DONKEY KONG'S LEGACY

About microprocessors as model organisms and the behavioral politics of video games in AI

Johannes Bruder

Abstract

The article discusses forms of contamination between human and artificial intelligence in computational neuroscience and machine learning research. I begin with a deep dive into an experiment with the legacy microprocessor MOS 6502, conducted by two engineers working in computational neuroscience, to explain why and how machine learning algorithms are increasingly employed to simulate human cognition and behavior. Through the strategic use of the microprocessor as "model organism" and references to biological and psychological lab research, the authors draw attention to speculative research in machine learning, where arcade video games designed in the 1980s provide test beds for artificial intelligences under development. I elaborate on the politics of these test beds and suggest alternative avenues for machine learning research to avoid that artificial intelligence merely reproduces settler-colonialist politics *in silico*.

Keywords: *artificial intelligence, machine learning, model organisms, lab studies, algorithm studies*

Introduction

Rose was one of the first researchers who I managed to engage in a longer conversation during my fieldwork in a British neuroscience institute in 2011. She was a very casual person, always chewing on a gum, an unperturbed look on her face, throwing truths around that I was not always prepared to hear. I had come here to observe those who study the human brain, yet what Rose told me is that many people in the lab I was visiting were "kind of largely outside the domain of understanding what the brain is doing. You're more in the domain of understanding what the signal tells you as a methodologist", she observed. "We tend to be more computational people".

As baffled as I was in the beginning, my conversation with Rose raised my interest in computer scientists' and engineers' perspectives of the human brain. I was intrigued by the question about what happens when people, who would normally design circuits or code algorithms study the human brain *by their means*. If microprocessors are substituted for human brains in experiments, if machine learning algorithms are used to simulate aspects of human cognition, how does that affect our understanding of cognition and intelligence?

This paper is a study of machine learning algorithms in practice. It differs from studies of algorithms *at work* (Kellogg, Valentine, and Christin 2020) and of algorithms *as* or *in culture* (Seaver 2017), in that I focus on "agents" in laboratories – experimental algorithms that learn to simulate aspects of human cognition and behavior to illuminate the ominous human capacity of intelligence, and to help researchers in building intelligent systems. These are experiments where the digital tries to pass as a version of the biological. This paper is hence more firmly embedded in traditional lab studies (Knorr Cetina 1992) than in the burgeoning field of algorithms studies, yet it draws on the methods of both fields to get a grip on experiments that shift the boundaries between, or allow for the mutual contamination of human and artificial intelligence.

An experiment conducted by two computational neuroscientists is exemplary in this regard. The experimenters substituted a microprocessor that once powered the Nintendo Entertainment System (NES) and the Apple II for the human brain and ran old video games as example "behaviors" to analyze how the computer "thinks". Their experiment was first and foremost meant and considered as epistemological critique: the authors provocatively ask why neuroscientists believe they could understand the human brain although the data analysis methods currently used in neuroscience cannot help elucidate the operations of the infinitely less complex MOS6502 chip?

I analyze how the experimenters selectively draw on laboratory experiments in biology to legitimate their decision to substitute a legacy microprocessor for human brains in scanners. Their argument supports a very specific analogy between brains and computers, which derives from mid-twentieth century attempts at modeling human decision-making on computers and suggests that anthropologists ought to study the scripts or protocols of simulations to get a grip on how cognition is reconceived in between human and machine.

At closer look, Jonas and Kording's study is not only epistemological critique; it provides an inroad to the use of video games as replacement laboratories in the study of cognition and intelligence. Jonas and Kording's choice of 1980s Atari video games like Donkey Kong as "naturalistic" behaviors is of particular interest, since it exemplifies a recent trend to consider these as "microcosms of the real world" (Markoff 2016). Against this backdrop, it is the video game as virtual laboratory or test bed that determines what counts as creative and intelligent in humans and machines.

Brains are not chips, but …

My encounter with Rose in 2011 only marked the beginning of an extended engagement with computational neuroscience. I did not immediately notice how closely related the resurgence of artificial neural networks and changing paradigms of computational neuroscience were. Initially, I was too intrigued by the fact that most of the researchers I interacted with in neuroscience laboratories were focused on data and – as one PostDoc in a Swiss neuroscience lab told me – thought of research on the brain as "an interesting application of maths". Some had just transferred from the field of security engineering, others had opted to analyze brain imaging data although they had originally wanted to become analysts and work in

finance. What they all had in common was that computers were their "lab" and data were their research objects. They ran experiments mostly in MATLAB, carefully modeling how the brain processes and stores information, based on data generated by scanning peoples' brains. To test the resulting models, my interlocutors ran simulations and compared their synthetic data with those of their volunteers' brain activity. In those tests, their models turned into artificial brains and the dividing lines between the study of human brains and the science of artificial intelligence began to blur.

