Special Issue: Buddhism and Free Will

Journal of Buddhist Ethics
ISSN 1076-9005
http://blogs.dickinson.edu/buddhistethics/
Volume 25, 2018

Buddhist Philosophy, Free Will, and Artificial Intelligence

James V. Luisi

Independent Scholar

Buddhist Philosophy, Free Will, and Artificial Intelligence

James V. Luisi[1]

Abstract

Can Buddhist philosophy and Western philosophical conceptions of free will intelligently inform each other? Repetti has described one possible Buddhist option of solving the free will problem by identifying a middle path between the extremes of rigid determinism, as understood by the hard determinist, and random indeterminism, as understood by the hard indeterminist. In support of this middle path option, I draw upon ideas from the fields of artificial intelligence, quantum computing, evolutionary psychology, and related fields that together render coherent the ideas that determinism may be non-rigid and that indeterminism may be non-random, on the one hand, and upon Buddhist ideas, such as interdependence, the four-cornered negation, and what Repetti describes as the Buddhist conception of causation as "wiggly," to argue that

[1] Independent scholar and author, AI researcher and designer, and IT enterprise architect. Email: jamesvluisi@gmail.com.

Buddhist philosophy has much to contribute to the field of artificial intelligence, on the other hand. Together, I suggest, the Buddhist philosopher and the software expert would form an ideal team to take on the task of constructing genuine artificial intelligence capable of the sort of conscious agency that human beings apparently possess.

Introduction

The field of artificial intelligence ("AI") is undergoing its first important renaissance. This is largely due to the lack of results that contemporary physicists and hardware experts continue to generate, such as with experiments to imitate hundreds of billions of neurons with tens of trillions of synapses that fail to produce intelligence, as if our brains were a uniform neuron-meatloaf that could initiate consciousness and intelligent thought the way a uniform system of binary circuits can run a software program. Recognizing the limits built into this paradigm, AI thought leadership is finally shifting to where it belonged from the beginning: software experts are side by side with philosophers, the latter of whom may be described as paradigm analysts and innovators. For millennia, Buddhist philosophers in particular have been analyzing paradigms of who we are, how we tick, and what our bugs and features are, to use a programming metaphor, and offering radically alternative paradigms in their place. In this paper, I offer some useful comparisons between Buddhist philosophy and AI that, viewed from a novel perspective, promise to shed light on the question of free will in ways that bear on that problem as reflected in a number of the articles in *Buddhist Perspectives on Free Will* (Repetti) and in this Special Issue of the *Journal of Buddhist Ethics*.

Software experts and philosophers, as a team working to create genuine AI, can be thought of in the same light as contemporary scientists

and the science fiction writers who preceded them. As quickly as science fiction writers dreamt up amazing ideas that captured the imagination of readers, a generation of scientists worked to make those dreams a reality. This is not to say that philosophers—Buddhist or otherwise—are no more than the conceptual equivalent of science fiction writers, but to a large extent both require a number of the same talents and proclivities. Philosophers must conceive complex ideas and how they operate within a consistent framework of predictable causal conditions and laws of logical reasoning. Scientists and software engineers share many of the same talents and proclivities. Just as scientists specialize in theoretical and applied sciences, so do software engineers. In software engineering, enterprise architects apply themselves to the theoretical world to envision the capabilities that will propel technology to the next level of efficiency and productivity, while solution architects must deliver these capabilities through a combination of developing new tools and/or leveraging existing tools.

To successfully address some of the challenges of AI, which we can sketch as the attempt to replicate genuine intelligence, minimally, and genuinely flexible agency, maximally, the task of the software expert and philosopher is not easy. To begin, the philosopher must constantly grapple with systems of thought that are in opposition to his or her own well-protected belief system (such as hard determinism or hard indeterminism, to use pertinent examples).

