Meditation for Machine Consciousnesses:
What Buddhism Can Tell Us About the Politics of Near-Future Technologies

Matthew J. Moore
California Polytechnic State University, San Luis Obispo



Matthew J. Moore
Professor
Department of Political Science
Cal Poly State University
1 Grand Avenue
San Luis Obispo, CA 93407-0328

805-756-2895
mmoore02@calpoly.edu

We live in interesting times with regard to technology. We are able to imagine, and almost able to see to how to create, technology that would fundamentally transform human existence. Some of these allegedly near-future technologies (NFTs) already exist (genetic engineering) or are close enough that it seems certain that we will be able to develop them through the ordinary course of technological improvement (robots capable of performing many tasks at a human level). Others remain speculative both in the sense that they don't exist yet and in the sense that we have no clear idea of how to develop them, though we have today what might ultimately turn out to be the primitive progenitors of such technologies (strong Artificial Intelligence, nanobots, and Whole Brain Emulation).

These NFTs stand out from other technologies for two important reasons. First, if they turn out to be possible, they promise the ability to act upon the world and ourselves to an unprecedented degree. Allow yourself a moment of sci-fi geek excitement, and imagine: a world where robots replaced dangerous and tedious labor, creating an inexpensive abundance of goods that humans with ample free time could enjoy; a world where debilitating genetic illnesses no longer afflicted millions; a world where computers that dwarf today's in speed and storage enabled us to solve the hardest problems in the natural sciences; a world in which virtually any illness or injury could be prevented or cured by microscopic nanobots constantly monitoring and repairing your body from the inside; perhaps even a world where human consciousness could be uploaded into a computer, and in that way transcend the limitations of our bodies and mortality.

The second way in which these NFTs are different from most other technologies is that they pose an unusual risk of escaping from our ability to control them, with potentially existential consequences. Thus: labor-replacing robots could result in massive unemployment and exacerbated inequality; genetic engineering could result in profound, heritable inequalities as

well as an arms race in which the wealthy seek constant changes to obtain fleeting relative advantages; the development of strong (or sentient) Artificial Intelligence would mean that we would be sharing the planet with another intelligent species, one that would be much smarter, faster, and more rapidly self-improving than us; nanobots could be hacked or programmed to cause harm, and self-replicating nanobots could replicate out of control; finally, merging human consciousness and computer platforms seems likely to result in unknowable changes in what it means to be human. I use the term "robopocalypse" to stand for all of these possible bad outcomes.

In this paper, I argue that there is a distinctively Buddhist approach to analyzing social and political issues, and that it offers a unique and uniquely helpful analysis of NFTs. The paper is part of a larger project whose current working title is *Mindfulness, Politics, and the Robopocalypse*. The overall argument of the larger project is that the various near-future technologies pose a variety of possible benefits and dangers, and that an approach to politics that is informed by mindfulness—the increasingly popular practice of non-judgmental awareness of present-moment experience that has been adapted from Buddhism--may be uniquely capable of helping us navigate those choices successfully. Although mindfulness practice does not lead to or especially support any particular ideological or partisan position, it does potentially have political implications.[1] First, mindfulness may help to make each of us calmer, less reactive, more deliberative citizens and consumers. Second, a polity that embraced the value of mindfulness practice would seek to nurture the social conditions that make mindfulness possible and obtainable, such as ensuring the ability of all of its citizens to meet their basic needs, attempting to ameliorate sources of social conflict and distress, and otherwise making sure that people have the time and energy to devote to being mindful citizens. That situation—mindful

citizens supported by policies aimed at encouraging mindfulness—seems to me to be the best-case scenario for having decisions made according to reason and principle, rather than being driven by impulse and reaction. Every plausible strategy for avoiding the robopocalypse requires humanity to decide not to do something tempting, and mindfulness seems to me to be the only hope for cultivating people who are able to do that. I'm not saying that mindfulness definitely will prevent the robopocalypse, but I am saying that it may be the only thing that can.