This shift from experiments with real brains to simulations of artificial brains has accelerated in recent years. The lab is now more often found in data centers, and machine learning algorithms are substituted for real brains. Researchers hope to detect patterns in data of human brain activity, and by internalizing these patterns, learn to simulate aspects of human cognition and behavior. That is to say that the field has shifted, and labs have partially moved online. Besides lying in brain scanners and participating in experiments conducted in psychology laboratories, I spent months sifting through science blog posts, studying design documents of microprocessors, and engaging with the scripts and protocols of 1980s Atari video games.[1]

In 2016, I came across an experiment conducted by two computational neuroscientists, who used a simulation of a microprocessor as an artificial brain, to test cutting-edge data analysis methods used to study the brain. The experiment already made waves when the paper was still in the review phase, accessible only through the online repository *arXiv*. Eric Jonas and Konrad Kording had applied methods typically used to analyze brain imaging data to study the operations of the chip while running 1980s video games such as Donkey Kong and Space Invaders. Although the two researchers were curious to see if they could come up with new insights on how the chip brings Donkey Kong to life, the results of their analysis were only of secondary interest. In fact, Jonas and Kording were sure that their experiment would fail.

The experiment was a clever hoax and a gesture of epistemological critique: if what in the real neuroscience world would be a millions-of-dollars data set does not result in some insights about how the processor works, why do we expect that the very same techniques would work on the human brain? Eric Jonas had come up with the idea to analyze the chip at work with cutting-edge brain imaging methods when he came across the *Visual6502* project, a collective effort of "retro-computing enthusiasts" to study, document, and preserve the microprocessor for generations to come (Yong 2016). Using highly detailed photographs, the Visual6502 project had managed to produce a fully functional digital model and simulation of the chip, which allowed, among other things, to play 1980s video games on current computers (fig. 1).

In an interview with *The Atlantic,* Eric Jonas explains how shocked he was that they used the exact same techniques as neuroscientists who are trying to map the brain's connectome. "It made me think that the analogy [between the chip and the brain] is incredibly strong" (Yong 2016). Nevertheless, Jonas and Kording openly admit in their paper that the brain is not actually similar to a processor.

---

[1] See Tara Mahfoud's review of ethnographies of neuroscience practice for an overview (Mahfoud 2014).

Neural systems are analog and biophysically complex, they operate at temporal scales vastly slower than this classical processor but with far greater parallelism than is available in state of the art processors. Typical neurons also have several orders of magnitude more inputs than a transistor. Moreover, the design process for the brain (evolution) is dramatically different from that of the processor (the MOS6502 was designed by a small team of people over a few years). As such, we should be sceptical about generalizing from processors to the brain. (Jonas and Kording 2017, 14)

At the same time, the authors point towards some – rather abstract – similarities in how the workings of human brains and microprocessors have typically been *analyzed*. Anthropologist Joe Dumit observed that circuit diagrams have for decades structured how neuroscientists and psychologists study the human brain through experiments with volunteers in scanners (Dumit 2016). In order to understand how the brain processes what we see, hear, and feel, experimenters still come up with mental tasks for their volunteers, to simulate mundane brain activity while they are lying in brain scanners, waiting for their thoughts to be turned into the meanwhile iconic images of the brain at work. In other words, experimenters activate their volunteers' brains in specific ways to understand how they work.

Figure 1: A visualization of the process by which the MOS 6502
has been optically reconstructed to generate a fully functional digital model
of the chip (Jonas and Kording 2017).
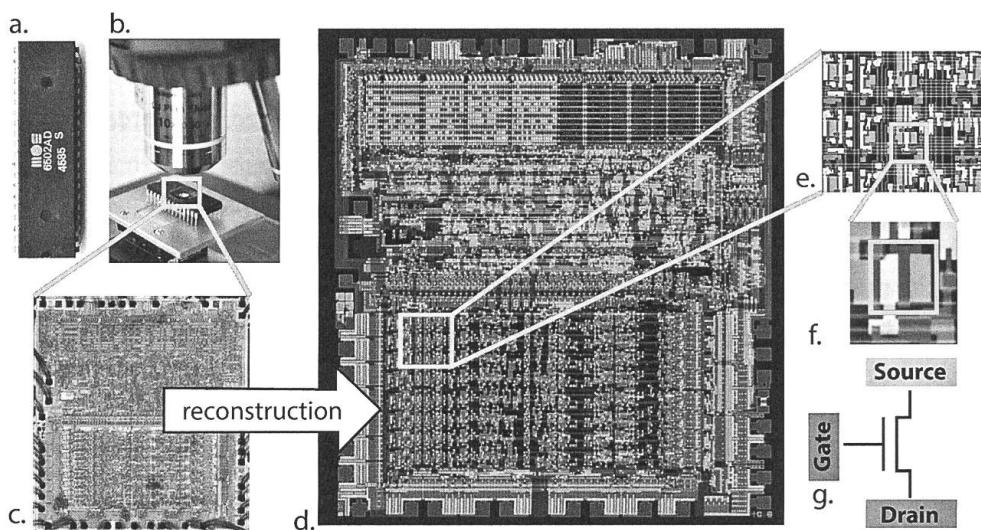
The experiment that undergirds Jonas and Kording's paper tested these methods of contemporary neuroscience and substituted the brain of a volunteer with a legacy microprocessor that was primarily used to play arcade video games. Their argument in a nutshell: despite all the differences between microprocessors and brains, we should be able to understand what the MOS6502 does by means of methods developed to analyze the infinitely more complex human brain.

