The real reasons we don’t have AGI yet

A response to David Deutsch’s recent article on AGI

October 8, 2012 by Ben Goertzel

As we noted in a recent post, physicist David Deutsch said the field of “artificial general intelligence” or AGI has made “no progress whatever during the entire six decades of its existence.” We asked Dr. Ben Goertzel, who introduced the term AGI and founded the AGI conference series, to respond. — Ed.

Like so many others, I’ve been extremely impressed and fascinated by physicist David Deutsch’s work on quantum computation — a field that he helped found and shape.

I also encountered Deutsch’s thinking once in a totally different context — while researching approaches to home schooling my children, I noticed his major role in the Taking Children Seriously movement, which advocates radical unschooling, and generally rates all coercion used against children as immoral.

In short, I have frequently admired Deutsch as a creative, gutsy, rational and intriguing thinker. So when I saw he had written an article entitled “Creative blocks: The very laws of physics imply that artificial intelligence must be possible. What’s holding us up?,” I was eager to read it and get his thoughts on my own main area of specialty, artificial general intelligence.

Oops.

I was curious what Deutsch would have to say about AGI and quantum computing. But he quickly dismisses Penrose and others who think human intelligence relies on neural quantum computing, quantum gravity computing, and what-not. Instead, his article begins with a long, detailed review of the well-known early history of computing, and then argues that the “long record of failure” of the AI field AGI-wise can only be remedied via a breakthrough in epistemology following on from the work of Karl Popper.

This bold, eccentric view of AGI is clearly presented in the article, but is not really argued for. This is understandable since we’re talking about a journalistic opinion piece here rather than a journal article or a monograph. But it makes it difficult to respond to Deutsch’s opinions other than by saying “Well, er, no” and then pointing out the stronger arguments that exist in favor of alternative perspectives more commonly held within the AGI research community.
I salute David Deutsch’s boldness, in writing and thinking about a field where he obviously doesn’t have much practical grounding. Sometimes the views of outsiders with very different backgrounds can yield surprising insights. But I don’t think this is one of those times. In fact, I think Deutsch’s perspective on AGI is badly mistaken, and if widely adopted, would slow down progress toward AGI dramatically.

The real reasons we don’t have AGI yet, I believe, have nothing to do with Popperian philosophy, and everything to do with:

- The weakness of current computer hardware (rapidly being remedied via exponential technological growth!)
- The relatively minimal funding allocated to AGI research (which, I agree with Deutsch, should be distinguished from “narrow AI” research on highly purpose-specific AI systems like IBM’s Jeopardy!-playing AI or Google’s self-driving cars).
- The integration bottleneck: the difficulty of integrating multiple complex components together to make a complex dynamical software system, in cases where the behavior of the integrated system depends sensitively on every one of the components.

**Assorted nitpicks, quibbles and major criticisms**

I'll begin here by pointing out some of the odd and/or erroneous positions that Deutsch maintains in his article. After that, I’ll briefly summarize my own alternative perspective on why we don’t have human-level AGI yet, as alluded to in the above three bullet points.

Deutsch begins by bemoaning the AI field’s “long record of failure” at creating AGI — without seriously considering the common counterargument that this record of failure isn’t very surprising, given the weakness of current computers relative to the human brain, and the far greater weakness of the computers available to earlier AI researchers. I actually agree with his statement that the AI field has generally misunderstood the nature of general intelligence. But I don’t think the rate of progress in the AI field, so far, is a very good argument in favor of this statement. There are too many other factors underlying this rate of progress, such as the nature of the available hardware.

He also makes a rather strange statement regarding the recent emergence of the AGI movement:

The field used to be called “AI” — artificial intelligence. But “AI” was gradually appropriated to describe all sorts of unrelated computer programs such as game players, search engines and chatbots, until the G for ‘general’ was added to make it possible to refer to the real thing again, but now with the implication that an AGI is just a smarter species of chatbot.

As the one who introduced the term AGI and founded the AGI conference series, I am perplexed by the reference to chatbots here. In a recent paper in *AAAI* magazine, resulting from the 2009 AGI Roadmap Workshop, a number of coauthors (including me) presented a host of different scenarios, tasks, and tests for assessing humanlike AGI systems.

The paper is titled “Mapping the Landscape of Human-Level General Intelligence,” and chatbots play a quite minor role in it. Deutsch is referring to the classical Turing test for measuring human-level AI (a test involving fooling human judges into believing a computer’s humanity, in a chat-room context). But the contemporary AGI community, like the mainstream AI community, tends to consider the Turing Test as a poor guide for research.
But perhaps he considers the other practical tests presented in our paper — like controlling a robot that attends and graduates from a human college — as basically the same thing as a “chatbot.” I suspect this might be the case, because he avers that AGI cannot possibly be defined purely behaviourally. In the classic ‘brain in a vat’ thought experiment, the brain, when temporarily disconnected from its input and output channels, is thinking, feeling, creating explanations — it has all the cognitive attributes of an AGI. So the relevant attributes of an AGI program do not consist only of the relationships between its inputs and outputs.

