Journal of Evolution & Technology

A peer-reviewed electronic journal
published by the Institute for Ethics
and
Emerging Technologies

ISSN 1541-0099

22(1) –November 2011

Misbehaving Machines: The Emulated Brains of Transhumanist Dreams

Corry Shores
Department of Philosophy
Catholic University of Leuven
corry.shores@hiw.kuleuven.be

Journal of Evolution and Technology - Vol. 22 Issue 1 – November 2011 - pgs 10-22

Abstract

Enhancement technologies may someday grant us capacities far beyond what we now consider humanly possible. Nick Bostrom and Anders Sandberg suggest that we might survive the deaths of our physical bodies by living as computer emulations. In 2008, they issued a report, or "roadmap," from a conference where experts in all relevant fields collaborated to determine the path to "whole brain emulation." Advancing this technology could also aid philosophical research. Their "roadmap" defends certain philosophical assumptions required for this technology's success, so by determining the reasons why it succeeds or fails, we can obtain empirical data for philosophical debates regarding our mind and selfhood. The scope ranges widely, so I merely survey some possibilities, namely, I argue that this technology could help us determine (1) if the mind is an emergent phenomenon, (2) if analog technology is necessary for brain emulation, and (3) if neural randomness is so wild that a complete emulation is impossible.

Introduction

Whole brain emulation succeeds if it merely replicates human neural functioning. Yet for Nick Bostrom and Anders Sandberg, its success increases when it perfectly replicates a specific person's brain. She might then survive the death of her physical body by living as a computer emulation. This prospect has transhumanist proponents. Philosophers who consider themselves transhumanists believe that our rapidly advancing human enhancement technologies could radically transform the human condition. One such transhumanist technology would allow our minds to think independently of our bodies, by being "uploaded" to a computer. Brain emulation, in its ultimate form, would then be a sort of mental uploading.

In 2008, Nick Bostrom and Anders Sandberg compiled the findings from a conference of philosophers, technicians and other experts who had gathered to formulate a “roadmap” of the individual steps and requirements that could plausibly develop this technology. Their vision for this technology’s advancement is based on a certain view of human consciousness and the mind-body relation. As I proceed, I will look more closely at these philosophical assumptions individually. For now let it suffice to say that I will adopt the basic framework of their philosophy of mind. Put simply, the authors and I regard human consciousness as a phenomenon emerging from the computational dynamics of some physical “machinery,” be it nervous tissue, silicon chips, or whatever else is capable of performing these complex operations. This involves a sort of “emergent dualism” where consciousness depends on the workings of its physical substrate while at the same time operating somehow at an emergent level. It means that minds are, on the one hand, embodied by their underlying “machinery,” while on the other hand, the mind is not limited to its given computational embodiment but can extend into other machines, even ones of a very different material composition.

Although I adopt these basic assumptions, I will explore research that calls into question certain other ones. For example, although the authors diminish the importance of analog computation and noise interference, there are findings and compelling arguments that suggest otherwise. As well, there is reason to think that the brain’s computational dynamics would not call for Bostrom’s and Sandberg’s hierarchical model for the mind’s emergence. And finally, I will argue on these bases that if brain emulation were to be carried out to its ultimate end of replicating some specific person’s mind, the resulting replica would still over time develop divergently from its original.

1. We are such stuff as digital dreams are made on

When writing of mental uploading, transhumanists often cite Hans Moravec’s *Mind children: The future of robot and human intelligence*. In this text, Moravec proposes his theory of *transmigration*, which involves extracting a person’s mind from her brain and storing it in computer hardware. To help us imagine one way this procedure might be performed, he narrates a futuristic scenario in which the transition from brain to computer is performed gradually and carefully. In this story, a patient is kept lucid while she undergoes an operation on her brain. After the top of her skull is removed, sophisticated devices monitor the activities of the neurons in a very narrow layer at the exposed surface of her brain tissue. Then, a computer program develops a model emulating these selected neurons’ behavior by finding their patterns and regularities. Eventually the emulation becomes so accurate that it mimics the activity of this top layer all on its own. The device then temporarily overrides the functioning of that thin neural region and lets the computer emulation take over the workings of that layer. If the patient confirms that she feels no change in her consciousness despite part of it already being computer controlled, then that top layer of neural tissue is permanently removed while the emulation continues to act in its place. This process is repeated for each deeper and deeper layer of brain tissue, until all of it has been removed. When the device is finally withdrawn from the skull, the emulated brain activity is taken away with it, causing the patient’s body to die. Yet supposedly, her consciousness remains, only now in the form of an emulation that has been given a robotic embodiment (Moravec 1988, 108-109).