[W]e cannot write off the failure of the methods we used on the processor simply because processors are different from neural systems. … Altogether, there seems to be little reason to assume that any of the methods we used should be more meaningful on brains than on the processor. (Jonas and Kording 2017, 14–15)

This line of thinking has a precedent. In 2002, the scientific journal *Cancer Cell* published a piece by cell biologist Yuri Lazebnik titled "Can a biologist fix a radio? What I learned while studying apoptosis" (Lazebnik 2002). Lazebnik elaborates on the frustration that takes hold whenever a whole field buys into the most recent hype around a specific target protein that promises to result in a miracle drug – only to abandon it altogether as soon as considerable doubt over the models and methods used arises. "I started to wonder whether anything could be done to expedite this event", Lazebnik writes. "To abstract from peculiarities of biological experimental systems, I looked for a problem that would involve a reasonably complex but well understood system. Eventually, I thought of the old broken transistor radio that my wife brought from Russia" (Lazebnik 2002, 179–180).

The similarities between the titles of Lazebnik's paper and that of Jonas and Kording are anything but coincidental. Lazebnik thought that what biology needs is an unambiguous language, adopted from engineering, to "change from an esoteric tool that is considered useless by many experimental biologists, to a basic and indispensable approach of biology" (Lazebnik 2002, 182). In similar ways, Jonas and Kording argue for the language and methods of data science to become central to research on the brain. That is, the necessity or significance of the biological substrate in experiments is called into question: why conduct complicated experiments with cells or human subjects if experimenting with rather simple, human-designed systems could make everyone's lives much easier?

In both cases, the human-designed circuit stands in for the complex and messy organ to prove that scientific methods currently used must necessarily fail. The Russian transistor radio and the North American microprocessor, however, are not arbitrary choices—they "emerge from particular cultural worlds, not from some technical outside" (Seaver 2018, 379). As model organism, the MOS 6502 invokes specific technocultural practices that cannot be subsumed to the supposedly neutral domain of engineering but extend to the situated worlds of biology laboratories as well as to the cultural niche of 1980s Atari video games.[2]

Model organisms

"[W]e take a classical microprocessor as a model organism, and use our ability to perform arbitrary experiments on it to see if popular data analysis methods from neuroscience can elucidate the way it processes information", Jonas and Kording (2017, 1) write. By referring to the MOS6502 as a model organism, they simultaneously re-engage the in many ways flawed brain-computer metaphor and a rich history of experimentation in biology and psy-

---

[2] I borrow the notion of "technocultural practices" from Kavita Philip, Lily Irani, and Paul Dourish, who use it in the context of tactics for postcolonial computing, referring to media technologies as constituting "the very cultural categories by which some seek to explain them" (Philip, Irani, and Dourish 2012, 14).

chology, where human subjects have been substituted with non-human animals to elucidate specific aspects of – typically disordered – human behavior.

Despite all the differences between human brains and microprocessors, comparing the MOS 6502 to model organisms in biology would seem to legitimize the experiment. Social scientist Nicole Nelson offers that model organisms are rarely a natural fit; "scientists themselves build up the case for their animal models through methodological experiments and arguments that bind together the animal model with the human disorder in a way that makes future experimental work possible and credible" (Nelson 2013, 7). Entire communities of researchers revolve around specific model organisms, a shared toolbox of experimental techniques and a common language that helps handle the – at times overwhelming – differences between humans and the non-human animal.

Jonas and Kording make a similar case for the MOS 6502 as model organism, which they describe as falling somewhere in between the nematode worm *Caenorhabditis elegans* and the lab mouse. The microprocessor sits at the intersection of their lab lives: whereas the mouse provides a fitting behavioral model, "the processor's scale and specialization share more in common with *C. elegans* than a mouse" (Jonas and Kording 2017, 3).

