Buddha in the Chinese Room: Empty Persons, Other Mindstreams, and the Strong AI Debate

JOSHUA STOLL
University of Hawai'i West O'ahu
js34@hawaii.edu

Abstract: The question of whether we can build machines that can think and feel has been with us since at least the time of Descartes. However, it has taken on a new sense of urgency and significance as our lives have become progressively more integrated with and dependent on artificially intelligent technologies. But what does 'artificial intelligence' mean exactly? Is it truly possible to build computers that can think and feel like us, and what would it mean if we could? This paper will explore John Searle's famous rejection of the possibility of such 'strong AI' through his Chinese Room Argument and how he handles several replies to his argument. We will then discuss these replies further in the context of Buddhist considerations with respect to the emptiness of persons (*pudgalanairātmya*), emptiness (*śūnyatā*) more generally, and the status of the succession of mental states in others (*santānāntara*)—especially as this pertains to Buddha's purported omniscience. Doing so will give us resources to examine the implications of Buddhist considerations for strong AI, thus giving us a sense of what a Buddhist perspective might say about the possibility of developing such technology.

Keywords: Strong AI, emptiness, other minds, Buddha's omniscience

DOI: https://dx.doi.org/10.15239/hijbs.03.02.05

I. Introduction

The question of whether machines can think, feel, and have conscious experiences has been with us since at least the time of Descartes. But they have taken on a new urgency and significance since around the middle of the twentieth century as technological advances began to produce machines that have seemed more and more capable of performing genuinely cognitive functions. Research in artificial intelligence and the development of apparently intelligent computer programs that can solve mathematical problems and play chess began in earnest around this time. Since then, advances in AI have been astounding, and our lives today are completely enveloped in such technologies. But what are the implications of such technologies? What can be said about the cognitive status of such programs and their implementations? Are they truly 'intelligent'? Are they really thinking about or understanding anything? Is it at all possible to build computers that will effectively be living, feeling, thinking persons like us, and what would this mean?

In response to such questions, John Searle famously developed an argument intended to show that 'any attempt to literally create artificial intentionality (strong AI) could not succeed just by designing programs'.[1] On the other hand, when asked about whether suitably designed computers could think or be sentient, H. H. the Dalai Lama remarked that 'It is very difficult to say that it's not a living being, that it doesn't have cognition, even from the Buddhist point of view'.[2] It appears, from these remarks, that the views of Searle and the Buddhist tradition as espoused by the Dalai Lama are at odds with respect to the possibility of building a thinking, feeling, con-

---

1 Searle, 'Minds, Brains, and Programs', 417.

2 Hayward and Varela, *Gentle Bridges*, 152.

scious machine. With this in mind, we will examine Searle's Chinese Room Argument (CRA) from a Buddhist perspective to elucidate the implications of Buddhist considerations more fully for the strong AI debate.

II. Searle Against Strong AI

Searle's most famous argument against strong AI has been dubbed the Chinese Room Argument (CRA). It is grounded in a thought experiment whereby (simplifying a little) Searle finds himself in a room, receiving printed questions in Chinese from outside the room. He has no knowledge of Chinese, or even that what he receives are questions printed in Chinese. He does, however, have a cache of similar symbols in the room with him as well as instructions in English (a 'program') for how to match certain strings of symbols to the strings of symbols he receives. Using scratch paper to make the transformations from one set of symbols to another based on the English instructions, he puts together a new string and sends it out of the room. What he sends out, unbeknownst to him, are answers in Chinese to the questions he received.

From the perspective of those outside of the room who feed it questions, the responses to their questions are indicative of someone who understands Chinese—the room effectively passes the Turing Test. Searle, however, has no knowledge of Chinese, and no knowledge that he is providing Chinese responses to Chinese questions. All he has are English instructions for matching strings of symbols which, for him, are little more than meaningless squiggles. Searle argues that since a computer program has nothing more than what Searle-in-the-room really has to go on—that is, a syntax, a set of rules by which to manipulate strings of symbols—and since Searle has no understanding of Chinese (i.e. the strings of symbols), he concludes that no computer program doing the same thing could understand the meaning of the strings of symbols it manipulates either.

Searle's stated target in the CRA is the position he calls strong AI. As he interprets it, this is the view that an appropriately programmed computer 'really *is* a mind', that 'it can literally be said to

understand and have other cognitive states'.[3] According to strong AI, 'the mind is to the brain as the program is to the hardware'.[4] The CRA therefore purports to show, as Searle later puts the point, that 'the mind could not be just a computer program, because the formal symbols of the computer program by themselves are not sufficient to guarantee the presence of the semantic content that occurs in actual minds'.[5] Since, according to Searle, programs are all syntax, and syntax is not sufficient for meaning, programs cannot be sufficient for the semantic contents of language-using minds. So suitably programmed computers cannot have a mental life in virtue of their programming alone.