Moravec believes that our minds can be transferred this way, because he does not adopt what he calls the *body-identity position*, which holds that the human individual can only be preserved if the continuity of its “body stuff” is maintained. He proposes instead what he terms the *pattern-identity* theory, which defines the essence of personhood as “the *pattern* and the *process* going on in my head and body, not the machinery supporting that process. If the process is preserved, I am preserved. The rest is mere jelly” (Moravec 1988, 108-109). He explains that over the course of our lives, our bodies regenerate themselves, and thus all the atoms present in our bodies at birth are replaced half-way through our life-spans; "only our pattern, and only some of it at that, stays with us until our death" (Moravec 1988, 117). It should not then be unreasonable to think that we may also inhabit a computerized robot-body that functions no differently than does our organic body.

This position suggests a paradoxical dualism, in which the mind is separate from the body, while also being the product of the patterns of biological brain processes. One clue for resolving the paradox seems to lie in this sentence: "though mind is entirely the consequence of interacting matter, the ability to copy it from one storage medium to another would give it an independence and an identity apart from the machinery that runs the program" (Moravec 1988, 117). The mind is an independent and separate entity that nonetheless is the *consequence of interacting matter*. On account of our neuronal structure and its organizational dynamic, an independent entity – our mind – *emerges*.

For N. Katherine Hayles, Moravec's description of mind transfer is a nightmare. She observes that mental uploading presupposes a cybernetic concept. Our selfhood extends into intersubjective systems lying beyond our body's bounds (Hayles 1999, 2). For example, Picasso in a sense places himself into his paintings, and then they reflect and communicate his identity to other selves. This could have been more fully accomplished if we precisely emulated his brain processes.

Hayles, who refers to thinkers like Moravec as "posthumanists," claims that they hold a view that "privileges information pattern over material instantiation" (Hayles 1999, 2). So according to this perspective, we are in no way bound to our bodies:

> the posthuman view configures human being so that it can be seamlessly articulated with intelligent machines. In the posthuman, there are no essential differences or absolute demarcations between bodily existence and computer simulation, cybernetic mechanism and biological organism, robot teleology and human goals. (Hayles 1999, 3)

In his article, "Gnosis in cyberspace? Body, mind and progress in posthumanism," Oliver Krueger writes that a basic tenet of posthumanism is the disparagement of the body in favor of a disembodied selfhood. He cites Hayles' characterization of posthumanism's fundamental presupposition that humans are like machines determined by their "pattern of information and not by their devaluated prosthesis-body" (Krueger 2005, 78). These thinkers whom Krueger refers to as posthumanists would like to overcome the realms of matter and corporeality in which the body resides so as to enter into a pure mental sphere that secures their immortality. They propose that the human mind be "scanned as a perfect simulation" so it may continue forever inside computer hardware (Krueger 2005, 77). In fact, Krueger explains, because posthumanist philosophy seeks the annihilation of biological evolution in favor of computer and machine evolution, their philosophy necessitates there be an immortal existence, and hence, "the idea of uploading human beings into an absolute virtual existence inside the storage of a computer takes the center stage of the posthumanist philosophy" (Krueger 2005, 80). William Bainbridge nicely articulates this belief:

> I suggest that machines will not replace humans, nor will humans become machines. These notions are too crude to capture what will really happen. Rather, humans will realize that they are by nature dynamic patterns of information, which can exist in many different material contexts. (Bainbridge 2007, 211)

Our minds, then, would be patterns that might be placed into other embodiments. So when computers attain this capacity, they will embody our minds by emulating them. Then no one, not even we ourselves, would know the difference between our originals and our copies.

2. Encoding all the sparks of nature

Bostrom and Sandberg do not favor Moravec's "invasive" sort of mind replication that involves surgery and the destruction of brain tissue (Bostrom and Sandberg 2008, 27). They propose instead *whole brain emulation*. To emulate someone's neural patterns, we first scan a particular brain to obtain precise detail of its structures and their interactions. Using this data, we program an emulation that will behave essentially the same as the original brain. Now first consider how a gnat's flight pattern seems irrational and random. However, the motion of a whole swarm is smooth, controlled, and intelligent, as though the whole group of gnats has a mind of its own. To emulate the swarm, perhaps we will not need to understand how the whole swarm thinks but instead merely learn the way one gnat behaves and interacts with other ones. When we combine thousands of these emulated gnats, the swarm's collective intelligence should thereby appear. Whole brain emulation presupposes this principle. The emulation will mimic the human brain's functioning on the cellular level, and then automatically, higher and higher orders of organization should spontaneously arise. Finally human consciousness might emerge at the highest level of organization.

