

Subverting the AI Superintelligence Myth: Bergson's Remarkable Model of Mind

Atlantis Rising #112, July/August

Humanity faces a crisis. At the rapid rate Artificial Intelligence (AI) is developing, we will quickly reach the point of Ray Kurzweil's "Singularity" – the point where AIs equal human intelligence. Then AIs will proceed beyond – to "Superintelligence." The superintelligent AIs will be so smart they will have the capability to destroy their creators. Unless, as a new theme in the AI world argues, we learn to impart human values to these machines – a process (the "values loading problem") now seen as extremely tricky – we are toast. Yup, like toast, we will be eaten for breakfast. Maybe we last until lunch.

The AI Onslaught

AI has made tremendous progress. AI programs are current champions in a multitude of games – chess, checkers, backgammon, Othello, even Jeopardy. Unfortunately, each of these programs has logic and algorithms very specific to the game; none are generalizable to a true, general intelligence (AGI) – what humans are considered to be. AI has seen three major approaches. The first is the standard "symbolic" programming with which most are familiar. It has led to theorem provers, language understanding programs, problem solvers. Another approach – evolutionary algorithms – has attempted to "evolve" programs. A third, neural networks, has numerous achievements, generally in applying statistical techniques to detecting patterns. It boasts nearly 150,000 academic papers. But all lead to a large black hole from which no clear exit is seen. On the other side stand two high, admittedly untaken hills – commonsense knowledge and true language comprehension.

The two hills are connected. Here is a commonsense problem: Given a 12" cubical box, a razorblade, rubber bands, staples, a pencil, string, toothpicks, a piece of cheese – create a mousetrap. We might make a "crossbow" where the pencil is inserted into a hole in the box's side, pulled back outside the box by the rubber bands, notched in place by the toothpick, and a string tied to the toothpick and the cheese. Or we might create a "beheader," with razorblade notched into the pencil, one pencil end anchored in the box corner, the pencil-axe raised up by the toothpick, with string attached again to toothpick and cheese. Both solutions require concrete knowledge and experience of physical dynamics, forces, properties of materials. Both are "analogic" solutions. The linguistic statement, "The mouse trap is a crossbow," is an analogy. Douglas Hofstadter of *Gödel, Escher, Bach* fame, in his recent tome (*Surfaces and Essences*), while showing indisputably that analogy is the foundational operation of human thought, ridicules all current AI language understanding programs (SIRI included). The problem is that language is based on analogy, and Hofstadter clearly doubts that computers, as we know them, can handle analogy at all.

But the black hole and untaken hills are not seen as stopping the superintelligence onslaught. Oxford professor, Nick Bostrom (*Superintelligence: Paths, Dangers, Strategies*), having taken us to the edge of the black hole in his book, veers away, placing his bet on whole brain emulation (WBE). WBE relies on the neuroscientists to map, via neural recordings, the functioning of the brain. But this is a shaky bet; WBE is an Everest. We face a 85 billion neuron brain with roughly 1000 types of neurons, the functions of none of these types we understand. We do not know basic facts such as how memory (our experience) is stored. We are quite certain that the brain is *not* using what we understand currently as "computation," but we do not know what this other form is. We face data from neural recordings that will be so massive, it will be in zeta-bytes, yet any interpretation will be completely dependent on a guiding theory – note, a theory – when we have none such. It will be, as neuroscientists Marcus and Freeman note (*The Future*

of the Brain), like trying to learn what a laptop is and does by taking electrical recordings of its components when we have no theory of, or knowledge of, the existence of something called “software.”

This is to say, we really have no clue what type of “device” the brain actually is. Yet Bostrom, with the rest of AI, sails serenely on from this subject of brain emulation, confident without a qualm that we will have recreated the brain as a silicon and wires device. Since it is certain that we will have electronics, we can speed up the transmission velocities, say, by 10,000x, thus allowing the device to quickly develop, creating and moving to a super intelligence and thus to the terrible threat of that future breakfast. But, but... what if the brain is a “device” that cannot be replicated in silicon at all?

Bergson’s Holographic Vision

Neuroscience, admitted Marcus and Freeman, has no understanding of how (or if) experience is stored in the brain. This is exacerbated by the fact that we have no theory of what experience is – we cannot explain the origin of our image of the external world – the coffee cup in front of us, with spoon stirring away, on the kitchen table. This image is our experience, and our experience is our consciousness – a word (consciousness) never once seen in Bostrom’s book. Yet this very subject forms part of that missing notion of “software.”