While Searle here puts strong AI in terms of programs being sufficient for cognitive states that exhibit intentionality, and therefore cognitive states that have semantic content, elsewhere he articulates strong AI in terms of consciousness. For example, he states that strong AI is the view that 'the conscious mind is a program'.[6] David Chalmers suggests that this is the 'root of the matter' since, for Searle, 'intentionality requires consciousness'.[7] Thus, if Searle can show that there is no conscious understanding of Chinese—either in Searle himself or in the Searle-in-the-room-system—he would *ipso facto* show that running such a program also lacks intentionality. At best, such a program might simulate conscious understanding of language, but simulation is not duplication: 'Why on earth would anyone believe that a computer simulation of understanding understood anything?'[8]

Searle's general point, then, is that no computer, just by virtue of implementing a program, could have any genuinely psychological properties.[9] Rather, consciousness, intentionality, and understand-

---

3 Searle, 'Minds, Brains, and Programs', 417.

4 Searle, 'Minds, Brains, and Programs', 421.

5 Searle, *The Mystery of Consciousness*, 10.

6 Searle, *The Mystery of Consciousness*, 9.

7 Chalmers, *The Conscious Mind*, 322–23.

8 Searle, 'Minds, Brains, and Programs', 423.

9 Preston, 'Introduction', 20–21.

ing are, according to Searle, the result of the 'causal powers' of the brain. They are biological phenomena in the same vein as digestion, lactation, and photosynthesis.[10] Thus, Searle rejects that the purely formal, syntactic properties of a program are sufficient for meaningful consciousness. The right kind of causal power, the kind found in the brain, is required. But, as Searle puts it, 'Strong AI is not about the specific capacities of computer hardware to produce emergent properties... Strong AI claims that implementing the right program *in any hardware at all* is constitutive of mental states'.[11] To this extent, Searle claims that strong AI is not just empirically false but also lacks clear sense. Syntactically defined computations, he argues, are 'observer-relative' unlike such physical processes as digestion and, by analogy, consciousness, which are intrinsic to nature.[12] As he forcefully puts the point, 'My present state of consciousness is intrinsic... I am conscious regardless of what anybody else thinks'.[13]

In the paper in which he first puts forth his arguments, Searle also responds to a number of replies to his thought experiment. We will focus, in particular, on three replies, though their concerns arguably overlap: the systems reply, the robot reply, and the other minds reply. Here, we will briefly recapitulate what these replies amount to and Searle's dismissals of them. We will then discuss these replies further in the context of Buddhist considerations with respect to the emptiness of persons (*pudgalanairātmya*), emptiness (*śūnyatā*) more generally, and the status of the succession of mental states in others (*santānāntara*)—especially as this pertains to Buddha's purported omniscience. In this way, we will attempt to unravel the implications of Buddhist considerations for strong AI.

The systems reply suggests that while Searle-in-the-room may not understand Chinese, the whole system—including the input of Chinese 'squiggles', the ledger of instructions in English (the pro-

---

[10] Searle, 'Minds, Brains, and Programs', 424.

[11] Searle, *The Mystery of Consciousness*, 13.

[12] Searle, *The Mystery of Consciousness*, 14; Searle, *Consciousness and Language*, 17.

[13] Searle, *The Mystery of Consciousness*, 15.

gram), a memory system (scratch notes that Searle makes in building a new string of symbols from the program and the input), a database of Chinese 'squiggles' for Searle to choose from in building a new string, etc.—does have an understanding of Chinese. Searle himself is, in this view, just the CPU. Searle dismisses this reply as follows: 'The idea is that while a person doesn't understand Chinese, somehow the *conjunction* of that person and bits of paper might understand Chinese'.[14]

He goes on to argue that even if Searle-in-the-room internalized all the elements of the system—memorized all the English instructions, the complete database of symbols in the room, and did all the transformations in his head—he still wouldn't know the meaning of the input or output strings of symbols. The issue here is where and how understanding occurs. Searle contends it must occur in and because of some physical mechanism that has the 'causal powers' of the brain, which are intrinsic to nature; it cannot occur simply due to the implementation of a syntax, that is, a computer program, the specification of which is dependent on the programmer.

The robot reply attempts to take up such causal concerns. Instead of Searle-in-the-room taking formal symbols (viz. the Chinese characters that are, for Searle, meaningless squiggles) as input and giving formal symbols as output, we put an analogous program into a robot. The robot has sensors such as cameras and microphones to take input from its environment and effectors to generate motor outputs. We can add to this what Searle calls the brain simulator reply: the sensors and effectors might be connected to an artificially constructed brain that simulates the sequences of neural firings of a Chinese speaker's brain. Indeed, perhaps we could build an entity that is physiologically isomorphic to us despite not being made of the same biological material.

Such an entity, it is argued, would understand Chinese and, more generally, would have meaningful conscious experiences, so far as anyone interacting with it could tell. It would pass what Stevan Harnad calls the Total Turing Test, since it wouldn't just exhibit

---

[14] Searle, 'Minds, Brains, and Programs', 419.

an apparent understand of language but would exhibit all the signs of life we associate with other people in our everyday interactions.[15] Searle, however, objects that such a response 'tacitly concedes that cognition is not solely a matter of formal symbol manipulation'[16] and therefore gives up the commitment to strong AI. Moreover, he goes on to argue, insofar as what the robot is doing with its artificial neurophysiological system is merely simulating the electrical sequences of a Chinese speaker's nervous system, all it has is the formal structure and not the causal powers of our neurophysiology.

The point of Searle's counterarguments to each of these replies is that an *implemented* program is, in itself *qua* program, nothing but syntax, and syntax alone cannot give us semantics. Semantics, that is, conscious understanding, requires the appropriate kinds of causal connections and causal powers, but the system in Searle's thought experiment, whether it is Searle-in-the-room or a robot with an artificial nervous system, only simulates the formal sequences of symbol manipulation or nerve firings. However, a number of commentators have argued that, though formal systems of logic may require a clean distinction between syntax and semantics, actual computational systems—i.e. implemented programs—are causally efficacious and so are not purely syntactic in the way of the artificial languages used in the study of logic.