Early in this technology's development, we should expect only simpler brain states, like wakefulness and sleep. But in its ultimate form, whole brain emulation would enable us to make back-up copies of our minds so we might then survive our body's death.

Bostrom's and Sandberg's terminological distinction between *emulation* and *simulation* indicates an important success criterion for whole brain emulation. Although both simulations and emulations model the original's *relevant properties*, the simulation would reproduce only some of them, while the emulation would replicate them all. So an emulation is a one-to-one modeling of the brain's functioning (Bostrom and Sandberg 2008, 7). Hillary Putnam calls this a *functional isomorphism*, which is "a correspondence between the states of one and the states of the other that preserves functional relations" (Putnam 1975, 291). The brain and its emulation are "black boxes": our only concern is the input/output patterns of these enclosed systems. We care nothing of their contents, which might as well be blackened from our view (Minsky 1972, 13). So if both systems respond with the same sequence of behaviors when we feed them the same sequence of stimuli, then they are functionally isomorphic. Hence the same mind can be realized in two physically different systems. Putnam writes, "a computer made of electrical components can be isomorphic to one made of cogs and wheels or to human clerks using paper and pencil" (Putnam 1975, 293). Their insides may differ drastically, but their outward behaviors must be identical. Hence, when a machine, software-program, alien life-form, or any other such alternately physically-realized operation-system is functionally isomorphic to the human brain, then we may conclude, says Putnam, that it shares a mind like ours (Putnam 1975, 292-293). This theory of mental embodiment is called *multiple realizability*: "the same mental property, state, or event can be implemented by different physical properties, states, and events" (Bostrom and Sandberg 2008, 14). David Chalmers recounts the interesting illustration of human neural dynamics being realized by communications between the people of China. We are to imagine each population member behaving like a single neuron of a human brain by using radio links to mimic neural synapses. In this way they would realize a functional organization that is isomorphic to the workings of a brain (Chalmers 1996, 97).

There are various levels of successfully attaining a functionally isomorphic mind, beginning with a simple “parts list” of the brain’s components along with the ways they interact. Yet, the highest levels are the most philosophically interesting, write Bostrom and Sandberg. When the technology achieves *individual brain emulation,* it produces emergent activity characteristic of that of one particular (fully functioning) brain. It is more similar to the activity of the original brain than any other brain. The highest form is a *personal identity emulation*: “a continuation of the original mind; either as numerically the same person, or as a surviving continuer thereof,” and we achieve such an emulation when it becomes rationally self-concerned for the brain it emulates (Bostrom and Sandberg 2008, 11).

3. Arising minds

Bostrom’s and Sandberg’s “Roadmap” presupposes a physicalist standpoint, which in the first place holds that everything has a physical basis. Minds, then, would emerge from the brain’s pattern of physical dynamics. So if you replicate this pattern-dynamic in some other physical medium, the same mental phenomena should likewise emerge. Bostrom and Sandberg write that “sufficient apparent success with [whole brain emulation] would provide persuasive evidence for *multiple realizability*” (Bostrom and Sandberg 2008, 14).

Our mind’s emergence requires a dynamic process that Paul Humphreys calls *diachronic pattern emergence* (Humphreys 2008, 438). According to emergentist theories, all reality is made-up of a single kind of stuff, but its parts aggregate and assemble into dynamic organizational patterns. The higher levels exhibit properties not found in the lower ones; however, there cannot be a higher order without lower ones underlying it (Clayton 2006, 2-3). Todd Feinberg uses the example of water to illustrate this. The H_2O molecule does not itself bear the properties of liquidity, wetness, and transparency, although an aggregate does (Feinberg 2001, 125). Emergent features go beyond what we may expect from the lower level, and hence the higher levels are greater than the sum of their parts.

Our minds emerge from the complex dynamic pattern of all our neurons communicating and computing in parallel. Roger Sperry offers compelling evidence. There are “split brain” patients whose right and left brain hemispheres are disconnected from one another, and nonetheless, they have maintained unified consciousness. However, there is no good account for this on the basis of neurological activity, because there is no longer normal communication between the two brain-halves (Clayton 2006, 20). For this reason, Sperry concludes that mental phenomena are emergent properties that “govern the flow of nerve impulse traffic.” According to Sperry, “Individual nerve impulses and other excitatory components of a cerebral activity pattern are simply carried along or shunted this way and that by the prevailing overall dynamics of the whole active process” (Sperry quoted in Clayton 2006, 20). Yet it works the other way as well:

> The conscious properties of cerebral patterns are directly dependent on the action of the component neural elements. Thus, a mutual interdependence is recognized between the sustaining physico-chemical processes and the enveloping conscious qualities. The neurophysiology, in other words, controls the mental effects, and the mental properties in turn control the neurophysiology. (Sperry quoted in Clayton 2006, 20)

In his book *The emergent self*, William Hasker provides a more detailed account specifically of how the mind can emerge from lower-level neuronal activity. From Sperry he obtains the notion that consciousness has causal influence acting “downward” upon the neural processes out of which the mind emerges (Hasker 1999, 180). If causation occurs exclusively within one layer, it is *intra-ordinal*; and, if one stratum has causal influence upon another, it is *trans-ordinal*

(O’Connor and Wong, 2006). When a higher level emerges, it does so on account of the lower level’s particular organization upwardly-causing it to come into being.