Hard determinism is the incompatibilist view that determinism entails only one future, given the laws of nature and previous conditions, and thus there are no real alternatives, no genuine choices or free will. Hard indeterminism is the incompatibilist view that indeterminism entails randomness, but nobody can coherently claim to author purely random choices. Combined, hard determinism and hard indeterminism entail hard incompatibilism, the view—espoused by Derk Pereboom, Sam Harris,

and others—that either way, whether determinism or indeterminism is true, there can be no free will. The philosopher in the software expert and philosopher AI team that is most likely to succeed must walk a narrow middle path between these two frameworks, which seem to form an exclusive dichotomy that is presently widely accepted as a paradigm.

As both a software expert and an independent philosopher, I suspect that if anyone hopes to design an AI program capable of genuinely intelligent agency, such a program must combine the sort of reliability typically associated with determinism and the creativity typically associated with indeterminism. For this reason, the Buddhist conception of free will that Repetti has identified is the one that the software expert needs most: a middle path between the extremes of the rigidity of (hard) determinism and the utter randomness of (hard) indeterminism. Whereas it should be intuitive how this abstract conception may support research into AI program design, I will spell out how this Buddhist position may be supported by quantum computing and related fields that will play a key role in AI research programs.

How May Quantum Computing Inform Us Here?

Perhaps one of the strongest arguments to the effect that neither rigid determinism nor chaotic indeterminism reflect appropriate paradigms for specifying the causal/functional features of free will can be illustrated with what contemporary AI has learned about quantum computing (QC). As we shall see, there are elements of QC that are deterministic and others that are probabilistic, but their interdependent operations circumvent and thus invalidate what I will call the "hard paradigm" of hard incompatibilism, a myopic view resting on a false dichotomy. It is clearly myopic insofar as anyone stuck in this hard paradigm has no reason to even try to

imagine either genuine AI or genuine free will. And it is clearly a false dichotomy, as the combined deterministic and probabilistic elements of QC make clear.

The use case[2] of QC is for solving certain types of problems in mere seconds or minutes that would otherwise take conventional computers up to trillions of years to compute. In mathematics, these problems are known as NP-complete problems, where NP stands for "non-deterministic polynomial time." The explanation proffered by some that free will may obtain from the quantum level for its randomness is unfortunately misleading, because upon close inspection it is anything but random. The better supported argument is that the quantum effect is part of the same phenomenon that has been detected in QC.

To expound, unlike the conventional computer bit with values of "0" or "1," each qubit in a QC platform has four possible states (i.e., "-1-1," "+1+1," "-1+1," and "+1-1"). This is somewhat analogous to Buddhist four-cornered logic (*catuṣkoṭi* in Sanskrit), which may be described as a "four-fold negation," "suite of four discrete functions," or "an indivisible quaternity." The *catuṣkoṭi* comes in both affirmative and negative versions, where the possible values of the affirmative version are (P), (NOT P), (P AND NOT P), and (NEITHER P NOR NOT P), and the possible values of the native version are NOT (P), NOT (NOT P), NOT (P AND NOT P), and NOT (NEITHER P NOR NOT P).

Let us pause for a moment to reorient after such possibly confusing logic, with a brief "debate" between two neurons.

N1: "Multiple negatives are really confusing."

N2: "You prefer multiple positives?"

[2] DEFINE "USE CASE" HERE.

N1: "Absolutely; their results are always positive."

N2: "Yeah, right."

The proof that the quantum level is not random is that an NP-complete problem with many variables can have an infinite number of possible values for each variable. Hence, the number of possible answers that satisfy a given formula is itself infinite. However, it can be shown mathematically that there is only one value for each of these variables that together yield the optimal result to the problem being solved. Hence, a quantum computer yielding the few near-optimal sets of values or the one and only set of optimal values is clearly not random, as a random process would become forever lost producing an infinite set of non-optimal answers.