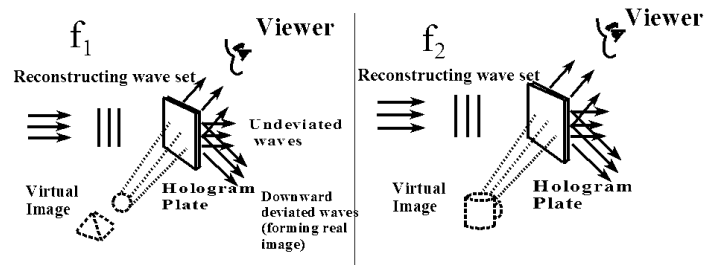
Nowhere in the brain, nor in a computer, do we see anything like our image of the coffee cup with its stirring spoon. In the brain, to pick a level of description, we see only neural-chemical flows. In a computer, we see flipping 1/0 bits, magnetic cores shifting on and off, etc. As external observers, we can *attribute* an image as existing over these flows or bit flips, but it is only an attribution; there is actually nothing like an image of the kitchen and its cup. This is the great “hard problem” of consciousness made famous by philosopher David Chalmers, still unsolved: given any neural or computer architecture, how do we account for the *qualia* of the perceived world? “Qualia” refers to the qualities in the image – the “whiteness” of the cup, the “silveriness” of the spoon. But everything in the image, to include its forms – the stirring spoon, the waving curtains – is “qualia.” The more general problem is accounting for our image of the external world – our experience. This was the starting point of the great French philosopher, Henri Bergson, in 1896, in his book, *Matter and Memory*.



Henri Bergson

Bergson, at his height, circa WWI, was the most famous philosopher in France. French newspapers debated whether he should move his lectures to the Paris opera hall so the ladies of society could hear “the great Bergson.” After his 1907 work, *Creative Evolution*, the

French army claimed to possess the Élan Vital – the force that drives evolution. He retired early due to academic resentment of his popularity, dying in 1940 from pneumonia in Nazi occupied Paris after waiting hours in a line to declare himself a Jew.



Holographic reconstruction. The reconstructive wave, modulated to frequency 1, reconstructs (or “specifies”) the stored wave front (image) from a pyramid/ball. Modulated to frequency 2, it now specifies the wave front from the cup.

In *Matter and Memory*, Bergson aimed to question whether the brain stores experience at all. Thus he noted that indeed we see nothing like a “photograph” of the external world (the kitchen and cup) within the brain. But he went on:

But is it not obvious that the photograph, if photograph there be, *is already taken, already developed in the very heart of things and at all points in space.* No metaphysics, no physics can escape this conclusion. Build up the universe with atoms: Each of them is subject to the action, variable in quantity and quality according to the distance, exerted on it by all material atoms. Bring in Faraday's centers of force: The lines of force emitted in every direction from every center bring to bear upon each the influence of the whole material world. Call up the Leibnizian monads: Each is the mirror of the universe.

Here, in the photograph “already developed in the very heart of things and at all points in space,” Bergson was stating – fifty years before holography’s discovery by Gabor – his prescient vision of the universe as a holographic field. In an earlier passage, he had previewed this, viewing the field as a vast interference pattern – a vast field of “real actions.” Any given “object” acts upon all other objects in the field, and is in turn acted upon by all other objects. It is in fact obliged:

...to transmit the whole of what it receives, to oppose every action with an equal and contrary reaction, to be, in short, merely the road by which pass, in every direction the modifications, or what can be termed *real actions* propagated throughout the immensity of the entire universe.

But Bergson envisioned the role of the brain differently from anything in current holographic lore – it is not, for example, as popularized by neuroscientist Karl Pribram, that the brain is simply a “hologram.” Following on the “photograph” passage he noted:

Only if when we consider any other given place in the universe we can regard the action of all matter as passing through it without resistance and without loss, and the

photograph of the whole as translucent: Here there is wanting behind the plate the black screen on which the image could be shown. Our “zones of indetermination” [organisms] play in some sort the part of that screen. They add nothing to what is there; they effect merely this: *That the real action passes through, the virtual action remains.*

In essence, Bergson is viewing the brain, over all its dynamics, as creating a *modulated reconstructive wave* passing through the holographic field. A reconstructive wave, modulated to a certain frequency and passing through a hologram, selects a set of information from the hologram plate – one wave front from a set of stored wave fronts – and specifies the source (a cup, a pyramid) of that original wave front. As a “wave” the brain also is specific to, or specifies, a portion of the information in the field, now by this process, an “image” of the field – the coffee cup on the table. For the brain, the selection principle is via information relating to the body’s ability to act. This is why Bergson stated that in essence perception is virtual action – we are seeing how we can act.

