Alex King – PHIL 3, Professor Daniel Dennett
Conway's Game of Life as an Intuition Pump for Consciousness, or, Everything I Know about Philosophy I Learned from Conway's Game of Life

I. An Appeal to Intuition

What does it mean for something to be lifelike? It's a surprisingly hard question to answer in a few words. We might say that if an entity "moves in a lifelike way", then it looks alive. But what is lifelike movement? How is a snail's steady movement any more lifelike than a raindrop falling from above? We ascribe an entity the attribute of being lifelike when it moves in a seemingly spontaneous manner, or when we ascribe an entity beliefs or desires that might encourage it to act in a certain way. Acknowledging that something appears lifelike in its behavior is known as adopting the *intentional stance*, and it is a useful perspective to take when trying to understand how to define life, and how one form of life might differ from another.

A fascinating example of apparent "life" is a simple evolving game known as Conway's Game of Life. Created by mathematician John Horton Conway in 1970[1], it is not much of a game in an interactive sense. Instead, an initial configuration is established on its two-dimensional grid of cells, and once started, a small set of rules takes over, allowing the game to proceed without any user input. To that end, it is really much more of a simulation: an environment created with a set of rules where a dimension of time allows us to see how entities progress in the environment. Conway's Game of Life is often referred to as Life; for the purpose of contrast with *physical life*, I will henceforth refer to it as *The Game*.

The rules of The Game are simple. Each alive cell with two or three alive neighbors will stay on during the next iteration, representing survival among a community of others. Each alive cell with one or fewer alive neighbors will be turned off during the next iteration, representing

[1] Gardner, Martin. "Mathematical Games: The Fantastic Combinations of John Conway's New Solitaire Game 'life'" *Scientific American* Oct. 1970: 120-23. *Ibiblio*. Web. 24 Apr. 2016.

death by isolation. Each alive cell with four or more alive neighbors will be turned off during the next iteration, representing death by competition and overpopulation. And finally, each dead cell with three alive neighbors will be turned on during the next iteration, representing reproduction. The rules are straightforward and easy to follow, such that one can play the game by hand on paper or on a board with physical pieces. In more recent years, the game has found a home as a web application, where all of the tedious counting and rule following can be automated, and the speed of simulation can be increased significantly. I encourage the unacquainted reader to experiment with one of these applications, especially one that provides interesting initial configuration templates.

Conway sought to design a game that met a few important characteristics, which are outlined in a 1970 article in *Scientific American* introducing the game. He did not want there to be a simple initial configuration of the board that could be mathematically proven to grow without limit. However, he wanted there to be initial configurations that *appeared* to grow without limit. Most importantly, he wanted there to be certain initial patterns that underwent considerable evolution before eventually stagnating, dying out completely, or repeating a cycle.[2] Conway's result is a game that meets these goals, and one that introduces considerable philosophical beauty.

Play The Game for a little while and you will recognize some of its emergent properties. First, and perhaps unsurprisingly, an empty simulation will remain empty forever. Nothingness persists; there is no spontaneity or random birth in The Game. Second, if seeded with one or more *on* cells, the viewer will observe the following characteristics. Cells are born, cells die, some shapes appear to persist as they move through space, (often called "gliders"), and in many cases,

[2] Ibid.

an arbitrary initial configuration will evolve from one configuration to a steadier one, whether it is a static still life or a loop (gliders are one such looping shape).

Conway's decision to call it a game of "life" is fitting. Cells are less frequently addressed as either *on* or *off* but rather either *alive* or *dead*, and it is hard not to interpret the movement of a moderately sized simulation as appearing "lifelike" in its movement, evolution, birth and death. The incredibly simple rules give rise to exceedingly complex systems. What if The Game is really modeling life, in the same way that we think of physical life? What if all of the same principles that apply to The Game also applied to physical life? For the purpose of argument, let us take the life parallel to its extreme and explore its implications for the physical world. The Game is just so intangibly lifelike that it *must* have something in common with our world. If we follow the metaphor and see where it takes us, we'll find a basic mental framework for understanding consciousness, intelligence, and free will in the physical world.