**An Approach to Buddhist Social and Political Analysis**

As you probably already know, Buddhism is a religion / philosophy taught by a man named *Siddhattha Gotama* some time between the fourth and sixth centuries BCE. After he achieved enlightenment, the Buddha began to preach about what he had learned. In his first sermon, he explained two ideas that became the basis of the religion: the Four Noble Truths and the Noble Eightfold Path. The Four Noble Truths are:

> *Dukkha* (suffering) – That life is persistently unsatisfactory; that there is no way to ensure that life does not contain unhappiness and dissatisfaction.
>
> *Samudaya* (origin) – That *dukkha* arises because of a conflict between what our minds want and what the world gives us. We want x but get y, and are resentful. We don't want z but get z anyway, and are angry. We want y and get y, but eventually lose it again and are broken-hearted.

*Niroda* (cessation) – That we could avoid *dukkha* if we could accept what the world gives us and not cling to what our minds want or flee from what they don't want.

*Magga* (path) – That we could learn to stop clinging to expectations by following the Noble Eightfold Path, which is cultivating: (1) right understanding; (2) right intention; (3) right speech; (4) right action; (5) right livelihood; (6) right effort; (7) right mindfulness; (8) right concentration.

For the purposes of social and political analysis, I find it helpful to reframe these basic ideas. First, the Buddha distinguished between pain (a bodily experience) and suffering (a mental experience), and argued that while some pain was inevitable in life, suffering was optional and could be overcome. Second, Buddhism rests on a simple and self-evident observation: all the different varieties of suffering, from boredom to anguish, have the same cause—the world gives you something different from what your mind wants. When that happens, there are three possible courses of action for you to pursue: change the world to match your mind; keep suffering; change your mind to accept the world.

We are always free to choose to keep suffering, but the Buddha thinks that no one who was aware of the choice would do so, though our ignorance usually prevents us from seeing the choice at all. Assuming for the sake of argument that we can see the choice in some particular circumstance, we are left to decide between changing the world and changing our mind. How should we decide which choice to pursue? Although the Buddha never says this directly, the answer seems obvious: choose the course of action that leads to the least suffering overall.

Traditionally, Buddhists have believed in karma and reincarnation, so that the choices you make in this life shape your next incarnation, for good or ill. The point of such a system, obviously, is to ensure moral equity—bad deeds always redound to your harm, and good deeds to your benefit. In such a system, acting so as to increase the suffering of others always ultimately increases your own suffering, and acting to decrease the suffering of others ultimately decreases your own suffering. Further, by decreasing your own suffering you become less likely to harm others (and thus increase their suffering), and by decreasing the suffering of others, they become less likely to harm you (and thus increase your suffering). The upshot is that overall actions that affect one's own suffering should have the same effect on the aggregate suffering of all sentient beings.

Many modern-day Buddhists in the West interpret reincarnation as a metaphor rather than taking it literally, and argue (or assume) that the consequences of your actions will ultimately affect you in this life. And such Buddhists seem to believe that it is still the case that what affects the suffering of the individual affects the aggregate suffering of everyone else (on average, in the long run), and vice versa. On one level, that's obviously false: any binary policy choice will create winners and losers, those who don't suffer from the outcome and those who do. But that may be too simple an analysis. I may genuinely be pained by losing a policy debate, but not as pained as I would be to lose the voluntary social cooperation of the winners, and thus I may be able to accept my loss with equanimity. Similarly, while it's probably true that we can imagine situations in which a majority oppresses a minority such that the majority's aggregate suffering is reduced more than the minority's aggregate suffering is increased, Buddhist social analysis encourages us to look at the long run—will simmering resentments lead to civil war? to generations of crime and desperation?

More generally, this Buddhist perspective appears to be a variant of consequentialism, though with the interesting twist that only suffering counts—increasing happiness isn't part of this utilitarian calculus. I don't have time here to try to work out fully whether this Buddhist semi-consequentialism can be defended against the standard critiques of consequentialist ethics, so for the moment I want to assume that it make sense at least at first glance (being a jerk will eventually come back to bite you, and being nice will make the world slightly nicer to live in), and see where it takes us.

Following this logic, if you are suffering, sometimes it is most appropriate to change the world. If you are cold, put on a sweater. If you are hungry, eat. If you are tired, sleep. The same calculation appears to operate in the realm of collective action as well. People acting as groups should take those actions that will minimize the aggregate suffering of the group (and anyone affected by the group's actions) in the long run. Thus the ancient Buddhist texts argue that government is necessary to prevent theft and violence, and imply that creating and maintaining government would cause less aggregate suffering than not having one.[2] Further, they argue that policies preventing extreme poverty cause less suffering than would allowing some people to live in economic distress, which would inevitably lead to crime and violence.