Chalmers, for example, notes that '*Implementations of programs*... are concrete systems with causal dynamics, and are not purely syntactic'.[17] Programs implemented in computational systems thus have causal properties that, as Margaret Boden puts it, give the programs 'a toehold in semantics'.[18] Thus, Searle's attack on strong AI is something of a strawman insofar as strong AI has functionalist, and not merely behaviorist, commitments. Georges Rey therefore points out that although programs are syntactically specifiable, and that this alone does not constitute a mind, implemented programs

---

[15] Harnad, 'Other bodies, other minds', 44.

[16] Searle, 'Minds, Brains, and Programs', 420.

[17] Chalmers, *The Conscious Mind*, 327.

[18] Boden, 'Escaping from the Chinese Room', 102.

still standardly possess a semantics grounded in 'their computational organization *and* their causal relations to the world'.[19] Indeed, insofar as Searle suggests that consciousness and intentionality have something to do with the causal powers of the brain, on analogy with the stomach and digestion, Rey points out that Searle might be a kind of functionalist himself.[20] We will return to these concerns with causality and functional organization in our discussion of Buddhist arguments for the emptiness or selflessness of persons.

The third reply that will concern us—the other minds reply—does return to the behaviorist considerations that Searle wants to reject. It suggests that we standardly attribute a conscious mind to others based on their behavior. Since the system in question passes the Turing Test—or better, since the physiologically isomorphic robot passes Harnad's Total Turing Test—there is no reason to not attribute understanding to the system or robot. For the robot just does whatever anyone else who understands Chinese does, right down to exhibiting the same sequence of neural firings in its robot brain. Searle brushes this reply off quickly: 'The problem in this discussion is not about how I know that other people have cognitive states, but rather what it is I am attributing to them when I attribute cognitive states to them'.[21]

However, this dismissal is perhaps too quick. For in rejecting the issue at hand in terms of the standard epistemological problem of other minds, Searle implicitly transitions to the conceptual problem of other minds. That is, he opens up the question of what it could possibly mean to say that an apparently physical object 'has' a mind. Whatever else it may mean, Searle argues that it cannot mean that the system or robot has a certain computational/functional organization alone since such an organization is, he assumes, merely syntactic. For Searle, what we attribute is a property—viz. consciousness/intentionality—that is naturally intrinsic to others, and therefore independent of outside observers. Again, we will return to this issue

---

[19] Rey, 'Searle's Misunderstanding', 219.

[20] Rey, 'Searle's Misunderstanding', 207.

[21] Searle, 'Minds, Brains, and Programs', 421.

in our discussion of Buddhist arguments surrounding the notion of other streams of conscious experience.

III. Empty Persons, Robots, and Systems

Perhaps the most straightforward place to start when bringing Buddhist philosophy into conversation with the philosophy of artificial intelligence is the former's rejection of a self-essence (*ātman*). It is often thought that explaining the continuity of one's experience requires reference to a permanent self that stabilizes one's existence across time and change. In contrast, the classical Buddhist analysis of the person offered by the Abhidharma schools is reductive, rejecting the existence of a self or an essence of the person by showing how suitably arranged and causally connected aggregates (*skandha*) condition the experience of such continuity. Because this account defines the self and, indeed, any given mental state, in terms of their causal roles, this analysis also has functionalist elements. As such, *skandha* theory is a good candidate for comparison with the sort of computationalist-functionalist position that Searle is trying to refute in the CRA.

That the self is empty, in this context, means that it lacks intrinsic reality; it is a convenient designator that allows us to make sense of functionally organized collections of fundamentally real parts. The mental states of persons are then defined in terms of the causes and conditions that aggregate together to give rise to the person's experiences, and how such experiences condition further experiences. But beyond such functional organizations, there is no experiencing person, only collections of parts causally conditioning each other in ways that make it useful to identify them with singular terms. These parts, functionally organized, condition the use of singular names, but these names are part of conventional discourse (*vyavahāra*) and indicate things that are only nominally real (*prajñaptisat*). They do not refer to any substantial reality (*dravyasat*) since they are only useful for referring to wholes that arise on the condition of a particular organization of parts that is of interest to a community of speakers. This kind of analysis, Mark Siderits suggests, is consistent with

the kind of 'technophysicalism' one finds in computationalist-functionalist analogies between minds and computers.[22] And although the Abhidharma analysis does emphasize the mental aggregates (*nāmaskandha*) of affect (*vedanā*), etc., Siderits notes that 'there is nothing in the analysis itself that precludes physical realizers'.[23] Indeed, we see such Buddhist philosophers as Saṃghabhadra and Buddhaghosa argue that sensory qualities are realized and change on the basis of changes in their corresponding physical substrates.[24]

Thus, H. H. the Dalai Lama states that a computer could be a candidate for being conscious: 'If the physical basis of the computer acquires the potential or the ability to serve as a basis for a continuum of consciousness'.[25] James Hughes, in his discussion of Buddhism and AI, elucidates this remark in a way that is reminiscent of the robot response to Searle's CRA. In order to build a conscious computer, we would need to first build a body with sensory components. That is, we would need a certain functional organization of physical aggregates (*rūpaskandha*) complete with sensory connections to the world. Without this concrete causal connection to the world, there can be no implementation of any programming. As Hughes puts it, 'To think like a human, AIs need to interact with the physical world through a body... This insight is very similar to the Buddhist observation that sense data drive the developing mind'.[26] We can see, then, that Buddhist philosophy has resources available to reject Searle's arguments along the lines of the robot reply by emphasizing that mental states are defined in terms of their causal-functional connections with their physical bases, and the causal connections between these bases and objects in the world.