Computational Properties of The Game

It's important that I establish a clear logical direction of implication for this metaphor, because one direction is far more exciting than the other. I claim that The Game can be seen a metaphor for physical life – but which contains the other? It's not especially profound to claim that physical life *can simulate* The Game; that has already been neatly accomplished. Note that I mean *to simulate* as in *to supervise*, implying that the supervisor is the creator of the simulation, not that it itself acts as the simulation. When discussing physical life, we should also mention the human brain as well, as it is contained in physical life. In this case, a human brain can also simulate the proceedings of The Game with ease, with the help of some basic offloading with pencil and paper.

The other direction is more interesting. It is profound, and provocative, to say that The Game *can simulate* physical life, and thus a human brain. If it seems extreme, return to the intuition of a complex simulation happening in The Game. Little striding agents are spawned from random initial configurations. Entities evolve, eventually leaving behind static structures that lay basis for further interaction. Just as the physical world has certain physical laws that have allowed for evolution of organisms over time, The Game has its own set of physical laws that appear to allow for the same basic loop of life, death, and evolution. Just because it is challenging to imagine a vastly complex configuration acting as a brain does not mean it could not happen in theory.

To make this argument more mathematically interesting, I will introduce one more property of The Game: it is Turing complete. Turing completeness is the property whereby a machine *can simulate* a single-taped Turing machine, which is an abstract machine that reads input as letters on a tape, and depending on the input and its current state, it can mark the tape, move left or right, and change state. From simple mechanical rules come great computational power – modern computer processors are no more computationally powerful than a Turing machine, just orders of magnitude faster and more complex in their architecture.[3] The Game was proven to be Turing complete after its inception in 1970, and later demonstrated to be so by Paul Rennell in 2001, who exhibited a Turing machine built entirely from on and off cells on The Game's two-dimensional grid.[4] This means that any computer program ever written, regardless of complexity, could be executed using a certain initial configuration drawn onto the grid. This is a worthwhile exercise to

[3] Dennett, Daniel. "Are We Explaining Consciousness Yet?" *Cognition* 79.1-2 (2001): 221-37. *ScienceDirect*. Web. 26 Apr. 2016.
[4] Rennell, Paul. "A Turing Machine In Conway's Game Life." *University of California, Irvine*. N.p., 30 Aug. 2001. Web. 26 Apr. 2016.

imagine, because it helps one think of how The Game's arbitrarily large board size can lead to arbitrarily complex simulations.

The Game's property of being Turing complete has exciting implications for our thought experiment. This means that we can establish a useful Turing equivalence:

- Conway's Game of Life *can simulate a* Turing machine
- A Turing machine *can simulate* Conway's Game of Life (via any computer running the simulation)

This means they are equal in computational power. A Turing machine can do anything Turing computable. Conway's Game of Life can simulate a Turing machine, achieving that same behavior, and it cannot accomplish anything else, else it would be more powerful than a Turing machine. We know that the physical world can model a Turing machine. Might a Turing machine be able to model the physical world, too?

II. Implications

What might this metaphor give us? I've described how The Game has characteristics of the physical world, and that it might not be coincidence, but rather because real life actually behaves in the same way. By understanding The Game, we can gain a powerful intuition about several philosophical questions, including the question of consciousness, the possibility of strong AI, and the presence, or lack thereof, of free will. I'll touch on each of these in the pages to come.

Concerning Consciousness

How does the human mind fit into our metaphor? From my mapping, I am claiming that The Game can simulate a human brain. Our theory suggests that consciousness is not magic, but rather a result of many, many things happening all at once. If we claim that The Game can model a conscious mind, it's immediately clear that something enormous has to be built from very simple parts. The Game is the simplest form of building blocks imaginable: on and off squares in two dimensions. We see that there is no magic, no intangible "experience" that is not completely visible as two-dimensional information on the game board. The implications for consciousness here are clear. There is no "hard problem" of consciousness: consciousness is built up by many, many small parts. Subjective experience, or *qualia*, cannot exist; there is only objective physical matter in this world, and any "experience" of seeing the color red is simply an informational transduction.