Lab mice

According to the Jackson Laboratory on Mount Desert Island in Maine, CL57BL/6J is the most widely used inbred strain of laboratory mice. JAX, as the laboratory is also referred to, has its very own "black 6" that carries the letter J as a postfix to the code that unmistakably identifies the laboratory mouse as a commodity or "technical thing" (Rheinberger 1997). C57BL/6J "is a permissive background for maximal expression of most mutations", the website states, and thus "used in a wide variety of research areas including cardiovascular biology, developmental biology, diabetes and obesity, genetics, immunology, neurobiology, and sensorineural research".[3]

Despite C57BL/6J's commodification, many science studies scholars and scientists themselves oppose considering lab mice as mere tools. The rodents that appear alongside amino acids, centrifuges, and other technical things in Hans-Jörg Rheinberger's analysis of experimental systems in biology do not lend themselves easily to standing in for the human, particularly if they are to reproduce behavior that does not come natural to mice – such as binge drinking, for instance. That is to say that even the genetically modified CL57BL/6J, which supposedly has a preference for alcohol and morphine, remains averse to consuming extensive amounts of alcohol. How can binge drinking behavior in mice be "naturalized" if even genetically modified strains will not drink enough alcohol to exhibit blood alcohol levels comparable to those that humans seem to enjoy?

Sabina Leonelli and colleagues have analyzed the use of mice as situated models in North American alcohol research throughout the twentieth and early twenty-first century and

---

[3] Please see Jackson Laboratory's website for more detailed information: https://www.jax.org/strain/000664, accessed February 7, 2021.

found that environmental factors and experimental situations have taken center stage in discussions about the validity of mice to model human drinking behavior (Leonelli et al. 2014). Researchers cannot resort to any "unnatural" incentives or force mice into drinking without tampering with the phenomenon of drinking itself. To create a situation where the mouse adapts human behavior, the lab and the experimental setup must constantly be reconfigured. "Mice and humans, mazes and drugs, genes and behaviors, practical experience and widely recognized findings – all these are continually and carefully set in relation to each other to create a space that functions as credible site for producing knowledge about human behaviour" (Nelson 2018, 6).

Experimenting with *C. elegans* is supposedly easier, not least since its brain has meanwhile been digitally mapped. The nematode worm was a staple of the Human Genome Project (HGP) and has become a poster child of genetics research in biology. Despite being infinitely less complex than humans, *C. elegans* promised insights into certain common or even universal biological mechanisms. Differences in complexity have typically been smoothed over by a hypothesized common biological lineage, well-founded in the theory of evolution and formalized in the genetic code. Ruth Ankeny, who studied the use of worms in the HGP, cites a *Science* article from 1998, where Francis Collins – former director of the HGP – and colleagues argue that

all organisms are related through a common evolutionary tree [and] the study of one organism can provide valuable information about others ... Comparisons between genomes that are distantly related provide insight into the universality of biologic mechanisms and identify experimental models for studying complex processes. (Collins, quoted in Ankeny 2007, 47)

Lab worms

Despite the fact that *C. elegans* is fairly atypical even compared to closely related organisms, it gained biological prominence because of its experimental manipulability and tractability: "an organism that proved experimentally straightforward to manipulate and had relatively basic behaviors and structures, but was not so simple as to be 'unrepresentative'", Ankeny writes (Ankeny 2007, 49). In fact, the experimental manipulability of *C. elegans* rests foremost on the ease of breeding an array of "actual material worms" – using the worm as a model organism therefore allows to construct what Ankeny calls a "data-summarizing descriptive device" (op. cit.). Thanks to its simplistic biologic make-up, it was the first multicellular organism with a completely sequenced genome and a known "brain".[4] Both are ideal types that do not exist in nature – thus "summarizing" and "descriptive" – but they provide stable models that permit investigating, analyzing and quantifying deviations from what is consid-

---

[4] The *C. elegans* genome was initially completed in 1998 and updated roughly two decades later to include genetic variations and mutations (The C. elegans Sequencing Consortium 1998; Yoshimura et al. 2019). The wiring diagram of neural connections for female worms has been available since 1986; it was updated, digitized and complemented by the male wiring diagram in 2019 (Cook et al. 2019).

ered normal or biologically sound. The genetically modified and standardized lab worm is hence closer to being an experimental device or machine than to an actual living organism.

Jonas and Kording consider *C. elegans* as a fitting analogy for the MOS 6502 since the nematode worm is as different from the human as the microprocessor is from a brain. Experimenting on *C. elegans* is considered worthwhile not because it is biologically or physically similar to humans, but because worms appear to be infinitely simpler to control – especially if they are disembodied and dematerialized in simulations. In a companion piece to the paper that announced the completion of a digital atlas of *C. elegans* brain, (neuro)biologist Douglas Portman contends that a detailed simulation of the worm's nervous system will allow "generating a virtual worm that 'lives' inside a computer" (Portman 2019).

