MIGHT HUMAN-LIKE INTELLIGENCE REQUIRE HYPERCOMPUTATION?

Ben Goertzel
Novamente LLC
ben@goertzel.org

Abstract

Hypercomputation, defined roughly as trans-Turing-machine computation, has been proposed by several authors as a potential foundation for human intelligence. A simple, natural formalization of the scientific process is introduced, and related to the hypothesis that hypercomputation is a (necessary or optional) component of human or other intelligence. The conclusion is reached that this hypothesis could plausibly be true, but even if so could never be verified scientifically. Possible ways of creating humanlike AI even if human intelligence requires hypercomputation are discussed, including imitation, intuition and chance.

Keywords: Hypercomputation, intelligence, neural networks

1. INTRODUCTION

Most AI theorists believe that humanlike intelligence can ultimately be achieved within a digital computer program. But some mavericks disagree, and have argued for a hypothesis that I will call HHI: the Hypercomputable Humanlike Intelligence hypothesis, which suggests that the crux of humanlike intelligence is some sort of mental manipulation of uncomputable entities – i.e., some process of "hypercomputation" [1-6]. Hypercomputable processes, as I use the term here, are ones that can never be simulated on any ordinary Turing-machine-equivalent computer. For a proper formalization of the notion of hypercomputation, see [1].

Roger Penrose [7] has gone even further in this direction, suggesting that some future theory of physics is going to reveal that the dynamics of the physical world is also based on hypercomputation. In this case, he proposes, humanlike intelligence (and general intelligence more broadly) would be demonstrated to be a hypercomputable consequence of hypercomputable physical reality.

I find HHI and related ideas intuitively perplexing, feeling something to be fundamentally problematic about the notion of "hypercomputable physics" and in general the intersection between hypercomputation and scientific measurement and theory. The core reason for my intuitive discomfort is as follows: Science as we know it is always, in the end, about finite sets of finite-precision data. And this raises the question: how could hypercomputable entities ever really be necessary to explain this finite data? Intuitively, it seems clear that such entities should never be necessary. Any finite dataset has a finite explanation.

But, following this intuitive argument further, the next question then becomes whether in some cases invoking hypercomputable entities is the *best* way to explain some finite dataset (even if not the only way). Can the best way of explaining some set of, say, 10 or 1000

or 1000000 numbers be "This hypercomputable process, whose details you can never (by the very nature of hypercomputation) communicate in ordinary language in a finite amount of time, generated these numbers."

I also find this latter possibility intuitively problematic – and my goal in this paper is to present a semi-formalized argument as to why it doesn't make sense, under appropriate and reasonable assumptions. I present a highly general semi-formalization of the scientific process, and then argue that the hypothesis of an hypercomputable explanation for some observed phenomenon can never be scientifically verifiable – because there will always be simpler alternative explanations that match the same data, yet don't involve hypercomputation.

I consider the arguments presented here fairly damning for Penrose's idea of hypercomputable physics. If accepted, these arguments imply that the presumption of a hypercomputable phenomenon as part of a scientific explanation of a set of physics data is always *unnecessary*.

The implications of the present ideas for AI and neuroscience are a little subtler, and are roughly as follows. In the following (as above) I will refer frequently to "humanlike" intelligence, meaning intelligence that is roughly human-level and roughly humanlike (but could be artificial rather than human). Most of the conclusions presented actually refer to general intelligence more broadly, but restricting discussion to humanlike intelligence for the moment will keep things simpler.

Firstly, it's entirely possible that human intelligence relies on hypercomputable phenomena ... the HHI is a coherent and sensible hypothesis. However, if HHI holds, there are serious limits on the ways in which this reliance can be probed using the tools of science.

The model of the scientific process presented here leads to the conclusion that HHI is *falsifiable* (via creating a Turing-machine-based humanlike AI system, whose intelligent behaviors are scientifically explicable in terms of its internals) but is not scientifically *verifiable* based on any possible set of scientific evidence.

Based on this, if HHI is true and humanlike intelligence does indeed rely on hypercomputable processes, then there's no reason to place particular confidence in attempts to use science to guide the implementation of humanlike intelligence on ordinary Turing computers, beyond the confidence one would place in attempts to guide such implementation by non-scientific methods.

However, in this case that humanlike intelligence does rely on hypercomputation, one might still be able to create an artificially intelligent software program or physical device imitating natural intelligences, by following one's own brain's hypercomputable intuition, or by a chance process. I will discuss these possibilities in more detail below.

Bringsjord and Zenzen [6] suggest that the failure of current AI programs to solve numerous problems that humans find simple, should be taken as evidence in favor of a hypercomputable foundation for humanlik intelligence. However, the present arguments imply that such failures cannot coherently be taken as scientific evidence in favor of HHI – though individuals may of course choose to interpret them as *intuitive* evidence in this respect.