To dive in a little deeper in the subatomic world of quantum computing, the programming of a quantum computer requires the aid of a conventional computer. The programming process consists of first developing a mathematical formula to represent the energy state of the system or problem in the form of a Boolean Satisfiability (SAT) problem. This step is analogous to the mind defining the bounds of a problem it intends to solve. The next step is to express the SAT formula within the quantum computer with its many variables that it will be challenged to optimize. Once placed into the quantum computer, each variable is set to an unknown state. Once the program is set inside the quantum computer, which is housed in an electromagnetically shielded container, it is cooled to 20mK, where 0 Kelvin is absolute zero. In comparison, interstellar space is a hundred times warmer at 2.75 Kelvin. The quantum computer then enters into its slowest computational step, "the annealing process," which takes approximately thirty seconds. This step is analogous to the mind pondering the possible solutions to the problem that it has defined. During this process, the variables that had been set to an unknown state

slowly shift to the lowest possible energy state. As the formula moves toward its lowest state, the values of each variable fluctuate in cooperation with one another, first toward a near-optimal solution, and then to the one and only optimal solution.

An interesting feature of QC is that the solution it arrives at has a probability of being the one and only optimal solution. If the annealing process did not have sufficient time (say, if thirty seconds were insufficient), the result would be a near-optimal solution, but just not *the* optimal or one most correct answer. That said, whether the annealing process was of a sufficient duration or not, its resulting answer set is neither purely random nor fully deterministic, but probabilistic. To determine if the result can be considered optimal, the annealing process is repeated multiple times to determine whether the result is consistent, or to determine if additional time may have been required. During the annealing process the answers cannot be observed without stopping the process and having to start all over again. As such, computing at the quantum level is prone to the observer effect in physics, where measurements may not be made without affecting the system (sometimes referred to as "quantum uncertainty"). But quantum uncertainty should not to be confused with the randomness typically associated with indeterminism.

While the neurons in the brain may not operate at temperatures of 20mK with elaborate electromagnetic shielding, the mind's functionally equivalent ability to anneal and arrive at various solutions that are neither deterministic nor random seems intuitively parallel to QC (albeit in a less perfect way, but appearing to function in an analogously probabilistic manner nonetheless). Thus, arguably, the mind is neither rigidly deterministic nor randomly chaotic. It is not rigidly deterministic, contrary to the hard determinist's consequence argument, which claims choices are lawfully necessitated consequences of previous conditions. However, it is still reliably subject to causes and conditions that support a softer form of

predictability governed by probabilistic laws of physics, where the outcome of each consequence can be neither certain nor completely random. Repetti has called this Buddhist conception, which steers a middle path between rigidity and chaos, "wiggly causation." Thus, the Buddhist view of wiggly causation and the view of causation understood in QC are mutually supporting.

Hence, as Repetti noted, the earliest Buddhist philosophers in the contemporary era to write about free will, who tried to describe Buddhist thinking about free will by saying their conception of causation was mostly but not exactly deterministic nor entirely indeterministic, were correct all along. They recognized this fact because without enough determinism there would be no reliable causal links, but without variation there could be no creativity, growth, or change. The Buddhist path requires both: it is concerned with manipulating the causal links to transform oneself into an enlightened, free being. The dichotomous determinism/indeterminism paradigm is misleading, for neither "fully fixed" nor "fully random" explain the events around us.

To break free of this misleading paradigm, we must reject the assertion that the only two choices are either all-rigid determinism or all-random indeterminism. This myopic and misleading paradigm has been vocally advocated by many in the scientific and philosophical communities with confidence, as if this position has been established. It has not. By contrast, those in search of truth will share all possible perspectives, put forth the strongest case supporting each of them, and explain the rationale for the perspective that they have concluded is most sound as supported by logic and observation.

If we accept that our available choices are dependent upon wiggly causation that has led to time t, and that the choice we may make at t+1 has only a probability of being made, then we have a path to becoming an enlightened being if that is what we choose to focus our energies upon,

whether the probabilities were high or low. Once we recognize that prior events merely created the circumstances that surround our next choice, and that they themselves had some probability of occurring, we can better understand that there must be some probability of our next decision, but that it is ours to choose regardless of how optimal or abysmal that choice may be. Otherwise, as Buddhist philosophy in general seems to accept as obviously true, immutably rigid karma would render the changes necessary for enlightenment impossible. This perspective is consistent with what we know about how the quantum level operates within a quantum computer, a perspective that is far more empowering than being trapped in a paradigm that removes all reasonable alternatives from consideration.