This view is supported by much cognitive science research. In the paper "Are We Explaining Consciousness Yet?", Daniel Dennett summarizes several scientific insights that give credibility to the theory of consciousness being an illusion of vast parallel processing, what he refers to as "fame in the brain." He cites Dehaene and Naccache's global workspace model, where it is suggested that the experience of consciousness comes from unconscious parallel processing of information, which can become conscious via "top-down attentional amplification into a brain-scale state of coherent activity."[5] Dennett is quick to note that "top-down" does not literally imply an "organizational summit" existing in the brain, but rather that it implies a cooperation of influences from many angles. Dennett likens consciousness to "political influence" and asserts that the ultimate capabilities of the conscious brain is achieved by a nesting of less-intelligent *homunculi*, with brain regions becoming smaller and smaller, less and less capable on their own, such that things eventually "bottom out." At this point, he posits, their functions could be replaced by machines.[6] Finally, he suggests that with this understanding of how mental computation takes place, there is no reason that one could not replace one's brain with "silicon chips and wires," and continue to be conscious just as before.[7]

---

[5] Dennett, Daniel. "Are We Explaining Consciousness Yet?"
[6] Ibid.
[7] Ibid.

Dennett's observations are grounded in brain research and exist in strong opposition to any theory purporting there to be more to the story of consciousness. This aligns precisely with what The Game demonstrates. Though it's hard to imagine exactly how a human brain might be drawn as a living organism in The Game, it's not hard to imagine its behavior: consciousness would not exist in any one place; it would simply be built up from a vast collection of simple two-dimensional machines, all iterating, pumping, signaling, and responding to external forces. Life can be observed from composed machinery. Why should consciousness be viewed any differently?

Concerning Strong AI

In the field of artificial intelligence (AI), There is a distinction between weak AI and strong AI. Weak AI is often plenty sophisticated, but its crutch is its narrow form of sophistication: weak AI is generally only capable of accomplishing one task, or a small family of similar tasks. For example, in March 2016, DeepMind's AlphaGo program beat Lee Sedol in a five match Go tournament, making international news and setting a new bar for the current state of AI capability.[8] By many measures, this was a triumph for the progression of AI, yet the program would still be defined as *weak* AI, because it was designed, trained, tested and tweaked to play Go, and only Go, with a high level of sophistication. Strong AI is a conceptual benchmark: artificial intelligence that matches general cognitive intelligence, such that a strong AI module could solve a variety of problems with a level of sophistication comparable to that of a human's.

In the essay "The Practical Requirements for Making a Conscious Robot", Daniel Dennett and co-authors describe the possibility of a conscious robot; i.e., strong AI. Dennett first responds to objections some may have about the impossibility of a conscious robot, including the argument

[8] Borowiec, Steven, and Tracey Lien. "AlphaGo Beats Human Go Champ in Milestone for Artificial Intelligence." *Los Angeles Times*. Los Angeles Times, 12 Mar. 2016. Web. 26 Apr. 2016.

that "robots are inorganic (by definition), and consciousness can exist only in an organic brain."[9] He counters, citing that organic compounds are proven to be no more mechanically capable than other physical medium. He goes on to explain the Cog project, focused on making a humanoid robot with the capacity to see and hear its surroundings. This paper is useful in establishing that strong AI is certainly a possibility, but it will rely on the entity itself having complex, nuanced capabilities to perceive the world and reason about it.

Paul Churchland and Patricia Churchland wrote about the prospect of strong AI in the 1990 essay "Could A Machine Think?". The authors recap the concept of the Turing test, and note that though Turing's original test had to do with question answering, the same principle of feasibility would be true if the test interacted with the world through vision, speech, or other mediums.[10] It is proposed that strong AI is achievable if the architecture of the machine becomes more like a human itself, citing three particular differences in design between standard computer architectures and the human brain. First, they note that the nervous system is massively parallel, much more than the scope of typical parallel computing. Second, they note that neurons are analog in their response rather than digital. Lastly, brain axons are not one-way, but rather omnidirectional, with "recurrent projections" allowing the brain to "modulate the character of its sensory processing."[11] With these three observations, the model of an *artificial neural network* is introduced, a method where a network of digital neurons recreates parallel computation. In an abstract sense, the neural network is capable of transforming any input vector to a corresponding output vector. Profoundly, it is noted that a neural network is computationally fault-tolerant; several connections can be disabled

