The Computer Never Was a Brain, or the Curious Death and Designs of John von Neumann

Benjamin Peters

> And it wasn't functioning anymore.
> *Edward Teller on John von Neumann's brain*

Even when it was most tempting, John von Neumann resisted the neuro-hubris of the computer-brain analogy he helped create. Indeed, in his deathbed lectures *The Computer & the Brain*, he grants certain aspects of the analogy "absolute implausibility".[1] The distance between neuroscience and computer science has grown more obvious since he was writing in 1956: nevertheless, the ripple effects of that troublesome analogy live on in not only the current moment of so-called "smart machines" but in modern approaches to designing not only the behavior of human cognition, but life and death itself, into machinery. This essay explores one moment in how twentieth-century information scientists and technologists came to situate designs on human behavior in computing, cognition, and communication discourse, and in the process encountered the limits of those design in the computer-brain analogy.

The literate world has endured, especially since the postwar period, a steady and ever increasing stream of articles, books, and pundits declaring the coming convergence of mind and machine in the age of artificial superintelligence.[2] Central to such futurism, of course, is the analogy between the computer and the brain collapsing into a singular reality. The details of these projected realities of such a computer-brain merger are, of course, many and diverse: Hermann von Helmholtz envisioned a nervous system in the transmission logics of telegraph networks of the late nineteenth century while commentators in the early twenty-first see democratic intelligence leavening

1 | John von Neumann: *The Computer & The Brain* [1958], Yale 2012, 72.

2 | Nick Bolstrom: *Superintelligence: Paths, Dangers, Strategies*, New York 2016.

crypto-currencies and blockchain, a distributed ledger central to many current ›smart‹ data imaginaries.[3] Throughout the annals of the twentieth-century age that connects these two examples, the dual images of the digital computer as a brain and the brain as a digital computer have proven hardy perennials for imagining a new class of superintelligent creatures—or, as the futurist Vernor Vinge wistfully put it, "the eminent creation of entities by technology of greater than human intelligence."[4]

Of course, the history of human attempts to mechanize the mind—and in particular to proffer mechanical designs for cognitive behavior—long predate the somewhat narrow history of artificial intelligence springing from von Neumann, Allen Newell, and Herbert Simon in the 1950s. To begin a potted history with a relatively late example, René Descartes held up the regularity of mechanical clocks and automata that moved and sang for the aristocratic seventeenth-century Europe as models of psychological capacity—indeed, the fact that he could reject these devices as ›mere‹ mechanism suggests, at a basic level, the emergence of the modern naturalist philosophical perception, soon championed by Thomas Hobbes *against* Descartes, that cognition and mechanical designs might share, in a prescient seeding of the Darwinian revolution to come, an irreducibly materialist nature; whereas Descartes sought to police the boundaries of humanity more closely, for Hobbes, all intelligence—animal, mechanical, social—is the product of mechanisms, and the mind is matter fit for thinking. (Perhaps the distance between the monarchy in his *Leviathan* and the Colossus computer of Bletchey Park is not so great.)

The subsequent two centuries saw the construction of mechanical automata such as chess-playing Turks, defecating ducks, and mechanical accounting systems, which in turn made for the slow interweaving of cognition, computation, and behavioral design across a range of innovations in the nineteenth century such as the programmable Jacquard looms, Charles Babbage's (never completed) Analytical Engine, (George) Boolean logic, the surgeon Alfred Smee's experiments on deducing instinct and reason from what he called "electrobiology," and the development of communication engineering and communication theory in the wake of the global spread of the electronic telegraph; these intellectual threads would converge such that, in 1887, *the American Journal of Psychology* saw no contradiction in Charles Sanders Peirce's speculations about "logical machines."[5] Suffice it to note that each of

3 | Timothy Lenoir: *"Helmholtz and the Materialities of Communication"*, in: Osiris 9 (1994), 184-207.

4 | Vernor Vinge: *"The Technological Singularity"*, *1993*, at https://www.frc.ri.cmu.edu/~hpm/book98/com.ch1/vinge.singularity.html (accessed 29.09.2017).