At the same time, we may wonder whether and to what extent such a robot can truly develop an autonomous sense of subjectivity. After all, any ability it would have to navigate and process the data its

---

22 Siderits, 'Buddhism and Technophysicalism', 311.

23 Siderits, 'Buddhism and Technophysicalism', 311.

24 Ganeri, *The Self*, 132–34.

25 Hayward and Varela, *Gentle Bridges*, 153.

26 Hughes, 'Compassionate AI', 71.

sensors take in would be derived from a programmer's algorithm. Perhaps, through embodied motility, such a robot can explore the sources of the input data of sensors and develop recognitional capacities (*saṃjñā*) through affordance learning.[27] Still, why should such capabilities imply that this robot is a conscious, feeling entity? As Charles Goodman points out, such a being may conceivably have '*saṃjñā*, conceptions, but no *vedanā*, or feeling-tones'.[28] An empty robot, with its 'internal states' defined in terms of its functional organization would thus still be devoid of what Searle thinks is crucial: *conscious* understanding, a phenomenology accompanying the recognitional capacity. It may be able to enactively embody recognitional capacities in its explorations of objects, but this would just contribute to a behavioral appearance of understanding based on algorithms written by programmers. There would still be no feeling, no value, no meaning for the robot; it is still just processing meaningless strings of data.

However, such considerations seem to come on the basis of the assumption that consciousness is an intrinsically real phenomenon, that feeling is not itself empty. For whether the robot is *really* conscious rather than just simulating consciousness comes down to something that is independent of its designers' algorithms. On this account, it would be instructive to examine the notion of emptiness (*śūnyatā*) as it is later developed in the Madhyamaka tradition. Doing so also gives us interesting Buddhist resources to consider Searle's arguments with respect to the systems reply. Recall that the systems reply agrees with Searle that Searle-in-the-room does not understand Chinese, but that nonetheless something does, namely the whole system of which Searle is just a part. What is at issue here is where understanding occurs: in the person or in the system that includes the person. Now, the emptiness of persons, for the Madhyamaka, is not just a result of the reduction of wholes to parts. While they agree with the Ābhidharmikas about the conventional status of the self, they disagree that the component parts are any more real or, rather, any less empty than the self. Emptiness is, for them, applicable to

---

[27] See Brody, this volume, 'Enaction, Convolution and Conceptualism'.

[28] See Goodman, this volume, 'A Buddhist Contribution to Artificial Intelligence?'.

the constituents that make up persons, as well as the causal relations between them, as much as to the individual self.

What the Madhyamaka thinkers mean by saying that the self, the causal powers, and the skandhas are empty is that they lack any kind of intrinsic reality (*svabhāva*). Indeed, for the Madhyamaka, all phenomena are empty in this sense, including emptiness itself. All phenomena arise in various dependency relations, including in dependence on conceptual construction, which means claims about their natures can only ever be conventionally true (*saṃvṛtisat*), that is, true in the context of customary discourse (*vyavahāra*). Otherwise, the concepts we apply to the world would have to be considered as being applicable to a reality beyond our conceptual resources. But since in this case we can never be in a position to coherently establish such applicability conditions, as Jan Westerhoff puts it, 'we would be necessarily unable to apply [our concepts] to anything'.[29]

Since everything lacks inherent nature or intrinsic existence, the Madhyamaka perspective would charge Searle's claim that consciousness is an intrinsically real phenomenon, and that something about the brain gives it the 'causal power' to bring about consciousness, with incoherence. Indeed, Searle's notion of 'causal power' is precisely the sort of reified notion of causality Nāgārjuna criticizes in the first chapter of his *Mūlamadhyamakakārikā*. The problem is that Searle takes these terms to be referring to a reality that is, as he puts it, 'observer independent'. The terms 'syntax', 'semantics', and 'computation', Searle argues, 'do not name intrinsic features of nature', unlike the terms 'consciousness', 'tectonic plate', and 'electron'.[30] But the idea that certain terms 'name intrinsic features of nature' implies that nature is, as Westerhoff puts it in his interpretation of Nāgārjuna's Madhyamaka, 'ready-made'.[31] And according to the Madhyamaka thinkers, any supposedly objective, observer-independent notion ends up, after careful analysis, resulting in unwanted consequences (*prasaṅga*).

---

[29] Westerhoff, *Nāgārjuna's Madhyamaka*, 207.

[30] Searle, *The Mystery of Consciousness*, 16.

[31] Westerhoff, *Nāgārjuna's Madhyamaka*, 59.