Now if the higher level can act independently of the lower level and also influence it downwardly, then perhaps not all instances of downward causation are first caused by rearrangements of the lower level’s constituents. Yet Hasker notes that the mind-body relation is further complicated by our minds’ dependence on our neural substrates. From Karl Popper, then, he derives the idea that the emergent mind is distinct from the brain, but yet inhabits it: if the brain were to be transplanted, the same mind would then occupy a new body (Hasker 1999, 187).

Moreover, Hasker rejects a Cartesian dualistic position that says the mind is somehow a separate element “added to” the brain from an exterior metaphysical realm. He believes that mental properties “manifest themselves when the appropriate material constituents are placed in special, highly complex relationships” (Hasker 1999, 189-190). He offers the analogy of magnetic fields, which he says are distinct from the magnets producing them; for, they occupy a much broader space. The magnetic field is generated because its “material constituents are arranged in a certain way – namely, when a sufficient number of the iron molecules are aligned so that their ‘micro-fields’ reinforce each other and produce a detectable overall field” (Hasker 1999, 190). Once generated, the field exerts its own causality, which affects not only the objects around it, but even that very magnet itself.

Hence Hasker’s analogy: just as the alignment of iron molecules produces a field, so too the particular organization of the brain’s neurons generates its field of consciousness (1999, 190). As a field, the mind bears physical extension, and is thus not akin to Descartes’ mind. Rather, the emergent consciousness-field permeates and haloes our brain-matter, occupying its space and traveling along with it (Hasker 1999, 192). Because this “soul-field” is in one way inherent in the neuronal arrangements, but in another way is independent from them, he terms his position *emergent dualism* (Hasker 1999, 194). Thus, he remains a mind-body dualist without encountering Descartes’ difficulty in accounting for the interaction between the mind and brain. In a similar way, William Lycan defends the idea that our minds can occupy space. He asks: “Why not suppose that minds are located where it feels as if they are located, in the head behind the eyes?” (Lycan 2009, 558).

Now let’s suppose that whole brain emulation continually fails to produce emergent mental phenomena, despite having developed incredible computational resources for doing so. This might lead us to favor Todd Feinberg’s argument that the mind does not emerge from the brain to a higher order. He builds his argument in part upon Searle’s distinction between two varieties of conscious emergence. Searle first has us consider a system made of a set of components, for example, a rock made up of a conglomerate of molecules. The rock will have features not found in any individual molecule; its weight of ten pounds is not found entirely in any molecular part. However, we can deduce or calculate the weight of the rock on the basis of the weights of its molecules. Yet, what about the solidity of the rock? This is an example of an emergent property that can be explained only in terms of the interactions among the elements (Searle 1992, 111). Consciousness, he argues, is an emergent property based on the interactions of neurons, but he disputes a more “adventurous conception,” which holds that emergent consciousness has capacities not explainable on the basis of the neurons’ interactivity: “the naïve idea here is that consciousness gets squirted out by the behaviour of the neurons in the brain, but once it has been squirted out, then it has a life of its own” (Searle 1992, 112). Feinberg will build from Searle’s position in order to argue for a non-hierarchical conception of mental emergence. So while Feinberg does in fact think consciousness results from the interaction of many complex layers of neural organization, no level emerges to a superior status. He offers the example of visual recognition and has us consider when we recognize our grandmother. One broad layer of neurons transmits information about the whole visual field. Another more selective layer picks-out lines. Then an even narrower layer detects shapes. Finally the information arrives at the "grandmother cell," which only fires when she is the one we see. But this does not make the grandmother cell emergently higher. Rather, all the neural layers of organization must work together simultaneously to achieve this recognition. The brain is a vast network of interconnected circuits, so we cannot say that any layer of organization emerges over-and-above the others (Feinberg 2001, 130-31).