By extension, we should choose to change our minds when that course of action is the one that leads to the least suffering overall. Like the Stoics and Epicureans, the Buddha argues that choosing to change one's mind is the path of least suffering in the overwhelming majority of cases. The method he taught for doing that consists of the eight courses of action mentioned above, and now we can see that what makes them "right" is that they lead to less suffering rather than more. Thus, for example, right speech consists of telling the truth and avoiding idle chatter, which seems like good advice for minimizing suffering.

The Noble Eightfold Path is traditionally divided into three groups: 1. *paññā* (wisdom) – understanding and intention; 2. *sīla* (virtue) – speech, action and livelihood; 3. *samādhi* (~meditation) -- effort, mindfulness, and concentration. In essence, *paññā* is about seeing the world and oneself correctly, *sīla* is about how to change the world, and *samādhi* is about how to change one's mind. The basic method of *samādhi* is to learn to accept the world as it is, by developing mindful awareness and non-judgmental acceptance of one's experience, both inner and outer. Meditation is how one develops these capacities.

In other teachings, the Buddha explains both how and why to meditate. In the *Satipaṭṭhāna Sutta*, the Buddha identified mindfulness (*sati*)—non-judgmental awareness of the present moment—as being an especially helpful path towards achieving *samādhi*. The Buddha describes how mindfulness arises from four foundations: awareness of the body (sensation); awareness of feeling (emotion); awareness of mind (thoughts); awareness of "phenomena" (*dhammās*). Here's a long-ish quote from the *Satipaṭṭhāna Sutta* in which the Buddha explains the practice and purpose of meditation:

> "Monks, this is the one-way path for the purification of beings, for the surmounting of sorrow and lamentation, for the passing away of pain and dejection, for the attainment of the true way, for the realization of Nibbāna [Sanskrit: Nirvana]—namely, the four establishments of mindfulness. What are the four? Here monks, a monk dwells contemplating the body in the body, ardent, clearly comprehending, and mindful, having subdued longing and dejection in regard to the world. [The same formula is repeated for feeling, mind, and phenomena.]….And how, monks, does a monk dwell contemplating the body in the body? Here a monk, gone to the forest, to the foot of a tree, or to an empty

hut, sits down; having folded his legs crosswise, straightened his body, and established mindfulness in front of him, just mindful he breathes in, mindful he breathes out. [Similar instructions are given for feeling, mind, and phenomena.]" (Bodhi 281-82)

In essence, meditation is the process of becoming aware of what our bodies and minds are doing when we are not intentionally trying to do anything. We learn to accept what is happening in our bodies and minds because we cannot control it, and indeed because trying to control it makes it worse (louder, more chaotic, harder to control). And by learning to accept what happens in our inner world, we can learn to accept what happens in the outer world. As above, most of the time our best strategy for reducing suffering will be to change our minds, and meditation is concentrated training in doing just that—changing our minds to accept rather than resist what we cannot change.

Although we rarely say these things out loud, meditation rests on a set of assumptions about the kinds of beings that will be meditating. First, meditation is a practice designed for consciousnesses that are capable of a high-degree of self-awareness, but that are also in the habit of not attending to most of their experience, either internal or external, and that may require extensive training and special conditions to be able to become aware of some aspects of their experience at all. We might say that meditation is designed for the *partially aware*. Second, meditation is for beings that can only control some of their experiences, both inner and outer, such that some (perhaps the vast majority?) of their experiences will always be out of their intentional control; thus we might say that meditation is for the *partially intentional*. Third, meditation is designed for consciousnesses that can *suffer*. That suffering need not be physical—indeed, according to the Buddha, suffering (as distinct from pain) is always mental or

psychological—but the consciousness must be capable of having conflicts between its preferences or intentions and the world around it, and it must experience those conflicts as being more than merely a logistical problem to be solved. The conflicts need to make it unhappy, or frustrated, or angry, or depressed or bring about some other undesirable state.

