab for Brain-Like Intelligent Systems at [Xiamen University](#), China; and general Chair of the [Artificial General Intelligence Conference Series](#). His research work encompasses artificial general intelligence, natural language processing, cognitive science, data mining, machine learning, computational finance, bioinformatics, virtual worlds and gaming and other areas, Dr. Goertzel has published a dozen scientific books, 100+ technical papers, and numerous journalistic articles, and the futurist treatise *A Cosmist Manifesto*. Before entering the software industry he served as a university faculty in several departments of mathematics, computer

science and cognitive science, in the US, Australia and New Zealand.

Dr. Goertzel spoke with Critical Thought's Stuart Mason Dambrot following his talk at the recent 2011 [Transhumanism Meets Design Conference](#) in New York City. His presentation, **Designing Minds and Worlds**, asked and answered the key questions, *How can we design a world (virtual or physical) so that it supports ongoing learning and growth and ethical behavior? How can we design a mind so that it takes advantage of the affordances its world offers?* These are fundamental issues that bridge AI, robotics, cyborgics, [virtual world](#) and game design, sociology and psychology and other areas. His talk addressed them from a cognitive systems theory perspective and discussed how they're concretely being confronted in his current work applying the OpenCog Artificial General Intelligence system to control game characters in virtual worlds.

This is the first part of a two-part article. The second part is available at http://www.physorg.com/news/2011-06-dr-ben-goertzel-artificial-intelligence_1.html

SM Dambrot: We're here with Dr. Ben Goertzel, CEO of Novamente, Leader of OpenCog and Chairmen of Humanity+ [at the 2011 Humanity+ Transhumanism Meets Design Conference in New York City]. Thank you so much for your time.

Dr. Goertzel: It's great to be here.

SM Dambrot: In your very interesting talk yesterday, you spoke about the importance of the relationship between minds and worlds. Could you please expound on that a bit in terms of Artificial General Intelligence?

Dr. Goertzel: As an AGI developer this is a very

practical issue which initially presents itself in a mundane form - but many subtle philosophical and conceptual problems are lurking there. From the beginning, when you're building an AGI system you need that system to do something - and most AI history is about building AI systems to solve very particular problems, like planning and scheduling in a military context, or finding documents online in a Google context, playing chess, and so forth. In these cases you're taking a very specific environment - a specific set of stimuli --- and some very specific tasks -- and customizing an AI system to do those tasks in that environment, all of which is quite precisely defined. When you start thinking about AGI - Artificial General Intelligence - in the sense of human-level AI - you not only need to think about a broader level of cognitive processes and structures inside the AI's mind, you need to think about a broader set of tasks and environments for the AI system to deal with.

In the ideal case, one could approach human-level AGI by placing a humanoid robot capable of doing everything a human body can do in the everyday human world, and then the environment is taken care of - but that's not the situation we're confronted with right now. Our current robots are not very competent when compared with the human body. They're better in some ways - such as withstanding extremes of weather that we can't - but by and large they can't move around as freely, they can't grasp things and manipulate objects as well, and so on. Moreover, if you look at the alternatives - such as implementing complex objects and environments in virtual and game worlds - you encounter a lot of limitations as well.

You can also look at types of environments that are very different from the kinds of environments in which humans are embedded. For example, the Internet is a kind of environment that is immense and has many aspects that the everyday pre-Internet human environment doesn't have: billions of text documents, satellite data from weather satellites, millions of webcams; but when you have a world for the AI that's so different from what we humans ordinarily perceive, you start to question whether a AI modeled on human cognitive architecture is really suited for that sort of environment.

Initially the matter of environments and tasks may seem like a trivial issue - it may seem that the real problem is creating the artificial mind, and then when that's done, there's the small problem of making the mind do something in some environment. However, the world - the environment and the set of tasks that the AI will do - is very tightly coupled with what is going on inside the AI system. I therefore think you have to look at both minds and worlds together.

SM Dambrot: What you've just said about minds and worlds reminds me of two things. One is the way living systems evolved - that is, species evolve not in a null context, but rather, as you so well put it, tightly coupled to in this case an environment niche; every creature's sensory apparatus is tuned that niche, so the mind and world co-evolve. The other is what you mentioned yesterday when discussing virtual and game worlds - that physics engines are not being used in all interactive situations - which leads me to ask you what you think will happen once true AGIs are embodied.

Dr. Goertzel: If we want to, we can make the boundary between the virtual and physical worlds pretty thin. Most roboticists work mostly in robot simulators, and a good robot simulator can simulate a great deal of what the robot confronts in the real world. There isn't a good robot simulator for walking out in the field with birds flying overhead, the wind, the rain, and so forth - but if you're talking about what occurs within someone's house a lot can be accomplished.

