

THE PRIMACY OF THE FIRST PERSON: REPLY TO RAY KURZWEIL

By William Dembski

I'd like to take this opportunity to thank Micah Sparacio for organizing the discussion of the recently published book *Are We Spiritual Machines?* as well as Ray Kurzweil for his response to my essay in that book and his willingness to take part in this discussion. My essay in that book was titled "Kurzweil's Impoverished Spirituality" and was essentially a stripped down version of a piece I had done for *First Things* in which I critiqued both Kurzweil and Nancey Murphy with regard to their materialistic account of human mentality and spirituality. The audience of that periodical would largely have been sympathetic to my case, being generally opposed to materialism. I see now, given Kurzweil's response, that it would have been better to provide a more thorough analysis of his project and a more detailed technical response to his claims. I doubt that it could have been done there with the thoroughness required in the space allotted, and I certainly won't attempt it here. Eventually I will get to it. I want here to outline where our projects diverge and why I find his project unconvincing.

First off, I want to clarify a point I made that Kurzweil seems to have interpreted quite differently from what I intended. I wrote that predictability is materialism's main virtue. He seems to have interpreted that claim as meaning that the materialist's world is particularly simple, where everything follows well-defined linear rules and comes out in the end in a neat and tidy package. In retrospect I can see how this interpretation was possible, but writing as a philosopher I had something completely different in mind. Philosophically speaking, materialism is a metaphysical position about what exists in the world, namely, matter or mass-energy or strings or quantum-gravitational fields (i.e., material and only material entities). In particular, materialism allows no extra-material factors. This restriction leads to an inherent predictability about what sorts of questions are legitimate to pose and what sorts of answers are acceptable. In the case of human mentality, it means that the mind must be entirely captured by the brain (indeed, what else could there be to the mind on materialist terms?).

Now the predictability of materialism as a philosophical project does not mean that material systems need be predictable. I'm all too aware of the literature on supervenience, holism, hierarchy, and above all emergence. Complex systems can give rise to unexpected emergent properties that cannot be reduced in some neat modular way (Michael Denton's account of protein folding in *Are We Spiritual Machines?* provides a case in point). Kurzweil attributes to me a modular understanding of complex systems, and complex machines in particular, but nothing I wrote in my essay requires such modularity. The issue remains whether mind and spirituality can adequately be characterized as an emergent property of a purely material system, and computation in particular. Kurzweil says yes, I say no.

Kurzweil's justification for saying yes seems to run roughly as follows: (1) There's overwhelming evidence that a purely material evolution has occurred and given rise to human beings and their consciousness. (2) The key to that material evolution is competition, essentially the Darwinian mechanism of random genetic change and natural selection. (3) That mechanism has a powerful analogue in computation, where it is known under the rubric of evolutionary computation (which includes genetic algorithms). (4) The complexity of the brain can be quantified and is shortly to be exceeded by the complexity of computation as available on humanly-built computers. (5) As a consequence, such computers, once suitably programmed, will exceed in cognitive capacities those of human beings. (6) Suitably programming such computers consists in endowing them with the right patterns. (7) Holistic approaches to programming that employ evolutionary computation and neural nets, for instance, will endow such computers with the right patterns, and in particular with consciousness as well as capacities that far outstrip those of humans.

Is Kurzweil a visionary outlining bold new possibilities that are just around the corner or is he an overly enthusiastic devotee of computation whose materialism has blinded him to the inherent limits of computation? By running through the previous seven points, let me indicate why I opt for the latter.

Ad (1). Despite the rhetoric about "overwhelming evidence" for a purely material evolution, the evidence is far from conclusive. The origin of life remains a great mystery on materialist terms as do certain key transitions in the history of life (e.g., origin of metazoans).

Ad (2). The Darwinian mechanism, despite its widespread acclaim, is increasingly recognized as not being able to resolve the problems with whose solution it has in the past been credited. In a 1996 review of Michael Behe's book *Darwin's Black Box*, James Shapiro, a molecular biologist at the University of Chicago, wrote: "There are no detailed Darwinian accounts for the evolution of any fundamental biochemical or cellular system, only a variety of wishful speculations. It is remarkable that Darwinism is accepted as a satisfactory explanation for such a vast subject -- evolution -- with so little rigorous examination of how well its basic theses work in illuminating specific instances of biological adaptation or diversity" (*National Review*, 16 September 1996). Five years later cell biologist Franklin Harold wrote a book for Oxford University Press titled *The Way of the Cell*. In virtually identical language, he notes: "There are presently no detailed Darwinian accounts of the evolution of any biochemical or cellular system, only a variety of wishful speculations."