A final assumption about meditation is that when we pay non-judgmental attention to our present-moment experience, we will learn things about ourselves and our experience of the world (both inner and outer) that will help us to become more able to accept the world as it is and not suffer from its failure to be as we would have it. One common experience in meditation is realizing that we have been semi-consciously fixating on something (a thought, a sensation, an emotion). Another common experience is that our bodies are experiencing tension or anxiety that we haven't registered consciously yet. Noticing those experiences, and accepting them without tying to change them, helps us to recognize patterns and habits that may be causing us a lot of suffering but that have been just below the threshold of conscious awareness. By learning to accept what is, we learn to see more of what is, and can learn to accept more and on a deeper level.

**Why Are We Developing These Near Future Technologies?**

From the Buddhist/mindfulness frame of analysis, we are developing these technologies because we implicitly believe that doing so will have more benefits than costs, and that it will make us better off in the long run. In other words, we apparently and implicitly believe that developing NFTs will reduce our suffering. What are the kinds of suffering that NFTs will ostensibly help with? Some are obvious and non-controversial: suffering caused by not having

the goods and services one needs, which could become more available through improvements in efficiency and productivity that NFTs might make possible; suffering caused by dangerous, injurious, tedious or otherwise harmful work that is currently necessary but that could be improved or even alleviated by the use of NFTs; suffering caused by inequality, which might be ameliorated by significantly increased production; suffering caused by illnesses and injuries that NFTs could better prevent, treat, or cure than we can do today; suffering caused by the lack of sources of pleasure, joy, entertainment, and wonder that might be possible with NFTs.

Some other kinds of suffering that motivate the effort to develop NFTs are not surprising, but may be more controversial: suffering caused by unfulfilled curiosity about what is possible and what exists beyond our current knowledge and capacities; suffering caused by fear that humanity and all of its achievements will ultimately be destroyed (by the death of our sun, for example) unless we find a way to send our bodies or our intelligence to other parts of the universe; suffering caused by not personally having a major impact on history; suffering caused by fear of humanity's failure to play its part in bringing about the evolution of a better species, and thereby moving the universe towards greater perfection.[3]

And some of the kinds of suffering are in essence suffering caused by being human, as we currently understand humanness: suffering from mortality and fear of death; suffering from being physically vulnerable; suffering from illness, injury, and aging; suffering from the limitations of our abilities to communicate with others; suffering from the need to compromise with others; suffering from the limitations of our brains, such as their limited and faulty ability to learn and remember; suffering from the limitations of our bodies, such as their limited strength or their need to sleep.

I think that it would be easy to be dismissive or critical of some of these sources of suffering, but I also think that that would be a mistake. Suffering is suffering. If someone is suffering from fear that they will not have a big enough impact on human history, which we might be tempted to criticize as megalomania or cultural privilege gone wild, we should treat that as just another source of suffering, in  part because dismissing someone's suffering is likely to increase it.

If we imagine the choice of whether or how to develop NFTs as being a balance scale, at the moment one pan is weighed far down with all of the suffering that NFTs might be able to diminish. Into the other pan we now have to put all the suffering that creating NFTs might bring about. Each NFT has its own particular dangers, but for the moment we can just talk about the dangers that they all share:

1. unequal access to the benefits of NFTs, both direct (whatever useful thing the NFT does or produces) and indirect (the profits to be made from NFTs and related enterprises);

2. that the NFTs, nearly uniquely among human technology, pose the risk of escaping from our capacity to control, with unknowable consequences;

3. disruption of existing patterns and institutions (employment, the family, etc.);

4. that some NFTs would expand our toolkit of means of killing all of humanity, either accidentally or deliberately;

5. that some NFTs (especially genetic engineering) pose the moral hazard of creating new collective action problems (for example, creating an arms race in which people seek more and more genetic changes to obtain a brief relative advantage over others);

6. that NFTs pose the danger of creating inequalities of power so dramatic that they would be all but impossible to overturn.

For the moment, I don't want to weigh in (if you'll forgive the pun) on which pan is heavier—that's a conclusion for the bigger project—but rather want to point out that one pan represents the aggregate suffering of deciding to just keep suffering (by not developing NFTs) and the other represents the aggregate suffering of trying to change the world (by developing NFTs). The third strategy, changing our minds, would have the effect of removing weights from the just-keep-suffering pan altogether. We learn to accept the world as it is through meditation—through the deliberate and sustained practice of accepting our own experience. The goal is to identify the pan that is the lightest, and changing our minds is one way to lighten the option of not developing NFTs.