Let us consider this wiggly causation conception from another angle. The notion of wiggly causation helps frame a thought experiment of what would happen if the Big Bang occurred in parallel in dozens of universes. Would the same exact thing happen over and over again, and would we have made all of the same identical decisions over and over again in each of those universes? A reasonable answer would be that there is *some* probability that everything would happen in exactly the same way across all of those universes, but what that probability is I do not know. To think it is one (100%) is just to assume total determinism is true. To think it is zero is just to assume that total indeterminism is true. However, based on our current world, I think it is safe to say it is neither zero nor 100%. Ours is a mixed-bag world.

There may be another way to raise a similar question. There is a subtle difference between the multiple Big Bangs question and the question of what would happen if we could rewind our universe to one thousand years ago to watch it unfold once again. Would it proceed in exactly the same way the second time? Again, if we assume the answer is yes, that is, a probability of one (100%), that betrays an assumption of determinism.

The probability that everything could happen in exactly the same way across that same universe as it replays time in a forward direction resembles the probability for multiple Big Bangs; that is, the probability of an identical unfolding within the rewound world would, likewise, not be 100%. Hard determinists, in this analysis, simply assume 100% determinism, but QC and quantum physics in general do not support that assumption.

To the software expert and philosopher team developing true AI, either way it doesn't matter. In fact, it might actually be helpful in the race for developing true AI if the competition, not guided by the insights of Buddhist philosophy, were preoccupied by the myopic paradigm of hard determinism's claim of 100% probability that it will result in exactly the same outcome, by hard indeterminism's claim of a zero probability of the same outcome due to randomness, and by the futility of their combination in the hard paradigm of hard indeterminism.

Thanks to wiggly causation, our brains appear to enjoy the possibility of a similar blend of, if not a collaboration between, mechanical processes and probabilistic processes, the combination of which facilitate free will. If Repetti is right, then Buddhist ethical practices of meditative disciplines increase the extent to which an otherwise largely mechanically operating, highly habituated, and conditioned brain is transformed into a more spontaneously operating, fluid, deconditioned brain that is ceaselessly responsive to the ever-new conditions and circumstances in which it is situated. This is intuitively what we would like an authentic AI to be like.

When we peer into our imaginations, various voices of reason pull us in different directions, each by distinct paradigms and beliefs, each acting in a manner consistent with its corresponding paradigm: we weigh the potential drawbacks and benefits with our desire, and we make a free decision to choose an action or no action at all. What is important is that we

make that decision based upon either our reasons, or based upon simply what we want regardless of what the reasons may have concluded if weighed mechanically.

As probability would have it, ironically, the Buddhist philosopher is the one who the software expert needs most: the Buddhist philosopher can see the limiting paradigms, rise above them, and guide the software expert away from the mind traps and toward the objective with the optimal set of solutions, like the quantum computer. The causation that explains our world, our brains, our free choices, and thus the possibility of genuine AI is neither entirely deterministic nor entirely indeterministic, but a wiggly collaboration between both.

The Fallacy of the Heap

Another useful perspective for the efforts of the software expert and philosopher AI team, and for the question of free will from the Buddhist perspective, is formed by an analysis of the fallacy of the heap. This fallacy makes the error of concluding that, just because one cannot identify some number of, say, grains of sand, N, such that N is not a heap, but N+1 is a heap, there is no such thing as a heap. But there are obviously heaps, in which case this inference is clearly erroneous. Here the issue is one that the software expert understands from large interactive computer systems that are constructed in a manner similar to a self-contained operating system. The reason that the analogy with an operating system is at all pertinent is that an operating system has modules that specialize in certain capabilities, and then collaborate with the other modules much like the modules of the brain specialize in certain capabilities and then collaborate with the other modules of the brain, with incredibly complex interdependencies that generate abilities exceeding those of the mere sum of the module parts.