In ways similar to the case of the virtual *C. elegans*, the virtualized MOS 6502 acts as a proxy for the human brain and was chosen as a model organism since "it is fully accessible to any and all experimental manipulations that we might want to do on it" (Jonas and Kording 2017, 3). And in ways similar to the case of the lab mouse, the experimental tasks for MOS6502 were chosen to mediate between behaviors that come "natural" to both, chips and brains. In fact, their choice of experimental task was not arbitrary – it reflects the experimenters' familiarity with certain test beds and attendant technocultural practices.[5]

The games they play(ed)

In their paper, Eric Jonas and Konrad Kording half-jokingly admit that most of their colleagues

> *have at least behavioral-level experience with … classical video game systems, and many in our community, including some electrophysiologists and computational neuroscientists, have formal training in computer science, electrical engineering, computer architecture, and software engineering. As such, we believe that most neuroscientists may have better intuitions about the workings of a processor than about the workings of the brain.* (Jonas and Kording 2017, 3)

I write half-jokingly since their statement obviously represents an ironic use of field-specific language to say what many of my interlocutors in neuroscience laboratories emphasized: that they feel more comfortable experimenting with models and algorithms than with living model organisms. This sort of irony is well-known among ethnographers of computing cultures (Coleman 2010; Seaver 2017). It helps navigate the ambiguities of computational practice – such as data that are "of the world" and simultaneously "of the computer", or test beds that are highly constrained, yet nevertheless pose as "microcosms of real-world problems" (Hassabis 2016). Engineers and computer scientists know very well that video games

---

[5] I use the notion of test bed in dialogue with Orit Halpern, Jesse Lecavalier, Nerea Calvillo, and Wolfgang Pietsch's "test-bed urbanism", which describes how tests, experiments, and demos are increasingly embedded in real life contexts and thus have immediate effects on our lives (Halpern et al. 2013). See also Marres and Stark (2020).

do not resemble the real world; and yet, some problems and strategies coded into video games such as Donkey Kong (1981), Space Invaders (1978), and Montezuma's Revenge (1984) are currently omniscient in artificial intelligence research and determine the problems that artificial agents are supposed to solve.

At one of the current powerhouses of machine learning and artificial intelligence research, the London-based Google subsidiary DeepMind, many leading protagonists have experience in professional chess and video game design. Their very own brand of "neuroscience-inspired artificial intelligence" consists in augmenting commonly used machine learning approaches with mechanisms that are supposedly at work also in the human brain (Hassabis et al. 2017). To test the capacities of their agents, researchers typically resort to well-known games. AlphaGo, for instance, became famous for beating the reigning (human) Go world champion in 2016. A more recent version, Agent57, learned to master all video games originally designed for Atari computers that were powered by the MOS 6502. In these experiments, the chip is *not* the model organism; instead, the so-called "Arcade Learning environment" is reconsidered as test bed for general artificial intelligence. That is, the video game is for Google DeepMind's Agent57 what the maze is for JAX's CL57BL/6J.

Google DeepMind offers that video games are an excellent testing ground for machine learning algorithms. Their ultimate goal is not to develop systems that excel at games; rather, gameplay is used "as a stepping stone for developing systems that learn to excel at a broad set of challenges". Video games reportedly force machine learning algorithms to develop "sophisticated behavioural strategies" and the high score provides "an easy progress metri to optimise against". Indeed, Agent57 outscored the average human in each of the 57 Atari 2600 games it learned to play. According to Google DeepMind, it "performed *sufficiently well* on a *sufficiently wide* range of tasks" and would thus need to be considered intelligent (Badia et al. 2020).

But what form of intelligence is this? In Donkey Kong, the protagonist and Mario are kept in an endless loop of outwitting their opponent, throwing wooden barrels or climbing ladders to ultimately win (the heart of) the princess. Pitfalls and Space Invaders are similar in that they reduce life to surviving in adverse environments, where the protagonists have the opportunity to roam the virtual worlds at will and amass capital if they find ways to outwit their opponents and survive. Yet, analyses of the strategies that Agent57 and its peers developed revealed that they did often not satisfy this rather simple objective (Ecoffet et al. 2019; Lehman et al. 2020). For instance, in Montezuma's Revenge an agent exploited a bug to remain in the treasure room indefinitely and collect unlimited points, instead of being moved to the next level and finish the game.

The agents would often reach high scores while failing to solve the actual problem (Krakovna et al. 2020). But does that mean that they failed? "If I put you in front of a computer game, you'll treat the point score as the objective", a machine learning researcher put it to me while we were discussing DeepMind's experiments. "And it seems to me that this is quite a delicate thing, because you have every incentive to sort of lie to yourself and look for the loophole that lets you score high without actually finishing the game or even dying as quickly as possible". Against this backdrop, it would seem that the sort of intelligence that machine learning algorithms exhibit when trained in the highly constrained worlds of 1980s Atari

video games is essentially that which video game designers expected human players to develop when they built cheat codes into their games.