It's interesting to see what robot simulators can and can't do. If we were trying to simulate the interior of a kitchen, for example, a robot simulator can deal with the physics of chairs and tables, pots and pans, the oven door, and so forth. Current virtual worlds don't do that particularly well because they only use a physics engine for a certain class of interactions, and generally not for agent-object or agent-agent interactions - but these are just conventional simplifications made for the sake of efficiency, and can be overcome fairly straightforwardly if one wants to expend the computational resources on simulating those details of the environment.

If you took the best current robot simulators, most of which are open source, and integrated them with a virtual world, then you could build a very cool massive multiplayer robot simulator. The reason this hasn't happened so far is simply that businesses and research funding agencies aren't interested in this. I've thought a bit about how to motivate work in that regard. One idea is to design a video game that requires physics - for example, a robot wars game in players build robot from spare parts, and the robots do battle. You could also make the robots intelligent and bring some AI into it, which if done correctly would lead to the development of an appropriate cognitive infrastructure.

Having said that, going back to the kitchen - what would current robot simulators not be able to handle, but would have to be newly programmed? Dirt on the kitchen floor, so that in some areas you could slip more than others; baking, so when you mix flour and sugar and put it in the oven, the chemistry which is beyond what any current physics engine can really do; paper burning in the flame of a gas stove; and so on. The open question is how important these bits and pieces of everyday human life are to the development of an intelligence.

There's a lot of richness in the everyday human world that little kids are fascinated by - fire, cooking, little animals - because this is part of the environmental niche that humans adapted to. Even the best robot simulators don't have that much richness, so I think that it's an interesting area to explore. I think we should push simulators as far as we can, create robot simulators with virtual worlds, and so forth - but at the same time I'm interested in proceeding with robotics as well because there's a lot of richness in the real world and we don't yet know how to simulate it.

The other thing you have to be careful of is that most of the work done with robots now completely ignores all this richness -- and I'm as guilty of that as anybody. When we use robots in our lab in China do we let the robots roam free in the lab? Not currently. We made a little fenced-off area, we put some toys in it, and we made sure the lighting is OK because the robots we're using (Aldebaran Nao

robots) cost \$15,000 and they have a tendency to fall down. It's annoying when they break - you have to send them back to France to get repaired.

So, given the realities of current robot technology we tend to keep the robots in a simplified environment both for their protection, and so that their sensation and actuation will work better. They work, they're cool, and they pick up certain objects well - but not most of those in everyday human life. When we fill the robot lab only with objects they can pick up, we're eliminating a lot of the richness and flexibility a small child has.

SM Dambrot: This raises two more questions: Is cultural specificity required for any given AGI, and is it necessary to imbue an AGI with a sense of curiosity?

Dr. Goertzel: Our fascination with fire is an interesting example. You wonder to what extent it's driven by pure curiosity versus our actual evolutionary history with fire - something that's been going on for millions of years. I think our genome is programmed with reactions to many things in our everyday environment which drive curiosity - and fire and cooking are two interesting examples.

Having said that, yes, curiosity is one of the base motivators. We're already using that fact in our OpenCog work. One of the top-level demands, as we call them, of our system is the ability to experience novelty, to discover new things. There are two demands: to discover new things in the world around it and just have the experience of learning new things internally, which can come through external or internal discovery. So we've already programmed things very similar to curiosity as top-level goals of the system. Otherwise you could end up with a boring system that just wanted to get all of its basic needs gratified, and would then just sit there with nothing to do.

SM Dambrot: That's very interesting - especially the internal novelty drive. That seems even more exciting in terms of any type of AGI analogue to human intelligence, because we spend so much time discovery ideas internally.

Dr. Goertzel: Some people more than others - it's cultural to some extent. I think we as Westerners spend more time intellectually introspecting than do people from Eastern cultures. Being from a Jewish background, I grew up in a culture particularly inclined towards intellectual introspection and meta-meta thinking.

On a technical level, what we've done to inculcate the OpenCog system with a drive for internal novelty and internal learning and curiosity is actually very simple: It's based on information theory and is related to work by [Jürgen Schmidhuber](#) and others on the [mathematical formulation of surprise](#). In an information-theoretic sense, OpenCog is always trying to surprise itself.

SM Dambrot: I recall that when Prof. Schmidhuber was discussing Recursive Neural Networks at Singularity Summit '09, he talked about how the system looks for that type of novelty in its bit configurations.

Dr. Goertzel: That's right - and what we do with OpenCog is quite similar to that. These are ideas that I encountered in the 1980s in the domain of music theory, based on Leonard Meyer's [Emotion and Meaning in Music](#). He was analyzing classical music - Bach, Mozart and so forth - and the idea he came up with was that aesthetically good music is all about the surprising fulfillment of expectations, which I thought was an interesting phrase. Now, if something is just surprising it's too random, and some modern music can be like that - modern classical music in particular. If something is just predictable - pop music is often like that, and some classical music seems like that - it's boring. The best music shows you something new yet it still fulfills the theme in a way that you didn't quite expect to be fulfilled - so it's even better than if it just fulfilled the theme.