---

[9] Dennett, D. C., F. Dretske, S. Shurville, A. Clark, I. Aleksander, and J. Cornwell. "The Practical Requirements for Making a Conscious Robot [and Discussion]." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 349.1689 (1994): 133-46. *JSTOR*. Web. 26 Apr. 2016.
[10] Churchland, Paul M., and Patricia Smith Churchland.
[11] Ibid.

without real negative impact on the output vector.[12] In practice, neural networks have been found to excel in complex machine learning tasks, such as image classification, perhaps because they work closer to how the human brain works (in the case of image classification, the network design has similarities to the structure of the visual cortex).[13] Churchland and Churchland argue that with this vast architecture, it is easier to reason about a computer program recreating a brain. Responding to Searle's tantalizing Chinese Room thought experiment, specifically its "parallel" incarnation, they state, "It is irrelevant that no unit in his system understands Chinese, since the same is true of nervous systems; no neuron in my brain understands English, although my whole brain does."[14]

Our theory suggests that strong AI and human intelligence are one in the same, because they are both concepts simulated on Turing machines. The Game can simulate life, so it can simulate the human brain. Our theory asserts that there is no magic that goes into making the human brain. Rather, what may separate us from achieving strong AI is not a mystery of consciousness and intelligence, but rather in the computational architecture of the mind. As new standards of AI performance continue to be set with brain-like neural networks, the reality of strong AI may not be far away.

Concerning Free Will

Having taken as truth that The Game can simulate a brain, what might we be able to learn about decision making and free will? One corollary of the rules of The Game is that a game state is either defined by *an evolution from a previous state*, or *the initial configuration.* In other words,

---

[12] Ibid.
[13] Goodfellow, Ian J., Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, and Vinay Shet. "Multi-digit Number Recognition from Street View Imagery Using Deep Convolutional Neural Networks." (2014): n. pag. ArXiv:1312.6082 [cs.CV]. Web. 6 Mar. 2016.
[14] Churchland, Paul M., and Patricia Smith Churchland.

everything that ever exists in the universe comes from what happened previously, unless it is the very first state of the universe. As humans, we may perceive the ability to make decisions freely, but where do our thoughts come from? We don't decide to think about something; we just think about it. Or, if we do decide to think about something, where did the thought come from to consider such a decision?

Sam Harris explores this consideration in the aptly titled *Free Will*, where he takes a hard-lined neuroscientific approach to bluntly suggest that we have no free will. In the first section, he states his effective thesis: "Free will *is* an illusion. Our wills are simply not of our own making. Thoughts and intentions emerge from background causes of which we are unaware and over which we exert no conscious control. We do not have the freedom we think we have."[15] He states that "there is simply no intellectually responsible position from which to deny this" and cites experiments proving that the human brain often contains information about decisions several seconds before decisions are consciously made.[16] He also takes care to consider quantum state uncertainty, affirming that even if such states exist inside the brain, nothing them would give it any more capacity to make independent decisions.[17]

Cutting-edge research bolsters the theory that free will may be an illusion created by the brain. In a research article published by *Psychological Science* in April 2016, Adam Bear and Paul Bloom of Yale University conducted experiments focused on understanding "postdictive" acts of the mind, where one feels as if one has made a choice *before* the choice has actually been made. One of their experiments presented participants with five white circles on a screen. Participants were asked to choose one of the five white circles before one at random was changed to red, and

[15] Harris, Sam. *Free Will*. New York: Free, 2012. EPUB file.
[16] Ibid.
[17] Ibid.

they were then asked to report whether they picked the correct dot or not, or if they did not have enough time to formulate a choice. When participants had enough time to formulate choices, they picked the correct circle 20% of the time, the predictable performance for randomized trials. However, when the circle turned red more quickly, participants picked correctly over 30% of the time.[18] In a summary of the study published in *Scientific American*, Bear writes: "This pattern of responding suggests that participants' minds had sometimes swapped the order of events in conscious awareness, creating an illusion that a choice had preceded the color change when, in fact, it was biased by it."[19] He goes on to suggest that this is a feature of the human brain, rather than a bug; by distorting our reality such that choice appears to precede action, (effectively constructing the illusion of free will), we believe that we can have effects on the world and are motivated to do so.