5 | Charles Sanders Peirce: *"Logical Machines"*, in: The American Journal of Psychology I (1887), 125-170.

these actors appears to have understood something that is all too often lost in contemporary debates about computing and the mind: namely, that artificial intelligence has long been and, even in the current moment of machine learning breakthroughs, largely remains fundamentally a *philosophy*—and a humane philosophy at that. For that reason, it is not a mistake, I argue, to welcome in the history of mind-computer analogies a broader family reunion of natural philosophers, political thinkers, and ethicists preoccupied with the question of designing human cognition. In particular, what can we learn from how humans, mortal creatures that we are, have confronted the limits of human cognition in designing the very automation set to overcome those limits?

The case of John von Neumann stands out in the mid-twentieth-century history of cybernetics, the mechanization of mind, and artificial intelligence. A Hungarian émigré and polymathic mathematician to the United States, von Neumann—whose work inspired his closest friend the mathematician Stanislaw Ulam (1958) to coin the term ›singularity‹ referring to technological change—is known for nearly continuous feats in mathematics, quantum physics, economics (co-inventing game theory), statistics, and computing, including designing the principal architecture for modern computing, especially the ENIAC.[6] No asocial outcast or isolated brain, von Neumann lived life to its earthy fullest, wearing the finest three-piece suits, recounting off-color jokes, and indulging hospitality, cordiality, and networks of power wherever he could. Such a *bon vivant* was capable of counting everything but calories, as his second spouse Klara von Neumann once quipped. He was also intimately involved in America's entrance into the nuclear and computer age, given that he served as a visiting consultant to the Manhattan Project and helped devise the implosion device for the first Trinity nuclear detonation and later the devastation wrecked upon Nagasaki.[7] It was likely in this very work designing the mechanisms of mega-death that he contracted the cancer that ended his life in 1957 at the age of 53.

Rarely more than one step removed from the legend and lore of the American superpower science, von Neumann's final published work brought together the issues of computers, minds, and (given his irradiated, failing health) mechanized death in the prestigious Silliman Lectures that he gave at Yale University in 1956, which was posthumously published in 1958 as the unfinished book *The Computer & the Brain*. This work, in turn, stimulated

6 | Stanislaw Ulam: "Tribute to John von Neumann", in: *Bulletin of the American Mathematical Society* 64/3/2 (May 1958), 5.

7 | Cf. Steve J. Heims: *John von Neumann and Norbert Wiener, from Mathematics to the Technologies of Life and Death,* Cambridge/MA 1980; Giorgio Israel, Ana Millan Gasca: *The World as a Mathematical Game: John von Neumann and Twentieth Century Science,* Basel 2009.

interest and imagination for artificial immortality in coming generations, while also drawing on the mathematical foundations laid by other scientists turning toward cognitive science, such as Warren McCullough's work on logic in neural networks, Norbert Wiener's 1948 *Cybernetics*, George Miller's contention in 1951 *Language & Communication* that language could be studied through information theory, and especially W. Ross Ashby's *Design for a Brain* as perhaps the behaviorist consolidation of adaptive behavior in the computer and the brain. Yielding slowly to the malignant bone cancer spreading from his left shoulder through his body in 1956, von Neumann extended his mental faculties in these lectures toward its own general problem: how might one extend the mind itself into the logical operations of the digital computers he designed?

Given the obvious urge to transcend death, we may encounter the temptation to repeat the line that perhaps "Johnny von Neumann, who knew so well how to live, did not know how to die".[8] The instinct to fault a human, on the brink of death, for pursuing artificial extensions to life is only part right, however—and in fact his text, despite being interpreted otherwise, complicates generations of subsequent naïve transhumanist readings and its critiques. Perhaps von Neumann did know how to die—perhaps not. What rings obvious from his lecturers is that, for him, never did the digital computer appear that much like a brain, nor the brain like a computer, despite his best attempts to analogize the two. Von Neumann, the same man often credited with sparking interest in the ›electronic brains‹ revolution, could not, in the end, bring himself to believe that the comparison between digital computers and human brains was ever much more than that—a comparison. The subsequent popularity of the analogy in the subsequent generation of artificial intelligence research as well as modern-day tech talk has not sufficiently reflected the fact that, in this deathbed text at least, the computer-brain analogy was built to be broken.