One version of the robopocalypse is the idea of the Singularity, championed by Ray Kurzweil (Kurzweil), who argues that in the next few decades technology will develop so rapidly that it will transform human existence. Two important factors in that transformation will be the development of strong (sentient) AI and the ability to merge or transfer human consciousness onto machine-based platforms (WBE). Kurzweil argues that humanity will in effect merge with technology, thereby humanizing machine sentience (which otherwise poses a grave danger[4]) and simultaneously transcending our biological limitations. I want to ask whether that's a good idea,

by beginning to ask what might be lost in such a merger. I argue that machine consciousnesses (AIs and WBEs) would be very unlikely to meditate, and that by choosing to develop consciousness in that direction, we would be giving up meditation as a tool. In other words, to try to alleviate suffering we would be permanently choosing to have only changing the world as our strategy, and by making that choice we would be trapping ourselves.

## Machine Consciousnesses

What I mean by a machine consciousness is a computer that is conscious—that is, that it is aware of and capable of thinking reflexively about itself. I hasten to say that there are not any machine consciousnesses today, and it's not obvious that we have any clear idea about how to create one. Indeed, many smart people think that machine consciousness is impossible even in principle. On the other hand, lots of other smart people think that machine consciousness may be possible, and there are certainly thousands of people working on trying to create one. I'm not going to try to resolve that question today, and will assume for the sake of argument that it is possible, so we can get on with speculating about the consequences. Among people who think that machine consciousness may be possible, there are two basic paths they think might get us there: 1. strong AI – creating computer programs that become self-aware; 2. whole-brain emulation – creating machine emulations that are able to reproduce or support every function of the human brain, including consciousness.

### *Strong AI*

Of course you've heard of AI, but it's worth saying out loud that AI is computer programs that can do various mental tasks that humans can do, up to and perhaps including being self-aware or conscious. I find it helpful to break AI into three sub-types:

(1) **Narrow-Weak AI** is programs that can do one or a few human tasks well. This is what we have now; examples are programs that can do voice recognition, read your lousy handwriting, respond to simple requests (usually for information), and win at chess, Jeopardy and Go.

(2) **Weak AI** is programs that can do most human tasks well, but are not self-aware. We don't have programs like this yet, but it seems very likely that they are possible, and that they will become available in your lifetime. We can see more or less how to create them, and have good reason to think that likely developments in technology and programming will make them possible.

(3) **Strong AI** is programs that can do most or all human tasks well and that are self-aware. We don't have programs like this yet, and we don't know how to create them. Indeed, we don't even understand where human consciousness comes from, and so don't have a model to try to replicate. If strong AI is possible, there are good reasons for thinking that it would be very dangerous. Strong AI would be able to reprogram itself (or its successors), and would be able to incorporate or access most human knowledge. In other words, it would gain in intelligence and ability extremely quickly. In short, strong AI would be smarter than you, faster than you, and able to improve much faster than you.

If such a thing were created, the Earth would probably swiftly belong to it, and we would just have to hope that it liked people.

But this paper isn't about strong AI destroying human civilization, but rather about a narrower question: would it meditate? Remember that meditation is for beings that are partially aware, partially intentional, and can suffer. How would any of that be relevant to a strong AI machine consciousness? Let's admit up front that we're in speculation-land here, since there are no such consciousnesses and we have no idea whether they're possible or how to create one. Nonetheless, I think you'll see that we can draw some reasonable inferences.

AIs will probably not be only *partially aware* (of themselves). It's hard to imagine that anyone (either a human or an AI) would program an AI to be able to access information that is available to it only with great effort or under special, rare circumstances. Obviously no program can be simultaneously processing every available source of data. But it would defeat the basic purposes of computation to make certain kinds of data difficult or impossible to access on demand. Thus, it seems very unlikely that strong AIs would have anything analogous to the human unconscious—the realm of experience and knowledge of which we are inconsistently and incompletely aware.

AIs will also probably not be partially intentional (that is, they will probably not be doing or experiencing lots of things without choosing to do so or being able to exercise intention over the experiences). It would be difficult to program a computer to function like that, and it's hard to see what the benefit would be.