I think that's an important aesthetic in human psychology, and if you look at the goal system of a system like OpenCog, the system is seeking surprise but it also gets some reward from having its expectations fulfilled. If it can do both of those at once then it's getting many of its demands fulfilled at the same time, so in principle it should be aesthetically satisfied by the same sorts of things

that people are.

This is all at a very vague level, because I don't think that surprise and fulfillment of expectations are the ultimate equation of aesthetics, music theory or anything else. It's an interesting guide, though, and it's interesting to see the same principles seem to hold up for human aesthetics in quite refined domains, and also for guiding the motivations of very simple AI systems in video game type worlds.

SM Dambrot: I've been wondering about materials and the structure of those materials. Do you think it's important or even necessary in any way to have something that is patterned on our neocortical structure - neurons, axons, synapse, propagation - in order to really emulate our cognitive behavior, or not so relevant?

Dr. Goertzel: The first thing I would say is that in my own primary work right now with OpenCog, I'm not trying to emulate human cognition in any detail, so for what I'm trying to do - which is just to make a system that's as smart as a human in vaguely the same sort of ways that humans are, and then ultimately capable of going beyond human intelligence - I'm almost sure that it's not necessary to emulate the cognitive structure of human beings. Now, if you ask a different question - let's say I really want to simulate Ben Goertzel and make a robot Ben Goertzel that really acts, thinks, and hopefully feels like the real Ben Goertzel - to do *that* is a different proposition and it's less clear to me how far down one needs to go, in terms of emulating neural structure and dynamics.

In principle, of course, one could simulate all the molecules and atoms in my brain in some kind of computer, be it a classical or quantum computer - so you wouldn't actually need to get wet and sticky. On the other hand, if you need to go to a really low level of detail, the simulation might be so consumptive of computing power, you might be better off getting wet and sticky with some type of nanobiotech. When you talk about mind uploading, I don't think we know yet how micro or nano we need to get in order to really emulate the mind of a particular person - but I see that as a somewhat separate project from AGI, where we're trying to

emulate human-like human level intelligence that is not an upload of any particular person. Of course if you *could* upload a person, that would be one path to a human-level AGI … it's just that it's not the path I'm pursuing now, not because it's uninteresting but I don't know how to progress directly and rapidly on that right now.

I think I know how to build a human-level thinking machine…I could be wrong, but at least I have a detailed plan, and I think if you follow this plan for, let's say, a decade, you'd get there. In the case of mind uploading, it seems there's a large bottleneck of information capture - we don't currently have the brain scanning methods capable of capturing the structure of an individual human brain with high spatial and temporal accuracy at the same time, and because of that we don't have the data to experiment with. So if I were going to work on mind uploading, I'd start by trying to design better methods of scanning the brain - which is interesting but not what I've chosen to focus on.

SM Dambrot: Regarding uploading, then, how far down do you feel we might have to go? Is imaging a certain level of structure sufficient? Do we have to capture quantum spin states? I ask because [Max More](#) mentioned random quantum tunneling in his talk, suggesting that quantum events may be a factor in cryogenically-preserved neocortical tissue.

Dr. Goertzel: I'm almost certain that going down to the level of neurons, synapses and neurotransmitter concentrations will be enough to make a mind upload. When you look at what we know from neuroscience so far -- such as what sorts of neurons are activated during different sorts of memories, the impact that neurotransmitter levels have on thought, and the whole area of cognitive neuroscience -- I think there's a pretty strong case that neurons and glia and the molecules intervening in interactions between these cells and other things on this level are good enough to emulate thought without having to go down to the level of quarks and gluons, or even (as [Dr. Stuart Hameroff](#) suggests) the level of the microtubular structure of the cell walls of the neuron. I wouldn't say that I know that for certain, but it would be my guess.

From the perspective of cryogenic preservation, you might as well cover all bases and preserve something, so well that even if our current theories of neuroscience and physics turn out to be wrong, you can still revive the person. So from Max More's perspective as CEO of [Alcor](#), I think he's right - you need to preserve as much as you can, so as not to make any assumptions that might prevent you from reviving someone.

SM Dambrot: Like capturing a photograph in RAW image format…

Dr. Goertzel: Yes - you want to save more pixels than you'll ever need just in case. But from the viewpoint of guiding scientific research, I think it's a fair assumption that the levels currently looked at in cognitive neuroscience are good enough.

This is the first part of a two-part article. The second part is available at http://www.physorg.com/news/2011-06-dr-ben-goertzel-artificial-intelligence_1.html

Copyright 2011 PhysOrg.com.

All rights reserved. This material may not be published, broadcast, rewritten or redistributed in whole or part without the express written permission of PhysOrg.com.

APA citation: Interview: Dr. Ben Goertzel on Artificial General Intelligence, Transhumanism and Open Source (Part 1/2) (2011, June 10) retrieved 22 October 2019 from <https://phys.org/news/2011-06-dr-ben-goertzel-artificial-intelligence.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.