Finally, Penrose [7] also relates hypercomputation to the philosophy of mathematics, arguing that human mathematical ability is evidence of our hypercomputional capability. In the final section of this paper I present a brief disputation of this perspective, arguing analogously to the above arguments about science and hypercomputation, to the effect that the existence of uncomputable entities in mathematics is a hypothesis whose truth is unnecessary for the proving of mathematical theorems. All mathematical theorem-proving can be carried out equally well via adoption of the hypothesis that

"uncomputability" in mathematics is just a label attached to certain structures produced by computable formal systems.

2. A SIMPLE FORMALIZATION OF THE SCIENTIFIC PROCESS

In this section I'll prepare the way for the main thrust of the paper via providing a simplified formalization of the process of science. This formalization is related to the philosophy of science outlined in [8] and [9]; but that reference also considers many aspects not discussed here. The general conclusions arrived at here are not particularly sensitive to the details of the formalization of the scientific process used, but even so, to make the argument clear I have chosen to take a formal (or at least semi-formal) approach.

Consider a community of agents that use some language L to communicate. By a language, what I mean here is simply a set of finite symbol-sequences ("expressions"), utilizing a finite set of symbols.

Assume that a dataset (i.e., considered as a finite set of finite-precision observations) can be expressed as a set of pairs of expressions in the language L. So a dataset D can be viewed as a set of pairs

$$((d_{11}, d_{12}), (d_{21},d_{22}) ,..., (d_{n1},d_{n2}))$$

or else as a pair D=(D1,D2) where

$$D1=(d_{11},...,d_{n1})$$
$$D2=(d_{12},...,d_{n2})$$

Then, define an explanation of a dataset D as a set E_D of expressions in L, so that if one agent A1 communicates E_D to another agent A2 that has seen D1 but not D2, nevertheless A2 is able to reproduce D2.

(One can extend the above formalization via looking at precise explanations versus imprecise ones, where an imprecise explanation means that A2 is able to reproduce D2 only approximately, but this doesn't affect the argument significantly, so I'll leave this complication out from here on.)

If D2 is large, then for E_D to be an interesting explanation, it should be more compact than D2.

Note that I am not requiring E_D to generate D2 from D1 on its own. I am requiring that A2 be able to generate D2 based on E_D and D1. Since A2 is an arbitrary member of the community of agents, the validity of an explanation, as I'm defining it here, is relative to the assumed community of agents.

Note also that, although expressions in L are always finitely describable, that doesn't mean that the agents A1, A2, etc. are. According to the framework I've set up here, these agents could be infinite, hypercomputable, and so forth. I'm not assuming anything special about the agents, but I am considering them in the special context of finite communications about finite observations.

The above is a simple formalization of the scientific process, in a general and abstract sense, which I propose not as an ultimate definition of science but rather as a working definition to support easy argumentation and discussion. According to this formalization, science is about communities of agents linguistically transmitting to each other knowledge about how to predict some commonly-perceived data, given some other commonly-perceived data.

3. THE (DUBIOUS) SCIENTIFIC VALUE OF THE HYPERCOMPUTABLE

Next, moving closer to the theme of the paper, I turn to consider the question of what use it might be for A2 to employ some hypercomputable entity U in the process of using E_D to generate D2 from D1. My contention is that, under some reasonable assumptions, there is no value to A2 in using hypercomputable entities in this context.

D1 and E_D are sets of L-expressions, and so is D2. So what A2 is faced with, is a problem of mapping one set of L-expressions into another.

Suppose that A2 uses some process P to carry out this mapping. Then, if we represent each set of L-expressions as a bit string (which may be done in a variety of different, straightforward ways), P is then a mapping from bit strings into bit strings. To keep things simple we can assume some maximum size cap on the size of the bit strings involved (corresponding for instance to the maximum size expression-set that can be uttered by any agent during a trillion years).

The question then becomes whether it is somehow useful for A2 to use some hypercomputable entity U to compute P, rather than using some sort of set of discrete operations comparable to a computer program.

One way to address this question is to introduce a notion of simplicity. The question then becomes whether it is simpler for A2 to use U to compute P, rather than using some computer program.

And this, then, boils down to one's choice of simplicity measure.

Consider the situation where A2 wants to tell A3 how to use U to compute P. In this case, A2 must represent U somehow in the language L.

In the simplest case, A2 may represent U directly in the language, using a single expression (which may then be included in other expressions). There will then be certain rules governing the use of U in the language, such that A2 can successfully, reliably communicate "use of U to compute P" to A3 only if these rules are followed. Call this rule-set R_U. Let us assume that R_U is a finite set of expressions, and may also be expressed in the language L.