Will AIs be able to suffer? Here the answer is more complicated. It seems very unlikely that AIs will be able to experience pain. Again, it's hard to see what the value of programming

an AI to experience pain would be. But if an AI is sentient—that is, if it is capable of thinking reflexively about itself—then it seems likely that it would avoid death and pursue continued existence, and that it would view the threat of future death or harm as being something to be avoided even at great cost. The question is: does that amount to suffering? I think the answer is: only if the AI can experience emotions. If imminent danger of death is merely a logistical problem to be avoided, the entity isn't suffering. The entity needs to experience a persistent non-desired state because of the danger, and that requires emotion.

While we have no idea whether strong AIs would have emotions, we have good reasons for thinking that experiencing emotions would at least not be part of an AI's intentional programming. I think we can infer this for two reasons. First, many (most? all?) human emotions have their roots in the body. A body-less strong AI would have no need for a fight-or-flight reaction, no need to bond with its parents or to demonstrate loyalty to a group. It would also not have a body that has those experiences independent of conscious intention. (And if an AI did need to do those things, it could learn how to fake them.) Second, emotions are in conflict with the logic-focused methods of computer reasoning. Imagine a computer program that is supposed to compare two numbers and print out the larger of the two, but that develops a strong emotional attachment to the smaller number and keeps printing out that one instead. More generally, experiencing emotions (as distinct from recognizing or being able to analyze the emotions of others) would introduce a degree of arbitrariness into a computer program that would self-defeatingly make the program worse at exactly the kinds of tasks that computers are great at. Thus, it seems very unlikely than an AI would experience emotions, and thus very unlikely that it would be able to suffer.

For all these reasons, it seems safe for us to infer that an AI would be very unlikely to meditate. It experiences no suffering to learn to minimize, and it has no semi-aware or semi-intentional experience to learn to notice and accept. What would be the point of meditating?

**WBEs**

What about Whole Brain Emulations (WBEs)—human consciousnesses uploaded to machine platforms? (For the moment, put aside the question of whether it's possible.) Would they be likely to meditate? To answer, we again need to speculate a little. Imagine that we were able to upload your consciousness to a computer. One immediate problem would be that there would no longer be any information coming into your "brain" from your body. When that happens, it seems to me, there are two basic possibilities: (1) you will freak out because your consciousness needs that input, and we will need to find a way to fake it; (2) you would not freak out because it turns out that your consciousness can get along just fine without having or experiencing a body.

If the second case obtains, then obviously WBEs would no longer have access to semi-aware, semi-intentional bodily experiences to focus on in learning how to accept what they cannot control. If the first case obtains, roughly the same thing happens because even though you would still be experiencing bodily sensations, etc., they would all be the result of deliberate choice, and they would teach you nothing about your own experience. Thus, for example, if we programmed your experience of "breathing" to be always calm and regular, or irregular and random, you would have something to pay attention to, but even careful attention would teach you nothing about yourself because the "breathing" has nothing to do with your consciousness.

(And if it did, for example by distracting or disrupting your thought process, presumably you would just reprogram it so that it didn't cause those problems.) Either way, you would lose non-judgmental present moment awareness of your body as a focus for training yourself to accept what you cannot change. The experience would be roughly the same as focusing on the ticking of a clock.

But perhaps that isn't so bad, since you would still have your unruly and chaotic mind to focus on. That's probably harder than focusing on your body, so it would be harder to learn to meditate, but it should still in principle be possible. Except...that proponents of WBE tend to see faithful replication of our existing consciousness as merely the first step in a longer process of changing and improving the ways our minds work. Consider these quotes from Ray Kurzweil:

> [T]here is only so much room in our skulls, so although Einstein played music he was not a world-class musician. Picasso did not write great poetry, and so on. As we re-create the human brain, we will not be limited in our ability to develop each skill. We will not have to compromise one area to enhance another. (Kurzweil 202)

> The most important application ... will be literally to expand our minds through the merger of biological and nonbiological intelligence. The first stage will be to augment our hundred trillion very slow interneuronal connections with high-speed virtual connections via nanorobot communication. This will provide us with the opportunity to greatly boost our pattern recognition abilities, memories, and overall thinking capacity, as well as to directly interface with powerful forms of nonbiological intelligence. The