This essay, which summarizes his lecturers in context of his death, lays out historical and theoretical commentary on why the computer-brain analogy has proven as resilient as it has misleading. In particular, it is precisely the felt mismatch in von Neumann's book between, on the one hand, his impulse to couple the computer and the brain and, on the other, his faithfulness to reason and evidence that renders him incapable of actually doing so that stands as a moving witness to the intertwined character of the transhumanist instinct to transcend death through technology as well as its antithesis, a thoroughgoing grounding in his own mortal condition. In other words, von Neumann sought a species-transcending union of computer and mind at the very moment his cancerous body could no longer sustain his mind, but it was precisely the limits of his mind and body that led him to assert no more than the flimsiest bridge

8 | MacRae: *John von Neumann*, 378.

between brain and computer, and thus, in the process, to conclude his own life of the mind.

Rereading the Text: Reinserting the "&" in von Neumann's *The Computer & The Brain*

Throughout this text *The Computer & the Brain,* von Neumann seemingly *wants* but cannot bring himself to analogize the computer and the brain—and in some ways it is the former fact that proves most interesting.[9] For example, he concludes a section exploring how precise a brain must be "when looking at the nervous system as at a computing machine" (he notes that, due to the complex networks in the brain, neural computation must be far more precise than a computer—perhaps 12 decimals of precision deep) by admitting that "this conclusion was well worth working out just because of, rather than in spite of, its absolute implausibility" (76). In the phrase "just because of" he admits that his conclusion—the brain, when viewed as a computer, must be far more precise than a computer—is nonsense, yet nevertheless some kind of *worthwhile* nonsense. Why, for von Neumann, is this and other "absolutely implausible" conclusions worth entertaining?

The whole book—a short 83 pages in its third edition (plus 51 pages of combined forewords from Klara von Neumann, Paul and Patricia Churchland, and Ray Kurzweil)—is an experiment upon the grand cybernetic analogy between biological and mechanical computational information systems. Some, but not all, of his interpreters accommodate this reading. In the 1957 edition, Klara von Neumann lays out the backstory of the lecturers writing and his death without commenting on the computer-brain analogy; in the 2000 edition, Paul and Patricia Church hedge their bets and stay ambiguous about the end analogy, while in the 2012 edition of the book, Ray Kurzweil reverses courses by widely crediting the text with naming the brain a digital machine, a year before publishing his own popular and speculative book *How to Create a Mind: The Secrets of Human Thought Revealed* (2013), which advances brain-computer parallels of mental data processing and brain algorithms to its full extent. Perhaps it was a stretch too far for Kurzweil to imagine the most famous brain of the twentieth-century pontificating about the *limits* of the brain, although he at least hedges that "there are very few discussions in this book that I find to be at significant odds with what we now understand." Indeed, despite Kurzweil's claims otherwise elsewhere, the last three generations of neuroscience have revealed, as von Neumann thought, just how unlike the two are. Given how limited our understanding of the brain, we may now see more clearly what

9 | Neumann, von: *The Computer & The Brain*.

von Neumann understood—the computer-brain analogy was broken before it was built. The most significant and neglected part in his famous title, *The Computer & the Brain*, in other words, is the "&"—the other two terms are not even approximate equivalents.

Indeed, in the text, von Neumann never concludes that the brain resembles a digital computer except in the most surface ways, and much less so does he argue that a digital computer is brain-like. In fact, von Neumann spends a clear majority of the short text *distinguishing* between the brain and the computer. His list of dissimilarities runs long: the computer processes serially and with high precision, while the brain does so in parallel and with low precision; the average neuron is exponentially smaller, slower, and more complex in its connections than the average transistor; the basic operations of computer and neuronal memory share a common vocabulary (for example, both have storage capacities embodied in automatic networks of active elements, whether nerve cells or "flip-flop" vacuum-tubes pairs), but little else; and, as he concludes, the brain and the computer presume wholly separate operating languages.