The upshot is that, unlike any functionality that has ever been programmed to date, this one can be achieved neither by a specification nor a test of the outputs. What is needed is nothing less than a breakthrough in philosophy. …

This is a variant of John Searle’s Chinese Room argument [video]. In his classic 1980 paper “Minds, Brains and Programs,” Searle considered the case of a person who knows only English, sitting alone in a room following English instructions for manipulating strings of Chinese characters. Does the person really understand Chinese?

To someone outside the room, it may appear so. But clearly, there is no real “understanding” going on. Searle takes this as an argument that intelligence cannot be defined using formal syntactic or programmatic terms, and that conversely, a computer program (which he views as “just following instructions”) cannot be said to be intelligent in the same sense as people.

Deutsch’s argument is sort of the reverse of Searle’s. In Deutsch’s brain-in-a-vat version, the intelligence is qualitatively there, even though there are no intelligent behaviors to observe. In Searle’s version, the intelligent behaviors can be observed, but there is no intelligence qualitatively there.

Everyone in the AI field has heard the Chinese Room argument and its variations many times before, and there is an endless literature on the topic. In 1991, computer scientist Pat Hayes half-seriously defined cognitive science as the ongoing research project of refuting Searle’s argument.

Deutsch attempts to use his variant of the Chinese Room argument to bolster his view that we can’t build an AGI without fully solving the philosophical problem of the nature of mind. But this seems just as problematic as Searle’s original argument. Searle tried to argue that computer programs can’t be intelligent in the same sense as people; Deutsch on the other hand, thinks computer programs can be intelligent in the same sense as people, but that his Chinese room variant shows we need new philosophy to tell us how to do so.

I classify this argument of Deutsch’s right up there with the idea that nobody can paint a beautiful painting without fully solving the philosophical problem of the nature of beauty. Somebody with no clear theory of beauty could make a very beautiful painting — they just couldn’t necessarily convince a skeptic that it was actually beautiful. Similarly, a complete theory of general intelligence is not necessary to create an AGI — though it might be necessary to convince a skeptic with a non-pragmatic philosophy of mind that one’s AGI is actually generally intelligent, rather than just “behaving generally intelligent.”

Of course, to the extent we theoretically understand general intelligence, the job of creating AGI is likely to be easier. But exactly what mix of formal theory, experiment, and informal qualitative understanding is going to guide the first successful creation of AGI, nobody now knows.
What Deutsch leads up to with this call for philosophical inquiry is even more perplexing:

Unfortunately, what we know about epistemology is contained largely in the work of the philosopher Karl Popper and is almost universally underrated and misunderstood (even — or perhaps especially — by philosophers). For example, it is still taken for granted by almost every authority that knowledge consists of justified, true beliefs and that, therefore, an AGI’s thinking must include some process during which it justifies some of its theories as true, or probable, while rejecting others as false or improbable.

This assertion seems a bit strange to me. Indeed, AGI researchers tend not to be terribly interested in Popperian epistemology. However, nor do they tend to be tied to the Aristotelian notion of knowledge as “justified true belief.” Actually, AGI researchers’ views of knowledge and belief are all over the map. Many AGI researchers prefer to avoid any explicit role for notions like theory, truth, or probability in their AGI systems.

He follows this with a Popperian argument against the view of intelligence as fundamentally about prediction, which seems to me not to get at the heart of the matter. Deutsch asserts that “in reality, only a tiny component of thinking is about prediction at all … the truth is that knowledge consists of conjectured explanations.”

But of course, those who view intelligence in terms of prediction would just counter-argue that the reason these conjectured explanations are useful is because they enable a system to better make predictions about what actions will let it achieve its goals in what contexts. What’s missing is an explanation of why Deutsch sees a contradiction between the “conjectured explanations” view of intelligence and the “predictions” view. Or is it merely a difference of emphasis?

In the end, Deutsch presents a view of AGI that comes very close to my own, and to the standard view in the AGI community:

An AGI is qualitatively, not quantitatively, different from all other computer programs. Without understanding that the functionality of an AGI is qualitatively different from that of any other kind of computer program, one is working in an entirely different field. If one works towards programs whose “thinking” is constitutionally incapable of violating predetermined constraints, one is trying to engineer away the defining attribute of an intelligent being, of a person: namely, creativity.

Yes. This is not a novel suggestion, it’s what basically everyone in the AGI community thinks; but it’s a point worth emphasizing.

But where he differs from nearly all AGI researchers is that he thinks what we need to create AGI is probably a single philosophical insight:

I can agree with the AGI-is-imminent camp: it is plausible that just a single idea stands between us and the breakthrough. But it will have to be one of the best ideas ever.

**The real reasons why we don’t have AGI yet**

Deutsch thinks the reason we don’t have human-level AGI yet is the lack of an adequate philosophy of mind to sufficiently, definitively refute puzzles like the Chinese Room or his brain-in-a-vat scenario, and that lead
us to a theoretical understanding of why brains are intelligent and how to make programs that emulate the key relevant properties of brains.

While I think that better, more fully-fleshed-out theories of mind would be helpful, I don’t think he has correctly identified the core reasons why we don’t have human-level AGI yet.