Some philosophers deploy a line of reasoning that loosely resembles the heap fallacy when they reason that just because neuroscientists cannot identify, say, the specific number (or pattern, cluster, location, etc.) of neurons at which point the accumulation of which there is consciousness, there is no such thing as consciousness. Though we cannot specify any such N, it is clear that there comes a point at which there is consciousness, and thus at which the whole brain is greater than the sum of its parts, which can collaborate in such effective ways: they obviously do establish consciousness and the ability to make decisions using reasons that free the brain's otherwise separately limiting modules from instinctual and hard-wired responses. The brain has given rise, in other words, to conscious agency or free will.

While physicists and hardware experts have been attempting to create AI with the most simple form of what may be described as the construction of a heap, by simply adding an ever increasing number of simulated neurons and simulated synapses into a computer hardware network, this can only take intelligent form if the heap can be arranged into modules that specialize and collaborate to deliver increasingly advanced capabilities that work together in the way that they do in our brains, e.g., pattern recognition, importance determination, analogical determinations, etc. (Luisi).

If we look at what may be described as the “heap paradigm” from the perspective of the evolution of intelligence as it appears, for example, in the different levels of intelligence of insects equipped with simple neural bundles, far before the formation of specialized modules within the brain, we see that the common honey bee represents a myriad of advancements over the wasp. As one example, unlike the wasp that will infinitely repeat a task until it is complete, potentially to its death, the honey bee has a sense of time, where, after approximately ten minutes, it will move on to its next task rather than being stuck in a hard-wired loop. As another

example, unlike the wasp the honey bee memorizes its flight path to food sources based upon the relative position of the hive and the sun, and upon returning to the hive the honey bee will convert the flying instructions into a dance that functions as a form of language. The other bees observe the dance, memorize it, and then convert it into flying instructions that guide them to the source of food.

Some seem to presuppose a heap paradigm in the foundations of hard determinism, assuming that if determinism is true, there can be no point at which neural complexity can generate the sort of genuine flexibility that could count as free will. One way to consider a refutation of the heap paradigm is with the following "debate" between two neurons that coexist in one of the first neural bundles, where Neuron 1 represents hard determinism and Neuron 2 represents soft determinism:

> N1: "I've studied the physics of the iron atom and conclude it is impossible for iron to become an engine that can create motion. The Big Bang created stars, which created iron atoms, then the stars exploded, and the resulting dust formed planets. The resulting iron atoms just sit wherever they were left."
>
> N2: "What about a complex collection of them? Can't enough iron atoms be configured in a way to create motion?"
>
> N1: "No. If you add a single iron atom to any collection of them that is not in motion, all you still have is a motionless pile of iron atoms."

While the hard determinist is in alignment with the physicist and hardware expert in their approach to creating artificial intelligence, what they fail to realize is that variability in software design matters. While it is true that both are required, it is the software that drives the hardware, not the other way around. The notion of simply adding neurons and synapses into a massive uniform container makes as much sense as believing

that human intelligence could emerge if the brain were one uniform mass of neural meatloaf, without modules developing that specialize and collaborate in a meaningful way.

In fact, the human brain has evolved dozens of specialized modules, such as Broca's Area, Wernicke's Area, the Motor Cortex, Primary and Secondary Visual Areas, and Olfactory Areas. These specialized modules collaborate to produce a result that the individual modules could not produce on their own. As an example, Wernicke's Area allows understanding of spoken and written language, and if damaged it causes the person to become unaware of their own speech and the speech of others. Broca's Area in the left frontal region receives impulses from Wernicke's Area and converts them into the motor commands that operate the muscles in the motor strip controlling the tongue and lips to speak clearly in the desired tone and volume. One of these specialized areas without the other eliminates either the ability to speak words or the ability to string a sensible sequence of words together.