With respect to the systems reply, the issue is that the system, for Searle, can only be *interpreted as possibly* understanding language (which he finds to be an absurd interpretation), but something intrinsic to Searle himself (namely something about how his brain works) has the causal capacity to bring about his *actually* understanding language—regardless of if anyone thinks this is the case. Thus, Searle supposes that consciousness and, by extension, the capacity to understand language, are intrinsically caused by his brain in a way that is independent of anyone's interests or concerns. But reducing semantics to causality in this way is problematic. It cannot be an intrinsic property of the 'causal power' of the brain without completely losing all semblance of being linguistic meaning. For such an account of semantics would, as Dan Arnold puts it, 'be intelligible only as itself dependent on *the perspective from which explanations must be offered*'.[32]

Thus, Searle's ontological claims about what really exists in nature independent of observers, namely a causal connection between consciousness and the brain, only makes sense in the context of a linguistic community. But his claim implies that whatever else being conscious of understanding language is, it is fundamentally something that is naturally intrinsic to the world independent of any community of language users. It is grounded first and foremost in the 'causal powers' of the brain. On Searle's view, as Vincent Descombes puts it, 'I cannot be said to understand French until it has been confirmed that my cranial cavity contains a brain (rather than an electronic system)'.[33] Yet no conscious understanding of linguistic meaning is possible for a brain alone.

Searle's rejection of the systems reply is arguably true—neither he himself nor his mere conjunction with pieces of paper understand Chinese. But it isn't because the appropriate causal powers are lacking. Rather, it is because neither Searle nor the system with which he is merely 'conjoined' are part of a Chinese-using community. Descombes thus highlights the fact that 'Neither the operator of the Chi-

---

[32] Arnold, *Brains, Buddhas, and Believing*, 216.

[33] Descombes, *The Mind's Provisions*, 130.

nese Room nor the entire system have any use for the ideograms'.[34] The emptiness (*śūnyatā*) of all phenomena means that whether understanding or consciousness are attributable to a computational system in the same way that they are attributable to any given person would be a matter of the conventions and interests of those who are using language with respect to these phenomena—and in this context, the extent to which the system or its operator have developed as a member of the language-using community. So, for Searle or the system to really be thought of as being able to understand Chinese, they would have to have learned to live amongst a Chinese-using community. Claiming that such understanding is intrinsic to some causal mechanism of the individual would result in the unwanted consequence that conscious understanding of language does not depend on a linguistic community.

IV. Other Mindstreams and the Other Minds Reply

Turning now to the other minds reply, we noted that Searle gave a quick dismissal of this take on the CRA. However, it was noted that his dismissal was perhaps too quick. Indeed, as Harnad points out, the problem of other minds and the problem of understanding what it would mean for artificial devices to have minds are closely connected.[35] Despite Searle's protest to the other minds reply, this is precisely what is at issue. For his argument against strong AI relies entirely on his first-person awareness of lacking an understanding of Chinese. Diane Proudfoot, summarizing Searle's point, thus says, 'Mostly Searle's claim is simply that *he* is the person manipulating the symbols and *he knows* he cannot read Chinese'.[36]

Buddhist philosophy, more than most other Indian philosophical schools, has been explicitly concerned with the problem of other minds, though their reasons for considering this issue are soterio-

---

[34] Descombes *The Mind's Provisions*, 132.

[35] Harnad, 'Other Bodies, Other Minds', 45.

[36] Proudfoot, 'Wittgenstein's Anticipation', 171.

logical and thus quite distinct from contemporary considerations. Indeed, the Buddha's teaching, particularly as interpreted by the Mahāyāna schools of thought, is especially concerned with liberating *others* from suffering—such is the bodhisattva ideal. Thus, *buddhadharma* is to be taught and preached to *others*. Moreover, one of Buddha's supernatural powers (*abhijñā*), afforded him by liberation, is said to be direct knowledge of other minds (*paracittajñāna*). Thus, for the Buddhists a lot rides on making sense of other streams of experience (*santānāntara*).

The existence and knowledge of other minds has been most thoroughly examined by the Yogācāra Buddhists, who are often read as taking consciousness to be the fundamental nature of reality. The defining feature of consciousness, as is articulated in later developments of the Yogācāra tradition, is its being intrinsically reflexive or self-aware (*svasaṃvitti*). That each moment of consciousness is, in itself, reflexively self-aware means that the content and meaning of any conscious episode is intrinsic to that conscious episode. This is the point of Dharmakīrti's well known *sahopalambhaniyama* argument: objects are characteristically constrained by their being co-cognized with the apprehending awareness. As Arnold puts it, the idea is that 'it's only in virtue of intrinsic properties of awareness that perceptual objects can *seem* in the first place to be distinguished by their independence from awareness'.[37] The very idea of an object external to consciousness is therefore constrained by its not being able to appear independently of being apprehended by a cognitive event. Thus, as Thomas Wood puts it, our cognitions 'cognize only themselves'.[38]

But if all that ultimately exists for the Yogācārin is a stream or series of reflexively self-aware mental events causally conditioned by previous mental events, they are now hard pressed to make sense of the emphasis on the suffering of a plurality of sentient beings that is presupposed by the Buddhist tradition. For it would seem to commit them to the view that there are no others, since the content of every cognition is intrinsic to the cognizing event itself. To counter this

---

[37] Arnold, *Brains, Buddhas, and Believing*, 175.