The main reason, I think, is simply that our hardware is far weaker than the human brain. It may actually be possible to create human-level AGI on current computer hardware, or even the hardware of five or ten years ago. But the process of experimenting with various proto-AGI approaches on current hardware is very slow, not just because proto-AGI programs run slowly, but because current software tools, engineered to handle the limitations of current hardware, are complex to use.

With faster hardware, we could have much easier to use software tools, and could explore AGI ideas much faster. Fortunately, this particular drag on progress toward advanced AGI is rapidly diminishing as computer hardware exponentially progresses.

Another reason is an AGI funding situation that’s slowly rising from poor to sub-mediocre. Look at the amount of resources society puts into, say, computer chip design, cancer research, or battery development. AGI gets a teeny tiny fraction of this. Software companies devote hundreds of man-years to creating products like word processors, video games, or operating systems; an AGI is much more complicated than any of these things, yet no AGI project has ever been given nearly the staff and funding level of projects like OS X, Microsoft Word, or World of Warcraft.

I have conjectured before that once some proto-AGI reaches a sufficient level of sophistication in its behavior, we will see an “AGI Sputnik” dynamic — where various countries and corporations compete to put more and more money and attention into AGI, trying to get there first. The question is, just how good does a proto-AGI have to be to reach the AGI Sputnik level?

The integration bottleneck

Weak hardware and poor funding would certainly be a good enough reason for not having achieved human-level AGI yet. But I don’t think they’re the only reason. I do think there is also a conceptual reason, which boils down to the following three points:

- Intelligence depends on the emergence of certain high-level structures and dynamics across a system’s whole knowledge base;
- We have not discovered any one algorithm or approach capable of yielding the emergence of these structures;
- Achieving the emergence of these structures within a system formed by integrating a number of different AI algorithms and structures is tricky. It requires careful attention to the manner in which these algorithms and structures are integrated; and so far, the integration has not been done in the correct way.

One might call this the “integration bottleneck.” This is not a consensus in the AGI community by any means — though it’s a common view among the sub-community concerned with “integrative AGI.” I’m not going to try to give a full, convincing argument for this perspective in this article. But I do want to point out that it’s a quite concrete alternative to Deutsch’s explanation, and has a lot more resonance with the work going on in the AGI field.
This “integration bottleneck” perspective also has some resonance with neuroscience. The human brain appears to be an integration of an assemblage of diverse structures and dynamics, built using common components and arranged according to a sensible cognitive architecture. However, its algorithms and structures have been honed by evolution to work closely together — they are very tightly inter-adapted, in somewhat the same way that the different organs of the body are adapted to work together. Due their close interoperation they give rise to the overall systemic behaviors that characterize human-like general intelligence.

So in this view, the main missing ingredient in AGI so far is “cognitive synergy”: the fitting-together of different intelligent components into an appropriate cognitive architecture, in such a way that the components richly and dynamically support and assist each other, interrelating very closely in a similar manner to the components of the brain or body and thus giving rise to appropriate emergent structures and dynamics.

The reason this sort of intimate integration has not yet been explored much is that it’s difficult on multiple levels, requiring the design of an architecture and its component algorithms with a view toward the structures and dynamics that will arise in the system once it is coupled with an appropriate environment. Typically, the AI algorithms and structures corresponding to different cognitive functions have been developed based on divergent theoretical principles, by disparate communities of researchers, and have been tuned for effective performance on different tasks in different environments.

Making such diverse components work together in a truly synergetic and cooperative way is a tall order, yet my own suspicion is that this — rather than some particular algorithm, structure or architectural principle — is the “secret sauce” needed to create human-level AGI based on technologies available today.

Achieving this sort of cognitive-synergetic integration of AGI components is the focus of the OpenCog AGI project that I co-founded several years ago. We’re a long way from human adult level AGI yet, but we have a detailed design and codebase and roadmap for getting there. Wish us luck!

Where to focus: engineering and computer science, or philosophy?

The difference between Deutsch’s perspective and my own is not a purely abstract matter; it does have practical consequence. If Deutsch’s perspective is correct, the best way for society to work toward AGI would be to give lots of funding to philosophers of mind. If my view is correct, on the other hand, most AGI funding should go to folks designing and building large-scale integrated AGI systems.

Until sufficiently advanced AGI has been achieved, it will be difficult to refute perspectives like Deutsch’s in a fully definitive way. But in the end, Deutsch has not made a strong case that the AGI field is helpless without a philosophical revolution.

I do think philosophy is important, and I look forward to the philosophy of mind and general intelligence evolving along with the development of better and better AGI systems.

But I think the best way to advance both philosophy of mind and AGI is to focus the bulk of our AGI-oriented efforts on actually building and experimenting with a variety of proto-AGI systems — using the tools and ideas we have now to explore concrete concepts, such as the integration bottleneck I’ve mentioned above. Fortunately, this is indeed the focus of a significant subset of the AGI research community.
And if you’re curious to learn more about what is going on in the AGI field today, I’d encourage you to come to the AGI-12 conference at Oxford, December 8–11, 2012.