The Buddhist paradigm of interdependence applies here as an argument against the hard determinist's attempt at a one-to-one reductionism: just because such a reductionism doesn't work is no reason to conclude that the holistic abilities that rest on these interdependencies do not exist. Thinking otherwise involves some sort of reverse heap fallacy: we cannot reduce the holistic interdependencies of conscious agency to some single neural component "N" for each such element of conscious agency and thus conclude there is no conscious agency.

During the long journey of evolution, a uniform meatloaf of a brain may have been attempted, but it obviously did not prove superior to the emergence of collaboration of specialized modules. The same effect may be observed in the emergence of packs of hunting animals, or in the formation of human teams that work particularly well together, where individually they could not assemble the skills and resources to accomplish

the same results. The Buddhist philosopher thus may disagree with the hard determinist who asserts that the process of adding neurons is simply extending the size of the deterministic, pachinko-machine-like mind-brain, never to create free will. Instead, the Buddhist paradigm of inter-dependence is consistent with the emergence of specialized modules that have the ability to collaborate in ways the hard determinist simply cannot imagine.

What Does the Evolution of Intelligence Inform Us about Here?

The world of the determinist is deeply rooted in the physical world, the world of physics and hardware, whereas the software expert is rooted in the world of ideas, decisions, and intellectual content, and hence more closely aligned to producing faculties of thought and reason. No matter how clever the hardware designer is, the design of a television can only provide a representation of the programming content being streamed into it. The television hardware by itself can neither create an episode of *I Love Lucy* nor *Game of Thrones*. It is the content that is streamed into the hardware as software encoded in the form of either analog or digital electronic signals that embodies the story, imagination, and drama of the programs that we enjoy and that determine ratings. Let us hear this debate played out between those neurons:

> N1: "As a physicist I can see specific areas of the brain light up depending upon what the subject is going to do just before he does it. It is all mechanical, like a pachinko machine where you can simply observe the outcome as the mechanical physics play out."
>
> N2: "So when you observe the physical brain, can you see the various thoughts and reasons being weighed inside the neurons?"

> N1: "The reasons and thoughts don't matter at all, it's all in the physical neurons. Everything operates in a universe of hard determinism, and as such, it is all determined by the physics of the neurons, even the thoughts. Let me show you a video that explains this."
>
> N2: "Great! The chair in front of the monitor is mine, and the chair in back is yours, so you can observe the hardware."
>
> N1: "But watching the hardware will not enable me to see the video."
>
> N2: "Hmmm. Your physical hardware is working just fine, but looks like your software has a bug."

A better television design might render the images in 3-D, or even in virtual reality, but the content is what captivates and drives the imagination, as it is with our lived experience. That said, we still need physicists and hardware experts to design better hardware to deliver content to our senses. But the skills for understanding the mind are in more capable hands when the Buddhist philosopher and the software expert apply their skills and specializations.

We would not want a brain surgeon to address the latest issues in the Windows operating system, and we wouldn't want a software developer performing brain surgery. Likewise, we would not want a Buddhist philosopher to repair our television, and we wouldn't want a neurologist or physicist to guide us along the path of enlightenment. As a software expert I understand, within the many voices and paradigms in my mind, and with my many years of experience in software and artificial intelligence, that the philosopher trained in Buddhist thought is what the software expert needs most, as it is the Buddhist philosopher who understands that our hardware is not the difference between the enlightened

and unenlightened individual. The same reasoning holds for conscious agency.

Works Cited

Harris, Sam. *Free Will*. Free Press, 2013.

Luisi, James V. *Sensitive by Nature: Understanding Intelligence and the Mind*. AuthorHouse, 2002.

Pereboom, Derk. *Living without Free Will*. Cambridge University Press, 2001.

Repetti, Rick, editor. *Buddhist Perspectives on Free Will: Agentless Agency?* Routledge, 2017.