[38] Wood, *Mind Only*, 93.

issue, Dharmakīrti, in his *Santānāntarasiddhi*, seeks to show that the Yogācārin idealist is no worse off on this account than realists. In effect, he argues that both the realist and the idealist use the same strategy to establish knowledge of other minds, namely inference. But the idealist's inference is lighter since it does not require taking steps to infer, first, the existence of physical bodies outside of the cognizing event, and secondly, the presence of mentality in such bodies. All it requires is the generalization of what is experienced in one's own case: that actions are caused by intentions to act.[39]

The later Buddhist thinker Ratnakīrti, in his *Santānāntaradūṣaṇa*, seeks to show that, from an ultimate perspective, Dharmakīrti's conclusion is incoherent. For if a difference is to be found between one's own mindstream and a mindstream belonging to someone else, some boundary must be manifest whereby we can compare similarities and dissimilarities between the two. But insofar as any given mental event is reflexively self-aware, no such boundary can ever be manifest. Indeed, given the Buddhist no-self (*anātman*) position, there is no distinct owner of any stream of mental events. Add to this the notion of the intrinsic reflexivity of all mental events, and there is ultimately no way of making a non-arbitrary distinction between one series of mental events and another, any more than we can non-arbitrarily divvy up waves in a roiling ocean. The very idea of one mind as opposed to another is, Ratnakīrti argues, unintelligible.

Our discussion of the problem of other minds in Yogācāra gives us resources to consider how some Buddhists might think of the other minds reply and Searle's response to it. On the one hand, Dharmakīrti's inference suggests that, at least at a conventional level, Buddhists would endorse the other minds reply. For if an artificially constructed being could behave in every way as any other apparently intelligent being does, then the inference to other minds would lead to successful practice, and this is all that conventional truth is for Dharmakīrti. Ratnakīrti's critique of Dharmakīrti's inference, on the other hand, brings up precisely the conceptual problem that Searle's critique of the other minds reply does. But it entails that the very notion of

---

[39] Wood, *Mind Only*, 108–9.

breaking up streaming mental events into this stream or that one is incoherent. 'For the Buddhist idealist,' Roy Perrett thus concludes, 'ultimately there is no way to be able to draw a distinction between one consciousness and another'.[40] Thus, according to the Yogācāra understanding of what constitutes ultimate reality, the very thing that Searle thinks is really in question in the CRA is incoherent because nothing intrinsic to the world allows for distinctions between one stream of consciousness and another.

Something like this conclusion seems to have already been suggested by both Vasubandhu (in the *Viṃśatikā*) and Dharmakīrti (in the *Santānāntarasiddhi*) in their respective discussions about Buddha's omniscience with regard to other minds. For they each tell us that Buddha's cognition is non-dual, without any distinction between cognizing subject and cognized object (*agrāhyagrāhaka*), and is thus ineffable and unthinkable.[41] As such, Buddha sees through the illusion of conventional discourse and directly perceives ultimate reality as simply the streaming succession of mental events (*santāna*). The question of what it means for there to be others is a question that only arises in the context of the illusions that are grounded in the appropriation (*upādāna*) of streaming mental events as essentially one's own self (*ātman*). Buddha's omniscience thus does not involve a distinction of other mindstreams from his own, and so the question of attributing mentality to sentient creatures is not one that can arise for Buddha. What does this mean, then, for Buddha's knowledge of other minds (*paracittajñāna*)? And how might it further help us think through what a Buddhist perspective on strong AI could look like?

We can find some help articulating this situation from Śāntarakṣita's *Tattvasaṃgraha* and Kamalaśīla's commentary on this work. In defending the Yogācārin point of view against a Buddhist realist who argues that their acceptance of both reflexivism and Buddha's knowledge of other minds is inconsistent, Śāntarakṣita says that Buddha 'has no cognitions' (*adarśana*)—that is, his cognition is free from

---

[40] Perrett, 'Buddhist Idealism', 67.

[41] Wood, *Mind Only*, 131.

dualistic conceptualizations—but is regarded as omniscient because 'He brings about the welfare of men'.[42] Kamalaśīla clarifies: 'By the force of his previous meditations, the Lord has no limitations; He is like the Kalpa-tree, bringing about the welfare of the entire universe'.[43]

Buddha's putative cognition of other minds, despite not distinguishing between his own and other minds, is explained here on the model of what Sarah McClintock calls 'spontaneous omniscience'.[44] By the force of the vows and meditative efforts of innumerable lifetimes, Buddha has developed wisdom and compassion to such a degree as to spontaneously—i.e. effortlessly and without deliberation—say and do what needs to be said and done to bring sentient beings closer to liberation. In short, he has cultivated unparalleled ethical skills akin to what Peter Hershock calls 'ethical improvisation', 'adaptive conduct that expands ethical horizons and progressively raises standards of ethical virtuosity'.[45] Buddha's putative cognition of other minds, then, is attributed to him by deluded non-Buddhas to whom he appears to spontaneously and effortlessly act in ways that exemplify ethical virtuosity, almost as if to be able to see directly what others need him to do. In this way, without having other minds as objects of cognitions, Buddha is said by ordinary folks to know other minds directly.

This understanding of Buddha's omniscience with respect to other minds suggests that Buddha is not that different from Searle-in-the-room in at least one respect: both are, in a sense, without cognition (*adarśana*). However, Searle-in-the-room lacks the relevant conscious awareness, namely, an understanding of Chinese, because—as we noted in our discussion of Madhyamaka thought and the systems reply—Searle has never shared in or grown through the conventions of Chinese language use. That is, he has never participated in the circumstances under which he could develop the skill

---

[42] Jha, *The Tattvasaṅgraha of Shāntarakṣita*, 973.

[43] Jha, *The Tattvasaṅgraha of Shāntarakṣita*, 973.

[44] McClintock, *Omniscience and the Rhetoric of Reason*, 353.