The suggestion that we have no free will is a radical and perhaps discomforting conclusion, but not one without intuitive understanding. If we try to imagine our vast, operating human brain built inside The Game, we can predict with absolute certainty exactly how the brain will progress during each iteration, and we can trace every action taken by the brain to something coming before it. This is the essence of determinism, and though we can't mathematically solve for the ultimate end or death of the universe, (see *The Halting Problem*), if we know the exact address of every particle in the universe and the laws that govern them, we can determine with certainty what will

[18] Bear, A., and P. Bloom. "A Simple Task Uncovers a Postdictive Illusion of Choice." *Psychological Science* (2016): n. pag. *SAGE Journals*. Web. 30 Apr. 2016.

[19] Bear, Adam. "What Neuroscience Says about Free Will." *Scientific American*. N.p., 28 Apr. 2016. Web. 30 Apr. 2016.

happen for the next finite number of iterations. We may have the illusion of choice and decision in our conscious minds, but it is important to consider that it may be precisely that: just an illusion.

III. Conclusion

The Game can be seen as a powerful and elegant intuition pump: a framework for thinking about how a process takes place. To succeed, an intuition pump should be so inherently convincing and intuitive that it's hard to deny its argument. I believe I've turned the knobs about as far as they can go with regards to Conway's Game of Life. What began as a playful metaphor for rule-based life and evolution has quickly grown into a framework that, if you choose to play along, seems to suggest answers to many questions regarding consciousness, intelligence, and free will. I have certainly pumped my own intuition with this thought experiment; when I seriously considered The Game's implications for determinism and free will, I found myself more convinced by the prospect than any other argument I had come across before.

I will not claim that this is an airtight theory of the universe, nor will I attempt to reconcile the apparent randomness of quantum states, or anything of that sort. What I will claim is that the big ideas are covered: the human world has physical laws that govern life, death, and evolution. The Game does as well. From those simple rules, computing completeness is achieved, and if "lifelike" behavior is observed, maybe we should take the hint. It's not so important that we look at The Game's little rule-based simulation and call it *life*. It's important to recognize that *life* might just be a little rule-based simulation. As for determining exactly what the initial game configuration was: I will leave that as an exercise to the reader.

References

Bear, Adam. "What Neuroscience Says about Free Will." *Scientific American*. N.p., 28 Apr. 2016. Web. 30 Apr. 2016.

Bear, A., and P. Bloom. "A Simple Task Uncovers a Postdictive Illusion of Choice." *Psychological Science* (2016): n. pag. *SAGE Journals*. Web. 30 Apr. 2016.

Borowiec, Steven, and Tracey Lien. "AlphaGo Beats Human Go Champ in Milestone for Artificial Intelligence." *Los Angeles Times*. Los Angeles Times, 12 Mar. 2016. Web. 26 Apr. 2016.

Churchland, Paul M., and Patricia Smith Churchland. "Could a Machine Think?" *Scientific American* 262.1 (1990): 32-37. Web. 26 Apr. 2016.

Dennett, D. C., F. Dretske, S. Shurville, A. Clark, I. Aleksander, and J. Cornwell. "The Practical Requirements for Making a Conscious Robot [and Discussion]." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 349.1689 (1994): 133-46. *JSTOR*. Web. 26 Apr. 2016.

Dennett, Daniel. "Are We Explaining Consciousness Yet?" *Cognition* 79.1-2 (2001): 221-37. *ScienceDirect*. Web. 26 Apr. 2016.

Gardner, Martin. "Mathematical Games: The Fantastic Combinations of John Conway's New Solitaire Game 'life'" *Scientific American* Oct. 1970: 120-23. *Ibiblio*. Web. 24 Apr. 2016.

Goodfellow, Ian J., Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, and Vinay Shet. "Multi-digit Number Recognition from Street View Imagery Using Deep Convolutional Neural Networks." (2014): n. pag. ArXiv:1312.6082 [cs.CV]. Web. 6 Mar. 2016.

Harris, Sam. *Free Will*. New York: Free, 2012. EPUB file.

Rennell, Paul. "A Turing Machine In Conway's Game Life." *University of California, Irvine*. N.p., 30 Aug. 2001. Web. 26 Apr. 2016.