His hypothesis for testing, however, is the opposite—to claim that the brain's "functioning is *prima facie* digital." This phrase *prima facie*—or that which appears true at first glance until later proven incorrect—crisply illustrates the distinction between desire to believe and the shortage of evidence for supporting the brain as anything more than, in a choice phrase, a superficially "digital organ." What was functionally digital then? Only the comparison between most elemental unit of neurons and transistors: every neuron has, von Neumann writes, "an essentially reproducible, unitary response to a rather wide variety of stimuli" (4)—and it was this unitary response, either a neuron will or will not emit a pulse, that "is clearly the description of the functioning of an organ in a digital machine" (43). This fact alone "therefore justifies the original assertion that the nervous system has a *prima facie* digital character."

His argument is structured around exploring the limits of this primary claim, which recognizes that the claim of the brain's *prima facie* digital quality involves "some idealizations and simplifications," a phrase he repeats several times throughout the text. Indeed, in light of these dissimilarities, he admits, "the digital character no longer stands out so clearly and unequivocally" (43-44). Later he notes that "neurons appear, when thus viewed, as the basic logical organs—and hence also as the basic digital organs." By this he seems to suggest that the on-off character of a neuron permits, in theory, their organization into the equivalent of "and" and "or" and "not" logic gates. However, as soon as he notes the theoretical possibility and its much-heralded consequence—that the brain could be the equivalent of a neural Turing machine (among other less idealized meetings of Claude Shannon's master's thesis with Warren McCulloch's work on neuronal networks)—he also immediately sets it aside. Instead of pursuing idealized computational theories, he describes actually-

existing biological complications in neural networks, noting that, depending on location of the connection along the neuron, the resulting pathways in the brain may produce "even more complicated quantitative and geometrical relationships that might be relevant" (55). Subsequent research into the neuronal networks have confirmed the extraordinarily complex and, in many ways, still opaque nature of mental gray matter.

He further compares these two most elemental units behind, in his paired phrasing, these "natural and artificial organs" or "artificial and natural automatons": namely, he evaluates the neuron and the transistor according to their speed, size, and organization of their discrete actions, with the preferred medium being (in an anticipation of countless media and communication advertisements) smaller, faster, and more immediate. His calculations conclude that the neuron is about a hundred million to a billion times smaller as well as over a thousand times (he estimates a factor of to) slower than the transistor. Anticipating that the transistor would shrink in subsequent generations (as it has), even the promise of the future (and our present) enjoying a potentially compatible scale of neuron-transistor does not distract von Neumann from noting that, even then, organic neurons and mechanical transistors are likely to be organized and structured very differently: "large and efficient natural automata are likely to be highly parallel, while large and efficient artificial automata [… are likely to be] serial" (51). This conclusion reaffirms the serial design of computer processors built into von Neumann computer architecture that routes one bit of data through the bus one at a time, while also licensing almost an endless number of possible alternative structures of neural networks. The brain and the computer, he insists, are almost sure to be organized differently: "Hence the logical approach and structure in natural automata may be expected to differ widely from those in artificial automata" (52); as a result, digital computers will have "systematically more severe" memory requirements than brains.

In other words, if one embraces the tenuous and superficial digital similarity between the neuron and the transistor (which von Neumann both courts and resists), one must also adopt a much broader definition of the term: indeed, the term *digital* would seem to describe any operation that introduces into a system a binary threshold or on-off mechanism. This position, which von Neumann first established at the first Macy Conference on Cybernetics in 1946, ushers in a previously neglected mind-bindingly broad class of other superficially digital techniques and sign practices: not only is the neuron *prima facie* digital, so too, by his definition, is the push button, bureaucrat's stamp, the flip of a coin, the strike of a drum, early lithic blades, and perhaps the horizon between heaven

and earth itself.[10] Each operation introduces a discrete on-off threshold; either the button is pushed, the paper stamped, the coin faces heads up, the drum struck, the material cut, the environment earthy—or it is not. To be digital, perhaps, is simply to be or not to be in a certain state. Even if one rejects this speculative vein of thinking (which I will set aside for another time), it is not clear how one could anoint the brain a holy digital organ without also naming many other systems—such as peripheral nervous systems in animals—*prima facie* digital.