[45] See Hershock, this volume, 'The Intelligence Revolution and the New Great Game'.

of Chinese language use to such a degree that he can effortlessly and spontaneously immerse himself in a Chinese language-using world. But Searle himself is presumably, like the rest of us ordinary folks, wrapped up in the continuous proliferation of conceptions that manifests as dualistic, discriminating cognition. Thus, he experiences suffering in the standard Buddhist sense of the dissatisfaction that is tied to simply being alive. There is presumably something it is like to be John Searle.

But, as Paul Griffiths contends, there is arguably nothing it is like to be Buddha.[46] This is because Buddha is said to have cultivated wisdom and compassion over many lifetimes to such a degree that he spontaneously does what needs to be done for the benefit of all sentient beings without thinking, deliberating, or even experiencing anything—at least insofar as we can make sense of the notion of 'experience' in conventional discourse. He has cultivated his ethical skill to such a degree that, like an improvising jazz musician, he adaptively acts without reproducing conceptual distinctions that section himself off from his actions and the world around him—he realizes the world, not as an object of cognition, but as an extension of his embodiment (*dharmakāya*). As such, it appears to us that he knows the minds of others, for he acts in just such a way as to aid us on the path. And we, out of the limitations of our conceptual capacities and conventional discourses, attribute such abilities to him.

It would seem, then, that a machine that passes the Turing Test, or even the Total Turing Test, would not be much different in respect of content, so far as Searle is concerned, from Buddha. In an important sense, there is nothing it is like to be either of them because neither cognizes any distinct semantic or mental content. But there is a crucial difference: Buddha's 'lack of cognition' is a highly cultivated skill, developed through the painstaking efforts of innumerable lifetimes. Buddha's skill allows him to break free of habitual modes of acting, thinking, and feeling, so as to spontaneously and effortlessly do whatever is needed to help us ordinary folk along the path to awakening. Machines, on the other hand,

---

[46] Griffiths, *On Being Buddha*, 193.

though they can develop themselves through learning algorithms to perform extraordinary feats beyond human capabilities within the domains for which they are designed, can only develop within the circumscribed domain delimited by human agents. As Douglas Duckworth puts it, 'machines, as cultural products, reflect the psyche and goals of their creators—our machines are an extension of ourselves and an expression of human values'.[47]

Thus, whereas Buddha, once mired in the proliferation of habitual tendencies, was able to extricate himself from these tendencies by intensive effort, the development of the machine is still at the mercy of the values, habits, and conceptual frameworks of those who develop them. One of the advantages human cognition appears to have over artificial intelligence is that the former can exhibit a creative flexibility that the latter cannot. Still, we are easily attached to our ways of being in the world so as to propagate habitual tendencies that limit our flexibility and adaptability. It is such attachment that grounds the conceptual proliferation (*prapañca*) of our discriminating cognitions, in turn trapping us in cycles of suffering. Buddha's omniscience is the product of having overcome this constant reproduction of such cognitions, and so is 'without cognition' (*adarśana*). And yet he displays a skillful intelligence beyond the capacity of any ordinary human bound up in the proliferation of their habitual ways of thinking, feeling, and acting. In this way, perhaps ironically, we ordinary humans with our habitual tendencies are, from the perspectiveless perspective of enlightenment, relatively mindless machines running a set of programs.

V. Conclusion

We have seen that Buddhist philosophy gives us resources to consider Searle's critique of strong AI along a number of dimensions. Perhaps the most important point we can glean from the Buddhist perspec-

---

[47] See Duckworth, this volume, 'Machine Learning, Plant Learning, and the Destabilization of Buddhist Psychology'.

tive comes from the emphasis on cultivating a skillful comportment, particularly as we found in our discussions of the Madhyamaka conception of emptiness (*śūnyatā*) and Buddha's direct knowledge of other minds (*paracittajñāna*). Insofar as a computational system can open-endedly develop a highly skillful mode of activity with respect to some domain, it is, on this perspective, sensible to attribute 'intelligence' to it insofar as this is how we use the term in customary discourse.

However, Searle's concern is precisely that such a way of thinking about intelligence gives in to behaviorism and denies the intrinsic reality of consciousness. Even if, as we saw in our discussion of the Abhidharma, we can develop a mobile robot with sensors than can learn about its environment through its actions, such an entity may still at best develop recognitional capacities without the 'what it is likeness' that gives intentional content a meaningful feel. Skillful comportment, for Searle, is not enough for 'real' intelligence; at best it simulates the structure of intelligence.

There is certainly something to this point. Such machines derive their apparently intelligent behavior from their programmers; whatever skills they exhibit, even if they go beyond human capacities within a given domain, are still delimited by the aims of their developers. Because of this, AI will always lack a certain flexibility and adaptability across domains. But still, this lack of flexibility is derived from our own. And the sense that something fundamentally, *objectively* distinguishes us from machines is itself due to the propagation of a methodological approach to mind that, to borrow phrasing from Christian Wittern, 'has obscured a whole set of other available methods'.[48] The framework in which such a distinction is made can itself be considered a kind of programming, the continuous repetition of a set of cultural products that obscures other possible modes of intelligence, other methods for understanding our world.