Then, the key question is whether we can have

$$complexity(U) < complexity(R_U)$$

That is, can U be less complex than the set of rules prescribing the use of its symbol S_U within the community of agents?

If we say NO, then it follows there is no use for A2 to use U internally to produce D2, in the sense that it would be simpler for A2 to just use R_U internally.

On the other hand, if we say YES, then according to the given complexity measure, it may be easier for A2 to internally make use of U, rather than to use R_U or something else finite.

So, if we choose to define complexity in terms of complexity of expression in the community's language L, then we conclude that hypercomputable entities are useless for science. Because, we can always replace any hypercomputable entity U with a set of rules for manipulating the symbol S_U corresponding to it.

If you don't like this complexity measure, you're of course free to propose another one, and argue why it's the right one to use to understand science – and then argue why this one supports the utility of hypercomputable explanations. The above discussion assumes that U is denoted in L by a single symbolic L-expression S_U, but the same basic argument holds if the expression of U in L is more complex.

4. IMPLICATIONS FOR ARTIFICIAL INTELLIGENCE

What do these arguments regarding scinece and hypercomputability have to say about HHI? They don't tell you anything definitive about whether humanlike intelligence can be achieved using traditional Turing computation. But what they do tell you is that, if

- humanlike intelligence can be studied using the communicational tools of science (that is, using finite sets of finite-precision observations, and

languages defined as finite strings on finite alphabets)
- one accepts the communication prior (length of linguistic expression as a measure of complexity)

then to the extent that humanlike intelligence is fundamentally hyper-computational, science is no use for studying it. Because science, as formalized here, can never distinguish between use of U and use of S_U. According to science, there will always be some computational explanation of any set of data, though whether this is the simplest explanation depends on one's choice of complexity measure.

However, none of this rules out the possibility that hypercomputation might in fact be a component, or even a critical component of humanlike intelligence. It just means that if this is so, neuroscience and AI are in significant part doomed sciences.

For instance, it seems possible that, in effect, all humans might have some kind of hypercomputational oracle machine in their brains. Zenil and Hernandez-Quiroz [5] have explored some mechanisms via which this might be the case, via developing mathematics of neural nets that use the full algorithmic information content of their real number weights, inputs and outputs.

And, if all people have the same or appropriately related internal neural oracle machines, then people could in effect communicate about the hypercomputable even though their language could never actually encapsulate what it is we're talking about. Discrete symbols could be used to communicate "control signals" of some sort from one person's neural hypercomputable process to another's.

In this case, some possible ways to create a humanlike AI system would be:

- **Imitation**: copying the relevant aspects of the way the human brain works without understanding them. Potentially, the copy could induce the right sort of hypercomputation process, for reasons we're unable to understand scientifically or communicate linguistically.
- **Intuitively**: just building the AI, guided by our internal hypercomputation processes, in ways that science and language can't encapsulate
- **Fortuitously**: just create an AI by luck, perhaps directed by some evolutionary process

However, if the arguments I've given above are accepted, what can't work in the case that humanlike intelligence requires hyper-computation is designing humanlike AI according to scientific principles. This restriction also rules out designing evolutionary processes for evolving humanlike intelligences -- but it doesn't rule out arriving at such evolutionary processes via imitation (e.g. of natural evolution) or intuition. In short, if intelligence is hypercomputable, then the only methods for creating AI are ones that do not involve scientifically understanding why the AI displays the intelligence it does.

Finally, it's worth noting one more implication of this whole line of argument, hinted at above, which is that the hypothesis that humanlike intelligence requires hypercomputation is falsifiable but not verifiable.

HHI is falsifiable via creating a Turing-computable humanlike AI that is "scientifically transparent" in the sense that the relationships between its internal operations and the humanlike-intelligence behaviors it displays are explained scientifically. It's interesting to note that the creation of a Turing-computable humanlike AI is not, in itself, a falsification of the hypothesis: because it's possible that one could, via chance or hypercomputable intuition, create a Turing-computable program that somehow elicits hypercomputable intelligence from some hypercomputable aspect of the universe that is not accessible to science. And it must be emphasized that this is not a

pragmatically farfetched idea. AI design in practice involves a combination of science and intuition; if the human brain involves hypercomputation, then this intuition might involve hypercomputation, and could lead to the programming of Turing-machine software that leads to interactions with hypercomputable (non-scientifically-accessible) aspects of the universe and hence gives rise to humanlike intelligence.

HHI is not scientifically verifiable because of the arguments given above: scientific observations of the behaviors (intelligent or otherwise) of systems are finite sets of finite-precision values, and we have seen above that there's no way to distinguish hypotheses involving hypercomputing from those that don't, using these sorts of datasets.