In what follows I build on this disputable digitality of the brain in order to reason that von Neumann—and so many before and since—gravitated to the digital brain-computer comparison, and not so many others, not because of any native operational similarities between a brain and a digital computer. Rather he offered the analogy for the romance of the comparison that reveals *post facto* a felt human instinct animating so much of digital technological discourse—namely the deep-seated desire to transform with cutting-edge technology the seat of the self, imagined by moderns to be the brain, into a kind of throne of information immortality.

The Final Confession of von Neumann

In 1956, the same year that von Neumann was invited to give the prestigious Silliman lecturers at Yale University, personal tragedy struck: von Neumann developed a malignant and ultimately fatal cancer likely sped by his exposure to irradiated materials that he encountered in his role consulting at the Manhattan Project. According to Klara von Neumann, even as his deteriorating health forced him, in 1956, to abandon other major responsibilities—including his recent appointment on a series of councils and commissions in DC—he nevertheless insisted on working on these lectures to the end, even though his infirmities kept him from delivering them in full (xlv-li). Edward Teller, ›father of the hydrogen bomb‹ reflected on von Neumann's passing. Teller stressed that the decline of a great brain might cause it suffering greater than all others. Indeed, the statement that "I think that he suffered from this loss more than I have seen any human to suffer in any other circumstance" is made all the more incredible by the fact it comes from Teller, an outspoken political advocate for nuclear bombs, even in times of peace—presumably the very person likely to have had the most cause to consider the scale and scope of human suffering. Teller's reversion, in his most intimate moment, to the functionalist language

10 | Claus Pias (ed.): *Cybernetics: The Macy Conferences 1946-1953*, Berlin/Zürich 2003, 60-68.

of the brain as "it wasn't functioning anymore", revealingly operationalizes intimate and tragic moments of human lives.

Whatever else one may say, it is clear that von Neumann did not decline peacefully. This desperate situation—crippled by pain and agonizing as he watched his mind, perhaps for the first time, begin to slip from his control—underscores just how significant it is, again, that von Neumann turned to offer, as best he could, a final (and in some ways first) argument about the brain and computer analogy. Indeed, the computer-brain analogy may be said to stand as von Neumann's final temptation, sacrament, and confession: perhaps no other choice would be more likely to reach beyond the living limits of mental work than von Neumann's choice to write on computers and brains approaching his own death. Von Neumann both held to and pressed beyond the limits of his secular worldview in his final days, calling for a Catholic Priest, taking his last sacraments while recounting both Pascal's wager and old Latin prayers from memory.[11] That the promise of Catholicism, alongside the computer-brain afterlife imaginary (e.g., that the computer might be enough like the brain to extend the activities of one's neuronal nets) moved a man both haunted and poisoned by the tools of nuclear mega-death that he helped develop is telling; however, more moving still is that, even in his deathbed witness, he maintained that his own most tempting grand cybernetic analogy was still, on balance, more wrong than right.

We cannot be sure why the text does not more either complete or critique the analogy. It could be because his faith in the metaphor failed. It could be because the body behind his mind failed. Perhaps it is—or even their merger into the same point: the brain is not like the digital computer except in the most superficial sense. Perhaps it was his increasingly crippled condition, (one colleague noted that "we could never keep up with the speed of his speaking until sadly in the last year in the hospital, we could") then that, in a final act of unintended grace, kept him from having to spell out, in writing, his final judgment on the commensurability of the brain and the computer.[12] Mortality—his most uncomputer-like condition—saves him from having to conclude his final confession.