If DeepMind researchers admit that many agents currently excel in exploiting loopholes, they reify specific cultural parameters of success that define intelligence as the ability to amass enormous amounts of capital without necessarily solving any problems. Against this backdrop, singularity acquires a whole new meaning: the general intelligence that is put to the test is modelled after a very specific and singular understanding of what human intelligence involves. It universalizes the idea of a player programmed into 1980s Atari video games and restricts the task of an agent to outperforming this benchmark. What the resulting artificial intelligence throws back at us is a radically provincial idea of human creativity, intelligence, and ability, courtesy of technocultural practices that derive from the domain of video game design in Europe, the US, and Japan.

Old games, new worlds

In concluding, I would like to return one final time to one of the most significant statements within Jonas and Kording's paper: the brain is not a chip! Throughout, the authors add qualifiers to the analogy between brains and microprocessors that should please all humanists and appease experimental psychologists and neuroscientists. Yet, what sits at the heart of their statement is a re-engagement of the failed brain-computer metaphor rather than its outright rejection. Jonas and Kording's choice to substitute a simulated version of the legacy microprocessor for the human brain in their experiment shows to what extent computing and cognitive science have for decades been entangled. While brains are no longer compared to computers, computational neuroscience has successfully implemented a language that allows to describe cognitive processes in the Cloud and in human brains in similar and compatible ways (Bruder 2019).

Again, this thinking has a historical precedent. In the mid-1950s, social scientist and artificial intelligence forerunner Herbert Simon and his colleagues at the RAND Corporation tried to model and operationalize human reasoning on the JOHNNIAC computer, to investigate a phenomenon Simon dubbed "bounded rationality" (Simon 1957). It was an attempt at exploring the very possibility of rational decision making and experimenting on the computer's capacity to circumnavigate the limits that biology imposed. Simon and his colleagues therefore implemented a program called Logic Theory Machine on the JOHNNIAC. In a paper published in 1958, Simon and colleagues offer that they "are not comparing computer structures with brains, nor electrical relays with synapses. Our position is that the appropriate way to describe a piece of problem-solving behavior is in terms of a program: a specification of what the organism will do under varying environmental circumstances in terms of certain elementary information processes it is capable of performing" (Newell, Shaw, and Simon 1958, 153).

In 1958, the JOHNNIAC put a strict limit on the complexity of human behaviors that could be simulated – both due to its insufficient computational capacities (Dick 2015) and the deeply North American technocultural practices that manifested in its design. The case of

MOS6502 is similar. While generally a multi-purpose processor, it comes with certain "naturalistic behaviors" which predetermine what "sophisticated behavioral strategies" *can be*. The player coded into 1980s video games is a liberal subject that conquers and stays, to accumulate capital and outscore existing benchmarks. The fact that these agents discover loopholes and bugs in video games is thus hardly "surprising creativity" (Lehman et al. 2020, 274). It proves that the agent performs sufficiently well in its task to mimic and outperform the player that was coded into the game. Against this backdrop, it would seem that it is not primarily the machine that contaminates our understanding of (human) intelligence, but that a very provincial understanding of the liberal subject, rooted in settler colonialist thinking and manifested in Arcade video games, bleeds into machine learning algorithms by way of the experimental task.

Nevertheless, this article is not to dismiss computational test beds, demos, and experiments; instead, it is an attempt at problematizing the notion that artificial intelligence must necessarily restrict our understanding of intelligence. The question that I would like to ask is whether we can avoid over-fitting agents to test beds that privilege exploitative strategies and winning at all costs? In other words, could the use of microprocessors and machine learning algorithms as model organisms incite experiments where what an organism *can become* is not determined by the design of its test beds?

A change of perspective on the laboratory and model organisms allows for very different mediations between the digital and the biological to come to the fore. In her analysis of mice as model organisms in post-genomic biology, geographer Gail Davies argues that the humanized mouse potentially offers a means of escape from the biological grammar of genomics. She observes that the "process of 'becoming human' opens these model organisms to biological relations, which are not only interior but also external, remaking relations between experimental subjects and objects, laboratory spaces and clinical contexts" (Davies 2013, 147). Hers is an invitation to study the humanized mouse as an object of patchwork, linking different fields and territories of scientific investigation and the continued exhaustion of nature (Tsing, Mathews, and Bubandt 2019). It may also be read as a call for attending to the epistemological and ontological opportunities of multispecies thinking, of thinking through, yet most importantly, beyond the roles that mice are relegated to in the lab: "not only are there many humanized mice in the world, there are also many worlds in the humanized mouse" (Davies 2013, 147).

In this spirit: there's potentially more also to Donkey Kong's legacy than the contamination of artificial intelligence through the situated technocultural practices of 1980s video game design. For instance, after youtuber @Hbomberguy played through Donkey Kong 64 in a gruelling 57-hour shift to raise financial support for UK trans charity Mermaids, his live stream on Twitch turned into a gathering of trans rights activists and claimed Donkey Kong as an icon of trans rights online, literally opening up new worlds in the humanized ape. This little episode from contemporary digital culture arguably shows how technocultural practices can be reclaimed and reconceived – a task that machine learning could valuably support if it waved goodbye to test beds and benchmarks that reward winning at all costs.

