

A Critical Analysis of Ray Kurzweil's *The Age of Spiritual Machines*¹

Roger White²

Loyola University New Orleans
Louisiana, USA

Not having the qualifications to discuss and comment on the technical issues of artificial intelligence raised by Ray Kurzweil (1999) in his latest book, it seems best if I limit myself to several of the theoretical problems raised by the author. The most provocative of these issues would seem to be his prediction that we will relatively soon be able to transfer human consciousness to computers. I will draw heavily upon the work of John Searle (1994, 1995, 1999) because he raises some rather interesting challenges to this prediction. The assumptions behind the arguments for the possibility of the transfer of the human mind to a computer are that 1.) The brain is a kind computer, 2.) consciousness is essentially a kind of computation. Given assumptions 1.) and 2.), a transfer of consciousness from one kind of computer to another should be feasible, given the proper technology. Searle, in challenging these assumptions, also challenges Kurzweil's prediction.

Ray Kurzweil argues that, towards the end of the next century, human beings will be able indeed to transfer their personalities from carbon based operating systems to more durable systems and that mortality will, thereby, cease to be a practical problem (1999, p. 280). This process will occur as a result of the development of non-invasive scanning technology (Kurzweil, 1999, p. 64), which will be able to copy the pattern of the carbon neural activity of our brain, and transfer that pattern to neural network systems operating on a much better substrate. Because these new systems will work much faster and will be more durable, we will be able to think much better and live much longer, almost forever. Thus, argues Kurzweil.

Searle, sitting on the opposite side of the debate, argues that it is not computation, which causes consciousness, but, rather, consciousness, which causes computation (1994, pp. 218–219). Computation, he implies, requires a set of symbols and symbols require a conscious brain. Without a conscious activity of this kind already in operation, on Searle's account, there are no symbols and, hence, no computation, just physical activities which may or may not lend themselves to the symbolic representations of a conscious observer (1995, pp. 64-65).

Searle invites you to look inside of a computer. Look as hard as you may, you will not find any zeros or ones (1994, p. 210). You will find, instead, physical systems designed to conduct electrical current in such a way that the positive and negative charges can be made to correspond with a binary code of zeros and ones. The zeros and ones are not in the machine. They are in our minds. Likewise, if you open up a human brain and look inside, you're not going to find any numbers. Rather, you will find a biological organism. (Searle, 1994, p. 90). The brain is not a computer, Searle concludes (1994, p. 247). Rather, computing is one among many things that the

conscious portion of the brain does (Searle, 1999, p. 38). This does not mean that there could not be other conscious entities besides the human brain, but merely that such entities would not be essentially computers.

The bases of Searle's arguments are fairly simple. Similar causes, he argues, yield similar effects. Disparate causes yield disparate effects (Searle, 1994, p. 218). Computers and conscious brains are disparate causes; therefore, computers and conscious brains yield disparate effects. The confusion, given Searle's account, seems to stem from a failure to distinguish between the metaphors we use to describe a physical reality and the reality itself, much like lapsing into the belief that a Weeping Willow tree is shedding actual tears. When we talk about neurotransmitters in the brain "processing information", we are using a metaphor. This is okay, given Searle's account, as long as we remember that we're using a metaphor, and not lapse into the belief that these physical processes are literally engaged in computation (Searle, 1999, p. 38).

To make my point, as it relates to the problem of natural computation, I draw upon the thinking of Ludwig Wittgenstein, who happens to be a favored philosopher for both Kurzweil (1999, pp. 59-60) and Searle (1999, p. 90). Computation proceeds according to certain rules, such as the algorithm for computing factorials, namely, $n \times (n - 1)$ where the factorial of $1 = 1$. (This also happens to be the algorithm Kurzweil uses in his book to explain how to build an intelligent machine. [p. 283]) Now Searle argues that rules are conscious devices which we use to guide conduct (1999, p. 216). I would say further that unconscious nature needs no rules to guide its conduct because unconscious nature never has to worry about straying from its correct path. It has no correct path. It does whatever it does. We are, indeed, able to represent certain relationships in nature according to mathematical formulas and get certain right and wrong answers, but the answers are right or wrong according to the rules we have established. Unconscious nature, however, never has to worry about being right or being wrong.

Since this is all rather abstract stuff, let me end with an example inspired by Wittgenstein that is concrete if, perhaps, unusual. In the college where I work, we have a rule for faculty meetings, and that rule is, "Be here on time." Suppose at the next faculty meeting, I arrive at precisely the time scheduled, but clad in a bathrobe and shower thongs, as well as being extremely drunk. Have I complied with the rule?

I have complied with its letter, I would argue, but not with its spirit. The rule, "be here on time," could be interpreted in accordance with a number of other rules, as can also be the case with the rules of mathematics, according to Wittgenstein (1958, p. 13). Some of the rules in the case I have presented are "be available for work; be able to make constructive comments; don't make 'dumb' comments; don't eat large and noisy food objects," and the list goes on and on. The number of rules is staggering and, perhaps, innumerable. In a similar fashion, rules, including the rules of computation, make sense in terms of other rules. They make sense in terms of a system of beliefs about proper and improper conduct in the field of computing. Those beliefs shade over into other non-computational areas as well. They arise out of, and are rooted in, the natural, social, and personal reality of our lives.

These realities, according to Searle, constitute an extensive and complex background to our consciousness (1994, p. 194). His arguments imply that to attempt to transfer consciousness to a computer is in effect to uproot it from the reality within which it is embedded, to tear it away from its background. Such an attempt will not, if Searle is correct, make that consciousness virtually immortal but will, in fact, destroy it. I, myself, find his points of sufficient interest to give them at least some thought before simply accepting Kurweil's prediction.

References

- Kurzweil, Ray (1999). *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, New York: Viking.
- Searle, J. R. (1994). *The Rediscovery of the Mind*. Cambridge, Massachusetts: The MIT Press.
- Searle, J. R. (1995). *The Construction of Social Reality*. New York: The Free Press.
- Searle, J. R. (1999, April 8). I Married a Computer. *The New York Review of Books*, 34–38.
- Wittgenstein, L. (1958). *The Blue and Brown Books*. New York: Harper Colophon Books, p. 13.

¹ Kurzweil, Ray (1999). *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, New York: Viking. (ISBN 0-670-88217-8)

² Dr. Roger White is an Assistant Professor of Political Science at Loyola University New Orleans. He can be reached at Campus Box 14, 6363 St. Charles Ave., New Orleans, LA 70118, USA. Email: rwhite@loyno.edu, Phone: (504) 865-2697, Fax: (504) 865-3883.