Ad (3). Why should we think that evolutionary computation is so powerful if its analogue in nature, the Darwinian mechanism, appears increasingly to suffer severe limitations. In fact, evolutionary computation only succeeds when intelligence carefully adapts the algorithm in question to the problem at hand and, in particular, carefully assigns a cost/error/fitness function to assess how well candidate solutions are doing. As Geoffrey Miller writes: "The fitness function must embody not only the engineer's conscious goals, but also her common sense. This common sense is largely intuitive and unconscious, so is hard to formalize into an explicit fitness function. Since genetic algorithm solutions are only as good as the fitness functions used to evolve them, careful development of appropriate fitness functions embodying all relevant design constraints, trade-offs and criteria is a key step in evolutionary engineering."

These last three points are of course too fast and easy. Materialist evolutionists will dispute them point for point, invoking co-evolution, the evolution of evolvability, and co-optation among other things. I've written several books now where I argue that these points are indeed valid and constitute strong countervailing evidence against materialist accounts of evolution. Indeed, the whole point of the intelligent design movement is to settle these three points against materialism and for biological design by an intelligence not reducible to material mechanisms.

It's the next four points that are more specific to Kurzweil's project. Since they depend on the previous three, Kurzweil's project is thrown into question to the degree that intelligent design proves successful. But I want here to argue briefly why Kurzweil's project is suspect even apart from the success of intelligent design in refuting evolution by purely material mechanisms.

Ad (4). Quantifying the complexity of the brain in terms of number of neurons or synaptic connections seems crude. Is this the right level of resolution? Suppose that resolution at the level of synaptic connections is adequate. How much processing is going on at these connections? How many neurotransmitters and inhibitors are being used. How many combinations of chemicals are possible? How many distinct states at a synaptic connection need to be represented computationally before no crucial information is lost? Is it significant that brain activity does not run according to a fixed clock time (as with present computers)? Does parallel processing in the brain where the processing does not run according to a fixed clock but things can be out of sync increase the level of complexity and if so by how much? All of these considerations cause me to doubt that a few more years of Moore's Law in operation are going to lead to a computer whose complexity captures or exceeds that of the human brain.

Ad (5), (6), and (7). I want to treat the final three points as a piece. Let's grant for the sake of argument that the complexity the brain can be matched computationally in the sense of hardware. Why should that lead us to think that we are computational systems or that we can program systems that will have the same

cognitive capacities as us? Think of it this way. Let's imagine we have a super-duper piece of computer hardware made of silicon and wires and the usual stuff that goes into computers. Let's assume that for any distinguishable brain state we could map it uniquely (and hopefully in some functionally appropriate way) onto a corresponding computational state. Why should we think that such a mapping underwrites a computational view of mind? I can, for instance, map spoken words uniquely onto written words. All that's needed here is some sort of appropriate lexicon that goes from voice prints to text. But having that correspondence does nothing to explain meaningful communication, whether spoken or written.

In the case of computational theories of mind what we have are minds with their full capacities, however they arise and whatever they are, and computers with their well-defined computational capacities. Assuming a one-to-one mapping of brain to computational states, why should we think that those capacities match up. Granted, the brain does things and computers do things. But why should we think they can do the same sorts of things? Consider Mozart and Muzio Clementi, rough musical contemporaries. Just because Mozart could in principle pen any composition that Clementi might and vice versa does not mean that they have the same capacities. Indeed, Mozart's capacities at musical composition far outstripped those of Clementi.

But according to Kurzweil, it's all in the "pattern." If the functionally equivalent pattern carried by a human brain could be represented in a computer, then the computer would be equivalent to the human being. And since all patterns need is a representational matrix of suitable complexity, once the complexity of computers matches that of brains, the only obstacle to getting computation to think and be conscious is finding the right pattern qua program to run on the computer.