5. IMPLICATIONS FOR THE PHILOSOPHY OF MATHEMATICS

In this section we briefly and even less formally note a relation between the above ideas and the philosophy of mathematics, in which the status of uncomputable mathematical entities is a long-debated issue (see [10] for an interesting recent review and contribution). There are philosophies such as intuitionism that aim to ban hypercomputable entities from mathematics [11], but traditional mathematics embraces them. This debate has been drawn into the AI/hypercomputation debate by Penrose, who has attempted to use the generality of human mathematical reasoning as an argument against the Turing computability of the human brain.

What the perspective presented above does, in this context, is clarifies the notion of "existence" underlying the question of whether uncomputable mathematical entities "exist." In doing so, it clarifies the sense in which a Turing computable system can do mathematics involving uncomputable entities.

Parallel to the above arguments about hypercomputation and science, one may argue that the hypothesis of existent uncomputable entities is never necessary to explain the truth of any theorem. I omit details here due to space limitations, but Leitsch et al's [3] arguments essentially establish this, in a carefully formalized way. All that theorem-proving requires is the existence of certain symbols that humans manipulate according to certain computable formal rules, and may label with words such as "uncomputable." Whether uncomputable entities exist is thus portrayed as a philosophical issue that is irrelevant to the evaluation and proof of theorems, in the same sense that hypercomputation is irrelevant to the evaluation of scientific hypotheses.

To make this clearer in the context of the preceding arguments about science, consider the example of differential calculus, which in its traditional formulations relies centrally on the real line, a set the majority of whose members are uncomputable numbers. When doing calculus, is one using the uncomputable numbers as "existent" entities, or is one merely assigning the textual label "uncomputable" to certain aspects of a discrete formal system?

The question comes down to whether we have

complexity(real number line R) < complexity(axioms defining R)

If NO, then it means the mathematical mind is better off using the axioms for R than using R directly. And, I suggest, that is what we actually do when using R in calculus. We don't use R as an "actual existing entity" in any strong sense, we use R as an abstract set of axioms.

What would YES mean? It would mean that somehow we, as hypercomputable beings, used R as an internal source of intuition about continuity ... not thus deriving any conclusions beyond the ones obtainable using the axioms about R, but deriving conclusions in a way that we found subjectively simpler. This seems to be in line with Penrose's ideas; yet as we've argued above, under some very reasonable and simple assumptions about the scientific process,

it seems not to be a scientifically testable hypothesis.

6. CONCLUSION

So what, ultimately, is the status of the HHI, the hypothesis that humanlike intelligence (or intelligence more generally) requires hypercomputation? It is not nonsense, it's not obviously false, it's not incoherent – it could be the way the world works. It is something that could be proved false via the success of a scientifically transparent Turing-machine based AI, but it's not something that could ever be verified scientifically.

If indeed humanlike intelligence does require hypercomputation, then what this means is that the scientific method – insofar as it involves communications in discrete-symbol-based language about finite sets of finite-precision observations – is fundamentally limited in scope. This would be very disappointing to those of us enamored of the power of science, but is not an incoherent possibility, and would not rule out the creation of humanlike AI systems on Turing-computers or other physical infrastructures, via nonscientific or partially scientific methods.

References

[1] Maass, W. (2002). On the computational power of neural microcircuit models: Pointers to the literature. In Jos R. Dorronsoro, editor, Proc. of ICANN 2002

[2] Maass, W. and P. Orponen, On the effect of analog noise on discrete-time analog computations, Advances in Neural Information Processing 11 Systems 9, 1997, 218224; journal version: Neural Computation 10, 1998

[3] Alexander Leitsch, Günter Schachner and Karl Svozil, ``How to Acknowledge Hypercomputation?", Complex Systems 18, 131-143, (2008)

[4] Copeland, J. (2002) Hypercomputation, Minds and machines, v. 12, pp. 461-502

[5] Zenil, Hector and Francisco Hernandez-Quiroz (2005). On the possible Computational Power of the Human Mind. Complexity, Science and Society Conference, 2005, University of Liverpool, UK.

[6] S. Bringsjord and M. Zenzen, Superminds, People harness hypercomputation and more, Kluwer, 2003.

[7] R. Penrose, Shadows of the Mind: A Search for the Missing Science of Consciousness, Oxford: Oxford University Press, 1994.

[8] Goertzel, Ben (2006). The Hidden Patterns. Brown Walker

[9] Goertzel, Ben (2004). A Social-Computational-Probabilist Philosophy of Science. Dynamical Psychology E-Journal.

[10] Kelly, K. (2004). Uncomputability: the problem of induction internalized. Theoretical Computer Science. Volume 317 , Issue 1-3

[11] Heyting, Arend (1971) [1956]. Intuitionism: An Introduction (3d rev. ed. ed.). Amsterdam: North-Holland Pub. Co