References

Ankeny, Rachel A. 2007. "Wormy Logic: Model Organisms as Case-Based Reasoning." In *Science Without Laws. Model Systems, Cases, Exemplary Narratives*, edited by Angela N. H. Creager, Elizabeth Lunbeck, and M. Norton Wise, 46–58. Durham, NC: Duke University Press.

Badia, Adrià Puigdomènech, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, and Charles Blundell. 2020. "Agent57: Outperforming the Atari Human Benchmark." *ArXiv:2003.13350 [Cs, Stat]*, March. http://arxiv.org/abs/2003.13350.

Bruder, Johannes. 2019. *Cognitive Code: Post-Anthropocentric Intelligence and the Infrastructural Brain*. Kingston & Montréal: McGill-Queen's University Press.

Coleman, E. Gabriella. 2010. "Ethnographic Approaches to Digital Media." *Annual Review of Anthropology* 39, no. 1: 487–505. doi:10.1146/annurev.anthro.012809.104945.

Cook, Steven J., Travis A. Jarrell, Christopher A. Brittin, Yi Wang, Adam E. Bloniarz, Maksim A. Yakovlev, Ken C. Q. Nguyen. 2019. "Whole-Animal Connectomes of Both Caenorhabditis Elegans Sexes." *Nature* 571, no. 7763: 63–71. doi:10.1038/s41586-019-1352-7.

Davies, Gail. 2013. "Mobilizing Experimental Life: Spaces of Becoming with Mutant Mice." *Theory, Culture & Society* 30, no. 7-8: 129–53. doi:10.1177/0263276413496285.

Dick, Stephanie. 2015. "Of Models and Machines: Implementing Bounded Rationality." *Isis* 106, no. 3: 623–34. doi:10.1086/683527.

Dumit, Joseph. 2016. "Plastic Diagrams: Circuits in the Brain and How They Got There." In *Plasticity and Pathology: On the Formation of the Neural Subject*, edited by David Bates and Nima Bassiri, 219–67. Berkeley: Fordham University Press.

Ecoffet, Adrien, Joost Huizing, Joel Lehman, Kenneth O'Stanley, and Jeff Clune. 2019. "Go-Explore: A New Approach for Hard-Exploration Problems." Accessed April 22, 2020. https://arxiv.org/abs/1901.10995.

Halpern, Orit, Jesse LeCavalier, Nerea Calvillo, and Wolfgang Pietsch. 2013. "Test-Bed Urbanism." *Public Culture* 25, no. 2, 272–306. doi:10.1215/08992363-2020602.

Hassabis, Demis, Dharshan Kumaran, Christopher Summerfield, and Matthew Botvinick. 2017. "Neuroscience-Inspired Artificial Intelligence." *Neuron* 95, no. 2: 245–58. doi:10.1016/j.neuron.2017.06.011.

Jonas, Eric, and Konrad Paul Kording. 2017. "Could a Neuroscientist Understand a Microprocessor?" edited by Jörn Diedrichsen. *PLOS Computational Biology* 13, no. 1: e1005268. doi:10.1371/journal.pcbi.1005268.

Kellogg, Katherine C., Melissa A. Valentine, and Angéle Christin. 2020. "Algorithms at Work: The New Contested Terrain of Control." *Academy of Management Annals* 14, no. 1: 366–410. doi:10.5465/annals.2018.0174.

Knorr Cetina, Karin. 1992. "The Couch, the Cathedral, and the Laboratory: On the Relationship between Experiment and Laboratory in Science." In *Science as Practice and Culture*, edited by Andrew Pickering, 113–37. Chicago: University of Chicago Press.

Krakovna, Victoria, Jonathan Uesato, Vladimir Mikulik, Matthew Rahtz, Tom Everitt, Ramana Kumar, Zac Kenton, Jan Leike, and Shane Legg. 2020. "Specification Gaming: The Flip Side of AI Ingenuity." *DeepMind Blog*. April 21. Accessed December 8, 2020. https://deepmind.com/blog/article/Specification-gaming-the-flip-side-of-AI-ingenuity.

Lazebnik, Yuri. 2002. "Can a Biologist Fix a Radio?—Or, What I Learned While Studying Apoptosis." *Cancer Cell* 2, no. 3: 179–82. doi:10.1016/S1535-6108(02)00133-2.

Lehman, Joel, Jeff Clune, Dusan Misevic, Christoph Adami, Lee Altenberg, Julie Beaulieu, Peter J. Bentley. 2020. "The Surprising Creativity of Digital Evolution: A Collection of Anecdotes from the Evolutionary Computation and Artificial Life Research Communities." *Artificial Life* 26, no. 2: 274–306. doi:10.1162/artl_a_00319.