Yet perhaps Feinberg looks too much among the iron atoms, so to speak, and so he never notices the surrounding magnetic field. Nonetheless, his objection may still be problematic for whole brain emulation, because Bostrom and Sandberg write:

> An important hypothesis for [whole brain emulation] is that in order to emulate the brain we do not need to understand the whole system, but rather we just need a database containing all necessary low-level information about the brain and knowledge of the local update rules that change brain states from moment to moment. (Bostrom and Sandberg 2008, 8)

But if Feinberg's holistic theory is correct, we cannot only emulate the lower levels and expect the rest to emerge spontaneously; for, we need already to understand the higher levels in order to program the lower ones. According to Thompson, Varela, and Rosch, "The brain is thus a highly cooperative system: the dense interconnections among its components entail that eventually everything going on will be a function of what all the components are doing" (Thompson et al. 1991, 94). For this reason, "the behavior of the whole system resembles a cocktail party conversation much more than a chain of command" (1991, 96).

If consciousness emerges from neural activity, perhaps it does so in a way that is not perfectly suited to the sort of emergentism that Bostrom and Sandberg use in their roadmap. Hence, pursuing the development of whole brain emulation might provide evidence indicating whether and how our minds relate to our brains.

4. Mental waves and pulses: analog vs. digital computation

In the recent past, many digital technologies have replaced analog ones, although a number of philosophers still argue for certain superiorities of analog computation. Digital, of course, uses discrete variables, such as our fingers or abacus beads, while analog's variables are continuous, as in the case of a slide rule. James Moor clarifies this distinction:

> in a digital computer information is represented by discrete elements and the computer progresses through a series of discrete states. In an analogue computer information is represented by continuous quantities and the computer processes information continuously. (Moor 1978, 217)

One notable advantage of analog is its "density" (Goodman 1968, 160-161). Between any two variables can be found another, but digital variables will always have gaps between them. For this reason, analog can compute an infinity of different values found within a finite range, while digital will always be missing variables between its units. In fact, Hava Siegelmann argues that analog is capable of a hyper-computation that no digital computer could possibly accomplish (Siegelmann 2003, 109).

Our emulated brain will receive simulated sense-signals. Does it matter if they are digital signals rather than analog? Many audiophiles swear by the unsurpassable superiority of analog recordings. Analog might be less precise, but it always flows like natural sound waves. Digital, even as it becomes more accurate, still sounds artificial and unnatural to them. In other words, there might be a qualitative difference to how we experience analog and digital stimuli, even though it might take a person with extra sensitivities to bring this difference to explicit awareness. Also, if the continuous and discrete are so fundamentally different, then maybe a brain computing in analog would experience a qualitatively different feel of consciousness than if the brain were instead computing in digital.

A "relevant property" of an audiophile's brain is its ability to discern analog from digital, and prefer one to the other. However, a digital emulation of the audiophile's brain might not be able to share its appreciation for analog, and also, perhaps digital emulations might even produce a mental awareness quite foreign to what humans normally experience. Bostrom's and Sandberg's brain emulation exclusively uses digital computation. Yet, they acknowledge that some argue analog and digital are qualitatively different, and they even admit that implementing analog in brain emulation could present profound difficulties (Bostrom and Sandberg 2008, 39). Nonetheless, they think there is no need to worry.

They first argue that brains are made of discrete atoms that must obey quantum mechanical rules, which force the atoms into discrete energy states. Moreover, these states could be limited by a discrete time-space (Bostrom and Sandberg 2008, 38). Although I am unable to comment on issues of quantum physics, let's presume for argument's sake that the world is fundamentally made-up of discrete parts. Bostrom and Sandberg also say that whole brain emulation's development would be profoundly hindered if quantum computation were needed to compute such incredibly tiny variations (Bostrom and Sandberg 2008, 39); however, this is where analog now already has the edge (Siegelmann 2003, 111).

Yet their next argument calls even that notion into question. They pose what is called "the argument from noise." Analog devices always take some physical form, and it is unavoidable that interferences and irregularities, called noise, will make the analog device imprecise. So analog might be capable of taking on an infinite range of variations; however, it will never be absolutely accurate, because noise always causes it to veer-off slightly from where it should be. Yet, digital has its own inaccuracies, because it is always missing variables between its discrete values. Nonetheless, digital is improving: little-by-little it is coming to handle more variables, and so it is filling in the gaps. Yet, digital will never be completely dense like analog, because values will always slip through its "fingers," so to speak. Analog's problem is that it will necessarily miss its mark to some degree. However, soon the magnitude between digital's smallest values will equal the magnitude that analog veers away from its proper course. Digital's blind spots would then be no greater than analog's smallest inaccuracies. So, we only need to wait for digital technology to improve enough so that it can compute the same values with equivalent precision. Both will be equally inaccurate, but for fundamentally different reasons.