> technology will also provide wireless communication from one brain to another. (Kurzweil 316)

> [I]f we are diligent in maintaining our mind file, making frequent backups, and porting to current formats and mediums, a form of immorality can be attained, at least for software-based humans. (Kurzweil 324-25)

In more general terms, WBE enthusiasts want to reshape the human mind in ways that make it more like a machine and less like a biological organism/system. For the moment I want to put aside the questions of whether that is in some ways desirable, or even coherent (human memory is very different from computer memory, and it's not at all obvious that the one could be made to work more like the other) to stay focused on the question of meditation. It seems to me that a WBE that is changing to be more like a computer and less like a biological system may be subject to several possible problems that would interfere with meditation:

Too Fast – Kurzweil dreams of making our minds much faster. But we don't understand what gives rise to our capacity for conscious recognition and deliberation. We don't know whether that capacity can handle information moving at higher speeds. Thus there may be a danger that we will increase the noise and chatter in our heads to a point where we can enjoy some benefits of greater speed, but also find it much harder to quiet our minds when we want to.

Too Much -- The idea of wireless communication among human minds sounds great, until you think about it. Even apart from some profound questions about whether the idea even makes sense (it isn't obvious to me that my mind contains "thoughts" that could be communicated until I intentionally choose words to speak or write), there is the more basic

question about whether our minds could handle all of the additional information and decision-making. I assume that mind-to-mind communication will be voluntary—something that you have to agree to do. But will there be persistent requests, a phone ringing in your mind forever? Or perhaps helpful notifications that seven people with whom we wirelessly communicate are currently available to mind-chat? Anyone who has a smartphone can predict the effects of making your mind into a smartphone—too many sources of possible distraction.

Too Little – One question that WBEs could finally settle definitively is how much of our "mental" experience is really a semi-conscious representation of our bodily experience. My suspicion is that it's an enormous percentage. If that's right, then we face a replication of the body problem outlined above—either we fake those experiences, in which case the resulting mental experiences are meaningless because they are merely a kind of mental white noise needed to keep our brains from freaking out, or we don't fake them, in which case a great deal of our unconscious, nonintentional mental experiences would disappear and no longer be available as a focal point for meditation. That might seem as if we would instantly have reached nirvana, since with the flip of a switch we could silence most of our "monkey mind" chatter. But the danger is that the opposite would happen, and we would be left only with the conscious, intentional aspects of our mental lives (and perhaps some associative bits and pieces), which would be unsuitable for meditative focus or learning how to accept what we cannot control.

Too Flat – I also suspect that WBEs would reveal that the vast majority of our emotional lives are also rooted in our bodies. What would fear be without the surge of adrenaline and cortisol? What would love be without pheromones and dopamine? Again, we face the choice of either faking those experiences and thus making them arbitrary and meaningless, or losing them

as a focus for meditation. (I feel compelled to say that losing the capacity to feel emotions seems bad to me for reasons greater than its interference with meditation.)

We can sum up all of these potential problems by saying that WBEs pose the general problem of either making many of our mental experiences meaningless because they will be the result of deliberate choice, or making them disappear altogether. Either way, WBEs seem to threaten our ability to learn to meditate, and in that way to threaten our ability to learn to accept what we cannot control and thus to suffer less. If we lose or diminish that ability, we in effect condemn ourselves to a particular future—one in which our initial choices to try to avoid suffering by changing the world eliminate the possibility of making different choices down the road. We will then either suffer without alternatives or have to keep changing the world to avoid suffering.

WBE enthusiasts might reply that, with all of these changes made already, we might either no longer be capable of suffering or no longer be vulnerable to the causes of suffering. The second retort seems self-evidently false. As the Buddha teaches, everything that comes into existence eventually goes out of existence, even a well-backed-up WBE. And when that moment comes, we will suffer because we can no longer change the world to match our minds, and we will have no other means of coping. The first response is more interesting—perhaps we will no longer suffer, because we will no longer experience emotions. Bill McKibben, in his excellent book about technology titled *Enough*, wryly notes that enthusiasts for this kind of transformation of humanity have an odd habit of arguing that we should work hard and sacrifice to create a future in which there are no beings that are recognizably human. It may well be utopia, but we won't be there to enjoy it (McKibben). In the same vein, perhaps transcending suffering by

shedding the ability to feel would be wonderful, but it is also profoundly different from anything that is recognizably human.