Such modes of intelligence include the kind of skillful, non-dual cognition that allows us to seamlessly get along in our world. We exhibit this kind of skillful comportment in the basic, mundane

---

[48] See Wittern, this volume, 'Zen, Motorcycles and Burning Buddhas'.

ways that we directly connect to our world without intellectualized dichotomization—for example in the spontaneous and effortless use of language. Indeed, beyond linguistic communication, we directly and affectively recognize feeling-tones (*vedanā*) in others without sectioning off their bodily comportment and consciousness from 'one's own' conscious episode in which this recognition occurs—it is originally all one event, like Martin Buber's I-Thou relation. The conceptual framework in which oneself is distinguished from another—one subject cognizing another as an object—is an abstraction from the seamless affective immediacy of our living situation. As we have seen from Ratnakīrti's argument, there is no non-arbitrary point from which we can distinguish streams of conscious episodes. From such a Buddhist perspective, whether a computational system can be said to be 'really' conscious is thus a question that only arises at the level of a dichotomizing conceptual framework.

Bibliography

Arnold, Dan. *Brains, Buddhas, and Believing: The Problem of Intentionality in Classical Buddhist and Cognitive Scientific Philosophy of Mind*. New York: Columbia University Press, 2012.

Boden, Margaret. 'Escaping from the Chinese Room'. In *The Philosophy of Artificial Intelligence*, edited by Margaret Boden, 89–104. New York: Oxford University Press, 1990.

Brody, Justin. 'Enaction, Convolution and Conceptualism: An AI-Based Exploration of Dharmakīrti's Perception and Conception'. *Hualin International Journal of Buddhist Studies* 3, no. 2 (2020): 1–26.

Chalmers, David. *The Conscious Mind*. New York: Oxford University Press, 1996.

Descombes, Vincent. *The Mind's Provisions: A Critique of Cognitivism*. Translated by Stephen Adam Schwartz. Princeton: Princeton University Press, 2001.

Duckworth, Douglas. 'A Buddhist Contribution to Artificial Intelligence?' *Hualin International Journal of Buddhist Studies* 3,

no. 2 (2020): 27–37.
Ganeri, Jonardon. *The Self: Naturalism, Consciousness, and the First-Person Stance*. New York: Oxford University Press, 2012.
Goodman, Charles. 'Machine Learning, Plant Learning, and the Destabilization of Buddhist Psychology'. *Hualin International Journal of Buddhist Studies* 3, no. 2 (2020): 38–61.
Griffiths, Paul. *On Being Buddha: The Classical Doctrine of Buddhahood*. Albany: SUNY Press, 1994.
Harnad, Stevan. 'Other bodies, other minds: A machine incarnation of an old philosophical problem'. *Minds and Machines* 1 (1991): 43–54.
Hayward, Jeremy, and Francisco Varela. *Gentle Bridges: Conversations with the Dalai Lama on the Sciences of Mind*. Boston: Shambala Publications, 1992.
Hershock, Peter D. 'The Intelligence Revolution and the New Great Game: A Buddhist Reflection on the Personal and Societal Predicaments of Big Data and Artificial Intelligence'. *Hualin International Journal of Buddhist Studies* 3, no. 2 (2020): 62–77.
Hughes, James. 'Compassionate AI and Selfless Robots: A Buddhist Approach'. In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Patrick Lin, Keith Abney, and George A. Bekey, 69–84. Cambridge: MIT Press, 2012.
Jha, Ganganatha. *The Tattvasaṅgraha of Shāntarakṣita with the Commentary of Kamalashīla, Vol. II*. Delhi: Motilal Banarsidass, 1986.
McClintock, Sarah. *Omniscience and the Rhetoric of Reason: Śāntarakṣita and Kamalaśīla on Rationality, Argumentation, and Religious Authority*. Somerville: Wisdom Publications, 2010.
Perrett, Roy. 'Buddhist Idealism and the Problem of Other Minds'. *Asian Philosophy* 27 (2017): 59–68.
Preston, John. 'Introduction'. In *Views into the Chinese Room*, edited by John Preston and Mark Bishop, 1–50. New York: Oxford University Press, 2002.
Preston, John, and Mark Bishop. *Views into the Chinese Room*. New York: Oxford University Press, 2002.
Proudfoot, Diane. 'Wittgenstein's Anticipation of the Chinese Room'. In *Views into the Chinese Room*, edited by John Preston

and Mark Bishop, 167–80. New York: Oxford University Press, 2002.

Rey, Georges. 'Searle's Misunderstanding of Functionalism and Strong AI'. In *Views into the Chinese Room*, edited by John Preston and Mark Bishop, 201–25. New York: Oxford University Press, 2002.

Searle, John. *Consciousness and Language*. New York: Cambridge University Press, 2002.

———. 'Minds, Brains, and Programs'. *Behavioral and Brain Sciences* 3, no. 3 (1980): 417–57.

———. *The Mystery of Consciousness*. New York: New York Review of Books, 1997.

Siderits, Mark. 'Buddhism and Technophysicalism: Is the Eightfold Path a Program?' *Philosophy East and West* 51, no 3 (2001): 307–14.

Westerhoff, Jan. *Nāgārjuna's Madhyamaka: A Philosophical Introduction*. New York: Oxford University Press, 2009.

Wittern, Christian. 'Zen, Motorcycles and Burning Buddhas'. *Hualin International Journal of Buddhist Studies* 3, no. 2 (2020): 102–28.

Wood, Thomas. *Mind Only: A Philosophical and Doctrinal Analysis of the Vijñānavāda*. Honolulu: University of Hawai'i Press, 1991.