Yet perhaps the argument from noise reduces the analog/digital distinction to a quantitative difference rather than a qualitative one, and analog is so prevalent in neural functioning that we should not so quickly brush it off. Note first that our nervous system's electrical signals are discrete pulses, like Morse code. In that sense they are digital. However, the frequency of the pulses can vary continuously (Jackendoff 1987, 33); for, the interval between two impulses may take any value (Müller et al. 1995, 5). This applies as well to our sense signals: as the stimulus varies continuously, the signal's frequency and voltage changes proportionally (Marieb and Hoehn 2007, 401). As well, there are many other neural quantities that are analog in this way.

Recent research suggests that the signal's amplitude is also graded and hence is analog (McCormick et.al 2006, 761). Also consider that our brains learn by adjusting the "weight" or computational significance of certain signal channels. A neuron's signal-inputs are summed, and when it reaches a specific threshold, the neuron fires its own signal, which then travels to other neurons where the process is repeated. Another way the neurons adapt is by altering this input threshold. Both these adjustments may take on a continuous range of values; hence analog computation seems fundamental to learning (Mead 1989, 353-54).

Fred Dretske gives reason to believe that our memories store information in analog. We may watch the setting sun and observe intently as it finally passes below the horizon. Yet we do not know *that* the sun has set until we convert the fluid continuum of sense impressions into concepts. These are discrete units of information, and thus they are digital (Dretske 1981, 142). However, we might later find ourselves in a situation where it is relevant to determine what we were doing *just before* the sun completely set. To make this assessment, we would need to recall our experience of the event and re-adjust our sensitivities for a new determination:

> as the needs, purposes, and circumstances of an organism change, it becomes necessary to alter the characteristics of the digital converter so as to exploit *more*, or *different*, pieces of information embedded in the sensory structures. (Dretske 1981, 143)

So in other words, because we can always go back into our memories to make more and more precise determinations, we must somehow be recording sense data in analog.

Bostrom and Sandberg make another computational assumption: they claim that no matter what the brain computes, a digital (Turing) computer could theoretically accomplish the same operation (Bostrom and Sandberg 2008, 7). However, note that we are emulating the brain's dynamics, and according to Terence Horgan, such dynamic systems use "continuous mathematics rather than discrete" (Horgan quoted in Schonbein 2005, 60). It is for this reason that Whit Schonbein claims analog neural networks would have more computational power than digital computers (Schonbein 2005, 61). Continuous systems have "infinitely precise values" that can "differ by an arbitrarily small degree," and yet like Bostrom and Sandberg, Schonbein critiques analog using the argument from noise (Schonbein 2005, 60). He says that analog computers are more powerful only in theory, but as soon as we build them, noise from the physical environment diminishes their accuracy (Schonbein 2005, 65-66). Curiously, he concludes that we should not for that reason dismiss analog but instead claims that analog neural networks, "while not offering greater computational power, may nonetheless offer something else" (2005, 68). However, he leaves it for another effort to say exactly what might be the unique value of analog computation.

A.F. Murray's research on neural-network learning supplies an answer: analog noise interference is significantly more effective than digital at aiding adaptation, because being "wrong" allows neurons to explore new possibilities for weights and connections (Murray 1991, 1547). This enables us to learn and adapt to a chaotically changing environment. So using digitally-simulated neural noise might be inadequate. Analog is better, because it affords our neurons an infinite array of alternate configurations (1991, 1547-1548). Hence in response to Bostrom's and Sandberg's argument from noise, I propose this argument *for* noise. Analog's inaccuracies take the form of continuous variation, and in my view, this is precisely what makes it necessary for whole brain emulation.

5. Even while men's minds are wild?

Neural noise can result from external interferences like magnetic fields or from internal random fluctuations (Ward 2002, 116-117). According to Steven Rose, our brain is an "uncertain" system on account of "random, indeterminate, and probabilistic" events that are essential to its functioning (Rose 1976, 93). Alex Pouget and his research team recently found that the mind's ability to compute complex calculations has much to do with its noise. Our neurons transmit varying signal-patterns even for the same stimulus (Pouget et al. 2006, 356), which allows us to probabilistically estimate margins of error when making split-second decisions, as for example when deciding what to do if our brakes fail as we speed toward a busy intersection (Pouget et al. 2008, 1142). Hence the brain's noisy irregularities seem to be one reason that it is such a powerful and effective computer.