## Conclusion

If AI and WBE, and the other NFTs more generally, threaten the problems I have suggested, what should we do about them? Logically, it seems that there only a few possibilities: (1) Embrace becoming post-human; (2) Develop the technologies but try to prevent bad outcomes; (3) Don't develop the technologies—that is, choose to renounce them in advance.

Off the cuff, (2) may seem like the most reasonable option, the path that we have usually chosen in the past. But there are good reasons for thinking that NFTs are not like other kinds of technology, because their very invention would mark a tipping point that would be difficult or impossible to control. The creation of a single AI would swiftly lead either to world domination by the AI or the power that could partially control it, or to a frenzied arms race to develop more AIs to counter the first, or possibly to all-out war. The creation of a single WBE would in essence bifurcate humanity into two competing species on very different evolutionary paths. The creation of a single nanobot would set off an arms race to create more, since the only thing that might be able to defend us from a hostile nanobot would be friendly nanobots. The creation of a single robot capable of replacing most human labor would mark the beginning of what seems likely to be the most profound transformation of human life since the beginning of agriculture. The birth of a single genetically engineered child would have a slower and more complicated impact, but whatever changes were made would be heritable, and would slowly change the

genetic make-up of humanity. The invention of any of the NFTs would set human kind off on a new path, with many more possibilities for self-destruction.

Thus, it seems as if renunciation is the best strategy—just don't pursue development of these technologies in the first place. On its face that seems hopelessly naive, even self-defeating, since (to paraphrase the NRA) if the nice people all decide not to develop these technologies, then only the bad people will end up with them (assuming that they're possible). This is where mindfulness seems to me to offer a glimmer of hope. What would it take to create a world in which everyone understood the dangers of NFTs and willingly chose not to pursue them? If NFTs are possible, and if their dangers are real, that's the same as asking: what would it take to create a world that has a future that's recognizably human? Well, for one thing, we would have to reduce the kinds of suffering that are currently encouraging people to develop NFTs. We would need to treat the reduction of want, desperation, and hopelessness as being urgent priorities. We would need to ensure that everyone has enough food, education, medical care, opportunities for remunerative and meaningful employment, and so on with other basic needs. We would need to act to reduce the sources of hatred and conflict, such as gross inequality, unremedied injustice, and, more generally, the ignoring or discounting of some people's suffering. Most broadly, we would need to create a world in which the obvious sources of human-controllable suffering are minimized. And we would simultaneously need the time and mental space to teach ourselves how to accept those things that we cannot change without creating more suffering in the long run, like our vulnerability, mortality, and finitude. We would need, in other words, a mindful society. Of course there's no guarantee that developing a mindful society will save the world, but from this perspective it seems like the only thing that might.

**Works Cited**


Bodhi, Bhikkhu, ed. *In the Buddha's Words: An Anthology of Discourses from the Pāli Canon*. Kindle ed. Somerville, Mass.: Wisdom Publications, 2005. Print.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014. Print.

Kurzweil, Ray. *The Singularity Is Near: When Humans Transcend Biology*. New York: Penguin Books, 2005. Print.

McKibben, Bill. *Enough: Staying Human in an Engineered Age*. New York: Henry Holt and Company, 2003. Print.

Moore, Matthew J. *Buddhism and Political Theory*. Oxford; New York: Oxford University Press, 2016. Print.

---. "Buddhism, Mindfulness, and Transformative Politics." *New Political Science* 38.2 (2016): 272-82. Print.

Rubin, Charles T. "Artificial Intelligence and Human Nature." *The New Atlantis*.Spring (2003): 88-100. Print.

---. *Eclipse of Man: Human Extinction and the Meaning of Progress*. New York: Encounter Books, 2014. Print.


[1] See (Moore "Buddhism, Mindfulness, and Transformative Politics")

[2] For an overview and critical reading, see (Moore *Buddhism and Political Theory*)

[3] On these themes, see (Rubin "Artificial Intelligence and Human Nature"; Rubin *Eclipse of Man: Human Extinction and the Meaning of Progress*)

[4] See (Bostrom)