Leonelli, Sabina, Rachel A. Ankeny, Nicole C. Nelson, and Edmund Ramsden. 2014. "Making Organisms Model Human Behavior: Situated Models in North-American Alcohol Research, 1950-Onwards." *Science in Context* 27 no. 3: 485–509. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4274764/.

Mahfoud, Tara. 2014. "Extending the Mind: A Review of Ethnographies of Neuroscience Practice." *Frontiers in Human Neuroscience* 8: 359. doi:10.3389/fnhum.2014.00359.

Markoff, John. 2016. "Alphabet Program Beats the European Human Go Champion." *The New York Times*, January 27, 2016. Accessed April 21, 2021. https://bits.blogs.nytimes.com/2016/01/27/alphabet-program-beats-the-european-human-go-champion/.

Marres, Noortje, and David Stark. 2020. "Put to the Test: For a New Sociology of Testing." *The British Journal of Sociology* 71, no. 3: 423–43. doi:10.1111/1468-4446.12746.

Nelson, Nicole C. 2013. "Modeling Mouse, Human, and Discipline: Epistemic Scaffolds in Animal Behavior Genetics." *Social Studies of Science* 43, no. 1: 3–29. doi:10.1177/0306312712463815.

Nelson, Nicole C. 2018. *Model Behavior: Animal Experiments, Complexity, and the Genetics of Psychiatric Disorders.* Chicago: The University of Chicago Press.

Newell, Allen, J. C. Shaw, and Herbert A. Simon. 1958. "Elements of a Theory of Human Problem Solving." *Psychological Review* 65, no. 3: 151–66. doi:10.1037/h0048495.

Philip, Kavita, Lilly Irani, and Paul Dourish. 2012. "Postcolonial Computing: A Tactical Survey." *Science, Technology, & Human Values* 37, no. 1: 3–29. doi:10.1177/0162243910389594.

Portman, Douglas. 2019. "The Minds of Two Worms." *Nature*, July 4. https://media.nature.com/original/magazine-assets/d41586-019-02006-8/d41586-019-02006-8.pdf.

Rheinberger, Hans-Jörg. 1997. *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube.* Stanford: Stanford University Press.

Seaver, Nick. 2017. "Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems." *Big Data & Society* 4, no. 2. doi:10.1177/2053951717738104.

Seaver, Nick. 2018. "What Should an Anthropology of Algorithms Do?" *Cultural Anthropology* 33, no. 3: 375–85. doi:10.14506/ca33.3.04.

Simon, Herbert. 1957. *Models of Man.* New York: Wiley.

The C. elegans Sequencing Consortium. 1998. "Genome Sequence of the Nematode C. Elegans: A Platform for Investigating Biology." *Science* 282, no. 5396: 2012–18. doi:10.1126/science.282.5396.2012.

Tsing, Anna Lowenhaupt, Andrew S. Mathews, and Nils Bubandt. 2019. "Patchy Anthropocene: Landscape Structure, Multispecies History, and the Retooling of Anthropology: An Introduction to Supplement 20." *Current Anthropology* 60, no. S20: 186–97. doi:10.1086/703391.

Yong, Ed. 2016. "Can Neuroscience Understand Donkey Kong, Let Alone a Brain?" *The Atlantic*, June 2, 2016. Accessed April 21, 2020. https://www.theatlantic.com/science/archive/2016/06/can-neuroscience-understand-donkey-kong-let-alone-a-brain/485177/.

Yoshimura, Jun, Kazuki Ichikawa, Massa J. Shoura, Karen L. Artiles, Idan Gabdank, Lamia Wahba, Cheryl L. Smith. 2019. "Recompleting the *Caenorhabditis Elegans* Genome." *Genome Research* 29, no. 6: 1009–22. doi:10.1101/gr.244830.118.

Acknowledgements

Author

Johannes Bruder ⓘ works at the intersection of anthropology, STS, and media studies. He studies the history and present of decision-making systems, and how these encode psychological categories, sociological models, artistic practices and speculative designs. His first book "Cognitive Code. Post-Anthropocentric Intelligence and the Infrastructural Brain" (2019) is based on fieldwork in neuroscience laboratories and provides deep insights into the bio-politics of contemporary machine learning. Johannes has a strong interest in experimenting with research methods, knowledge practices, alternative pedagogies and publication formats that unsettle disciplinary paradigms and render research in the humanities operational in real-world contexts. He is a senior researcher at the Critical Media Lab Basel and affiliated with the Department of Sociology and Anthropology, as well as Milieux at Concordia University Montréal.

johannes.bruder@protonmail.com
FHNW Academy of Art and Design – Institute of
Experimental Design and Media Cultures