Some also theorize that noise is essential to the human brain's creativity. Johnson-Laird claims that creative mental processes are never predictable (Johnson-Laird 1987, 256). On this basis, he suggests that one way to make computers think creatively would be to have them alter their own functioning by submitting their own programs to artificially-generated random variations (Johnson-Laird 1993, 119-120). This would produce what Ben Goertzel refers to as "a complex combination of random chance with strict, deterministic rules" (Goertzel 1994, 119). According to Daniel Dennett, this indeterminism is precisely what endows us with what we call free will (Dartnall 1994, 37). Likewise, Bostrom and Sandberg suggest we introduce random noise into our emulation by using pseudo-random number generators. They are not truly random, because eventually the pattern will repeat. However, if it takes a very long time before the repetitions appear, then probably it would be sufficiently close to real randomness.

Yet perhaps there is more to consider. Lawrence Ward reviews findings that suggest we may characterize our neural irregularities as *pink noise*, which is also called *1/f noise* (Ward 2002, 145-153). Benoit Mandelbrot classifies such *1/f* noise as what he terms "wild randomness" and "wild variation" (Mandelbrot and Hudson 2004, 39-41). This sort of random might not be so easily simulated, and Mandelbrot gives two reasons for this. 1) In wild randomness, there are events that defy the normal random distribution of the bell curve. He cites a number of stock market events that are astronomically improbable, even though such occurrences in fact happen quite frequently in natural systems despite their seeming impossibility. There is no way to predict when they will happen or how drastic they will be (Mandelbrot and Hudson 2004, 4). And 2), each event is random and yet it is not independent from the rest, like each toss of a coin is. One seemingly small anomalous event will echo like reverberations at unpredictable intervals into the future (Mandelbrot and Hudson 2004, 181-185). For these reasons, he considers wild variation to be a state of indeterminism that is *qualitatively* different than the usual mild variations we encounter at the casino; for, there is infinite variance in the distributions of wild randomness. Anything can happen at any time and to any degree of severity, so this sort of random might not be so easily emulated (Mandelbrot 1997, 128). In *The (mis)behavior of markets* Mandelbrot and Hudson write, "the fluctuation from one value to the next is limitless and frightening" (2004, 39-41). This is the wildness of our brains.

Yet let's suppose that the brain's wild randomness can be adequately emulated. Will whole brain emulation still attain its fullest success of perfectly replicating a specific person's own identity? Bostrom and Sandberg recognize that neural noise will prevent precise one-to-one emulation; however, they think that the noise will not prevent the emulation from producing meaningful brain states (Bostrom and Sandberg 2008, 7).

To pursue further the personal identity question, let's imagine that we want to emulate a certain casino slot machine. A relevant property is its unpredictability, so do we want the emulation and the original to both give consistently the same outcomes? That would happen if we precisely duplicate all the original’s relevant *physical* properties. Yet, what about its essential unpredictability? The physically accurate reproduction could predict in advance all the original’s forthcoming read-outs. Or instead, would a more faithful copy of the original produce its own distinct set of unpredictable outcomes? Then we would be replicating the original’s most important relevant property of being governed by chance.

The problem is that the brain’s *1/f* noise is *wildly* random. So suppose we emulate some person’s brain perfectly, and suppose further that the original person and her emulation identify so much that they cannot distinguish themselves from one another. Yet, if both minds are subject to wild variations, then their consciousness and identity might come to differ more than just slightly. They could even veer off wildly.

So, to successfully emulate a brain, we might need to emulate this wild neural randomness. However, that seems to remove the possibility that the emulation will continue on as the original person. Perhaps our very effort to emulate a specific human brain results in our producing an entirely different person altogether.

Conclusion

Whether this technology succeeds or fails, it can still advance a number of philosophical debates. It could suggest to us if our minds emerge from our brains, or if the philosophy of artificial intelligence should consider analog computation more seriously. Moreover, we might learn whether our brain’s randomness is responsible for creativity, adaptation, and free choice, and if this randomness is the reason our personal identities cannot be duplicated. If in the end we see that it is bound to fail, we might learn what makes our human minds unlike computers. Yet if it succeeds, would this not mean that our minds could in fact survive the deaths of our physical bodies, and might not we be able to create new human minds by merely writing a program for a new person?

Acknowledgments

I wish to thank Russell Blackford, Ullrich Melle, and Aziz Zambak for their contributions to this paper.

Bibliography

Bainbridge, W. 2007. Converging technologies and human destiny. *The Journal of Medicine and Philosophy* 32(3): 197-216.

Bostrom, N., and A. Sandberg. 2008. Whole brain emulation: A roadmap. *Technical Report* #2008 ‐ 3, Future of Humanity Institute, Oxford University.

Chalmers, D. 1996. *The conscious mind: In search of a fundamental theory.* Oxford: Oxford University Press.

Clayton, P. 2006. Conceptual foundations of emergence theory. In *The re-emergence of emergence: The emergentist hypothesis from science to religion*, ed. P. Clayton and P. Davies, 1-34. Oxford: Oxford University Press.