Critical here is Bergson’s model of time. In watching the spoon stirring, the coffee swirling or a fly buzzing by, we are viewing the *past* – a portion of the past transformation of the holographic field. But Bergson viewed this transformation (time) as indivisible. There are no separate, mutually external “instants” in this transformation, where each “instant” passes away into non-existence (the “past”) as the next instant (the “present”) arrives. He viewed this transformation as a melody where each note (read, “instant”) interpenetrates the next forming an organic continuity, and where each note reflects the entire preceding series. With this view of the time-transformation of the universal field, we now have the basis for a form of “primary memory” that allows us to see stirring spoons, leaves twisting and falling or flies buzzing by, for the brain’s specification then is simply to a portion of this (past) indivisible motion of the field – a motion that does not and has not fallen into the past, ceasing to exist, and therefore need not be stored somehow in the brain to preserve it.

What this says is that perception – our experience – is not occurring solely within the brain. Rather, it is a specification of events that are precisely where they appear – within the ever transforming external field. It is also a specification that is at a scale of time determined by the brain’s dynamics – a “buzzing” fly versus a fly slowly flapping his wings like a heron. Therefore, since experience is not occurring solely within the brain, experience cannot be solely stored there in the first place. This is why current theory is having such trouble determining how the brain stores experience. To retrieve this experience, to over simplify, the brain is again acting as a reconstructive wave passing through the 4-D holographic field. Modulated to the proper pattern, past events are reconstructed. Damage to neural mechanisms underlying this retrieval-modulation make it appear that “stored” experiences are “erased.”

A Real, Concrete Dynamics

Clearly, in this view, the brain is a quite different “device,” and perception, memory and thought are employing a completely different, yet unknown, form of computation. To create a reconstructive wave effectively specifying information (as an image) in the holographic field, the brain must be achieving a very concrete dynamics via very specific materials and forces. We cannot create an AC motor out of rubber bands, toothpicks and cheese; we need precise materials interacting in a precise way to create an electric field of force. This begs the question of whether the brain-device – being simultaneously a very concrete wave – can be embodied in silicon, wire and transistors at all, but rather, to support such a reconstructive wave, all the biological stuff comprising the brain, with all its quantum dynamics, is absolutely required. In other words, it is not a question of speculating, as Bostrom and AI, whether we will achieve human equivalence in 2075, 2100 or 3100, it is a question of what the ultimate “device” we ultimately create (brain/body – AI version) will look like. This answer completely determines whether controlling such a device, or “loading values,” or significantly increasing its intelligence, is going to be any problem or reality at all.

That “values loading” problem – it is based in our commonsense knowledge and experience – so problematic to AI – where we learn the “value” of the flat of a knife (as opposed to a noodle) for

stirring coffee, or of rubber bands for launching pencils. But it is worse, for to achieve intelligence, the brain as a self organizing dynamic system requires several years. Per Jean Piaget, the great child development theorist (and student of Bergson), it takes two years to enable the child to achieve the base concepts of causality, object, space and time, also the capability of explicit memory (conscious localization of past events in time), and even the ability to *symbolize* the events of the world. It requires to the age of seven to achieve concrete operations which include further grasp of space, time and number, and to the age of twelve for formal operations which include forms of logic and thought we take for granted. In other words, the brain is not only an organized structure, but a structure changing its organization along a complex (and not understood) trajectory purposed to achieve these conceptual and logical capabilities.

As to worries about this trajectory for intelligence development – for nature a necessity – AI has immunized itself, yet if the brain is a reconstructive wave via biological stuff, it is questionable what “speeding up” this development process means. I suspect there will be plenty of time to teach our superintelligent AI not only to read that breakfast menu, but to learn not to choose us.

More on Bergson/AI: *Time and Memory: A Primer on the Scientific Mysticism of Consciousness* (Amazon).