Conclusion

At a glance, it seems perfectly evident why a leading polymath, on the brink of collapse, might experiment upon a brighter technological future in which the mind might be freed from the mortal flesh, indeed precisely the excruciating

11 | MacRae: *John von Neumann,* 377-380.

12 | Ibid., 378.

pain and death he was suffering from the bone cancer he likely contracted while working with nuclear materials in the Manhattan Project during World War II. His deathbed lecturers—partial and short as they are—also suggest his iconic and complicated embodiment in the time and place of the early nuclear age. His death, as his life, witnesses not only humanity's instinct for intellectual transcendence but also the wretched blessedness of its embodied finitude. His death signals a stirring reminder that even the most "open minds" in the midcentury were willing to calculate mega-deaths of others without ever fully reconciling with one's own.[13] His final text at once seeks, refuses, and ultimately falls short in analogizing the superhuman union of the computer and the brain.

Let us, with von Neumann's resistance to his own analogy in mind, diagnose and treat the strain of neuro-hubris in contemporary computing thought. To be clear, I see no reason to be against the brain-computer analogy as an analogy itself. There's *ipso facto* nothing wrong with analogizing the brain with the most recent technology, provided we acknowledge it is bound to get the technology wrong. The spirit of Hebrew clay, the Roman aqueduct, the hydraulics of the humors (and its eventual Cartesian pump), the medieval catapult, Freud's steam engine, Helmholtz' telegraph, and today the holograph, among a host of other new media, have all been compared to the brain and its neural system. (New media cannot help but evoke old concerns.) Still, a simple glance at brains and computers should suggest the literal and lexical comparison to be nonsense: mammalian brains—equipped with maternal, sound, smell, sight, touch, and other reflexive instincts from birth—share little in common with the rules, lexicons, algorithms, memories, subroutines, decoders, processors, and buffers that run computers. Even on its own terms, the computer remains indisputably unlike the brain. We may even put a twist onto Vannevar Bush's famous 1945 essay, in which he envisions the potential of computing memory to do so much more than destroy: perhaps the most sustainable model of computing, both past and future, lies in abandoning models "as we may think."

Instead, we may now set aside the computer-brain analogy, and with it refresh scrutiny of its larger frames: the evolutionary analogy that humans are animals and the cybernetic analogy that humans are machines. Of course, in a strict sense, neither is wrong: humans are both animals and machines, yet none of this has clarified our ongoing search for ways to design the mechanisms of self-control—be it the psyche, the computer, or civil society: such designs for the devices for their control have variously been understood over the ages as embodied spirits and helmsmen, hydraulic and mechanical devices, electronic and optical metaphors, and—the coin of the current media environment—

13 | Jamie Cohen-Cole: *The Open Mind: Politics and the Science of Human Nature*, Chicago 2014.

cumputational mechanisms. That humans find so many analogies so fertile is itself a healthful reminder of where the computer falls far short of the human brain; perhaps the most human feature, as we see in von Neumann, is the capacity to see more than exists in the design of a machine; and *that* quality of interpretation—the very quality that appeals to the transhuman and to peek, with von Neumann, beyond the finitude of our beings—is ultimately precisely what makes us human.

Of course, none of this will keep visionary researchers and tech propagandists—from futurists like Kurzweil to physicist Stephen Hawkins to the philosopher Nicholas Bolstrom—from offering up new waves of analogous terms for understanding the design of smart machines: recent claims holds that the computer and the brain operate by shared "field-programmable gate arrays" and that neurons and computational primitives are bound to unite as the "software" of human consciousness is uploaded to superintelligent computer networks, a kind of digital deification. In so doing these claims will continue the transhumanist instinct, if not the conclusion, of von Neumann as well as one of the central tasks, and hubrises, of the information age—the convergence of computing, communication, and cognition designs.[14]

14 | My thanks to Mark Brewin, Lincoln Cannon, Joli Jensen, Tamara Kneese, and Christina Vagt for their helpful comments and criticisms.