Dartnall, T. 1994. Introduction: On having a mind of your own. In *Artificial intelligence and creativity: An interdisciplinary approach*, ed. T. Dartnall, 29-42. Dordrecht: Kluwer Academic Publishers.

Dretske, F. 1981. *Knowledge and the flow of information*. Cambridge: MIT Press.

Feinberg, T. 2001. Why the mind is not a radically emergent feature of the brain. In *The emergence of consciousness*, ed. A. Freeman, 123-46. Thorverton, United Kingdom: Imprint Academic.

Goertzel, B. 1994. *Chaotic logic: Language, thought, and reality from the perspective of complex systems science*. London: Plenum Press.

Goodman, N. 1968. *Languages of art: An approach to a theory of symbols*. New York: The Bobbs-Merrill Company.

Hasker, W. 1999. *The emergent self*. London: Cornell University Press.

Hayles, N. K. 1999. *How we became posthuman: Virtual bodies in cybernetics, literature, and informatics*. Chicago: University of Chicago Press.

Humphreys, P. 2008. Synchronic and diachronic emergence. *Minds and Machines* 18(4): 431-42.

Jackendoff, R. 1987. *Consciousness and the computational mind*. London: MIT Press.

Johnson-Laird, P. 1988. *The computer and the mind: An introduction to cognitive science*. Cambridge: Harvard University Press.

Johnson-Laird, P. 1993. *Human and machine thinking*. London: Lawrence Erlbaum Associates.

Krueger, O. 2005. Gnosis in cyberspace? Body, mind and progress in posthumanism. *Journal of Evolution and Technology* 14(2): 77-89.

Lycan, W. 2009. Giving dualism its due. *Australasian Journal of Philosophy* 87(4): 551-63.

Mandelbrot, B. 1997. *Fractals and scaling in finance: Discontinuity, concentration, risk*. Berlin: Springer.

Mandelbrot, B., and R. Hudson. 2004. *The (mis)behavior of markets: A fractal view of risk, ruin, and reward*. New York: Basic Books.

Marieb, E., and K. Hoehn. 2007. *Human anatomy and physiology*. London: Pearson.

McCormick, D., Y. Shu, A. Hasenstaub, A. Duque, and Y. Yuguo. 2006. Modulation of intracortical synaptic potentials by presynaptic somatic membrane potential. *Nature* 441: 761-65.

Mead, C. 1989. *Analog VLSI and neural systems*. Amsterdam: Addison-Wesley Publishing Company.

Minsky, M. 1972. *Computation: Finite and infinite machines*. London: Prentice-Hall International, Inc.

Moor, J. 1978. Three myths of computer science. *The British Journal for the Philosophy of Science* 23(3): 213-222.

Moravec, H. 1988. *Mind children: The future of robot and human intelligence*. Cambridge: Harvard University Press.

Müller, B., J. Reinhardt, and M. Strickland. 1995. *Neural networks: An introduction*. Berlin: Springer.

Murray, A. 1991. Analogue noise-enhanced learning in neural network circuits. *Electronics Letters* 27(17): 1546-1546.

O'Connor, T., and H. Wong. 2006. Emergent properties. In *The Stanford encyclopedia of philosophy (winter 2006 edition)*, ed. E.N. Zalta. http://plato.stanford.edu/entries/properties-emergent/ (accessed 04-October-2010).

Pouget, A., P. Latham, and B. Averbeck. 2006. Neural correlations, population coding and computation. *Nature Review Neuroscience* 7: 358-366.

Pouget, A., J. Beck, W. Ma, R. Kiani, T. Hanks, A. Churchland, J. Roitman, M. Shadlen, and P. Latham. 2008. Probabilistic population codes for Bayesian decision making. *Neuron* 60(6): 1142-52.

Putnam, H. 1975. *Mind, language, and reality*. Cambridge: Cambridge University Press.

Rose, S. 1976. *The conscious brain*. Harmondsworth, Middlesex: Penguin Books.

Schonbein, W. 2005. Cognition and the power of continuous dynamical systems. *Mind and Machines* 15(1): 57-71.

Searle, J. 1992. *The rediscovery of the mind.* Cambridge, Massachusetts: The MIT Press.

Siegelmann, H. 2003. Neural and super-Turing computing. *Minds and Machines* 13(1): 103-114.

Thompson, E., F. Varela, and E. Rosch. 1991. *The embodied mind: Cognitive science and human experience*. Cambridge, Massachusetts: The MIT Press.

Ward, L. 2002. *Dynamical cognitive science*. London: MIT Press.