This seems to be the heart of Kurzweil's argument, and it is specious. The fault can be found by looking a bit closer at these patterns and the material that embodies them. In his response to me, Kurzweil writes, "We are unable to really 'touch' matter and energy directly, but we do directly experience the patterns underlying 'things'." What I want to focus on in this sentence is the "we." It is WE who perceive patterns embodied in material things. But who here is the "we"? In Kurzweil's review of Wolfram's new book (*A New Kind of Science*), Kurzweil refers approvingly to Marvin Minsky's theory of intelligence as a "society of mind" in which "intelligence may result from a hierarchy of simpler intelligences with simple agents not unlike cellular automata at the base." These intelligences, at whatever level in the hierarchy, are thus themselves patterns (that, after all, is what all the pictures of cellular automata in Wolfram's book exemplify -- patterns). Thus, within Kurzweil's scheme, it's patterns all the way down -- patterns interpreting other patterns. The first person perspective is thus irretrievably lost or submerged in a sea of nonpersonal patterns.

Now perhaps such a reduction of the first person is possible and in fact the way things are. But if so, how can we know it? What we know for now is that first-person agents exhibiting a unity of consciousness display capacities that far outstrip anything computers are currently able to accomplish. What's more, within that first person perspective, patterns do not become self-interpreting, self-organizing, or self-validating. Rather, it is the first person that makes sense of patterns. Patterns are embodied in material and then interpreted by, from all we can tell for now, an irreducible first person.

From this vantage, we might think of the brain as a "dynamic text" whereas a book can be thought of as a "static text." The patterns in the book need to be intact if the book is to be readable. Certain damage to the book will not effect the readability and thus interpretability of the text. But others will. Water damage, for instance, usually mars a book's appearance substantially but doesn't affect its readability. A torn page from a crucial passage, on the other hand, destroys the sense of the book. We can therefore think of the mind as that which makes sense of and interprets the brain and human body. We can thus think of the brain as "thinking" in much the same way as a book "informing." A book informs when a reader with suitable background knowledge reads it. Likewise the brain thinks when an intelligence suitably connected to that brain interacts with it.

From a materialist perspective this will seem completely crazy. But for the nonmaterialist materialism is itself completely crazy. Thus, when Kurzweil at the end of his response to me remarks, "If Dembski's intelligence-enhancing extra-material stuff really exists, then I'd like to know where I can get some," I

would respond that he already has it. In fact, the "I" that would like to "get some" is the very extra-material stuff that he is asking for. I would make the same point about Wolfram. In his preface he remarks, "The creation of this book and the science it describes has been a vast personal undertaking, spanning the better part of half my life so far." But there is no place for this first person perspective that creates and perseveres within Wolfram's science. His is a world of patterns and not persons. Kurzweil sees this problem with Wolfram's science when he remarks in his review of Wolfram's book that Wolfram's cellular automata never evolve into interesting things like "insects, or humans, or Chopin preludes, or anything else that we might consider of a higher order of complexity than the streaks and intermingling triangles that we see in these images."

For that Kurzweil thinks we need additionally a Darwinian style of evolution. This provides Kurzweil with his designer substitute. But why should we think that this materialist designer substitute will give us computers that will, for instance, write Chopin preludes (from scratch, not just by copying them)? According to Kurzweil, we simply need to reverse-engineer human consciousness using evolutionary computation and neural nets (albeit, he offers no detailed testable proposal how this is to be done). Kurzweil admits that no modular programming approach is going to solve this problem. All we need, according to him, is sufficiently powerful computer hardware and then run on it a suitably detailed evolutionary computation that takes into account key aspects of brain structure and function. And voila, human consciousness and thought snap into place.

Perhaps it will happen that way and I'll be terribly disappointed. But frankly I'm not losing any sleep at night worrying about this possibility. The fact is that evolutionary computation has hardly proved itself to be an all purpose computational tool. Rather, it itself depends on careful design constraints. I'm fond of quoting Geoffrey Miller on this point: "Genetic algorithms are rather robust search methods for [simple problems] and small design spaces. But for hard problems and very large design spaces, designing a good genetic algorithm is very, very difficult. All the expertise that human engineers would use in confronting a design problem -- their knowledge base, engineering principles, analysis tools, invention heuristics and common sense -- must be built into the genetic algorithm. Just as there is no general-purpose engineer, there is no general-purpose genetic algorithm."

What have evolutionary computation and neural nets bought us? They've given us some nifty pattern recognition software. They've given us novel airplane wing designs, crooked wire genetic antennas, and world-class chess playing programs. They've solved some interesting problems. What I haven't found them to do, however, is invent complex multipart devices whose components all need to be in place for the devices to function. Nor have I found them to solve the frame problem, understand language above the level of a three-year old, or compose a Chopin prelude. In short, nothing I've seen to date leads me to believe that intelligence can properly be subsumed under complexity or computation.