THE NEW ATLANTIS
A JOURNAL OF TECHNOLOGY & SOCIETY

Artificial Intelligence and Human Nature

Charles T. Rubin

> *What awaits is not oblivion but rather a future which, from our present vantage point, is best described by the words "postbiological" or even "supernatural." It is a world in which the human race has been swept away by a tide of cultural change, usurped by its own artificial progeny.*
>
> –Hans Moravec, Mind Children

> *We are dreaming a strange, waking dream; an inevitably brief interlude sandwiched between the long age of low-tech humanity on the one hand, and the age of human beings transcended on the other … We will find our niche on Earth crowded out by a better and more competitive organism. Yet this is not the end of humanity, only its physical existence as a biological life form.*
>
> –Gregory Paul and Earl D. Cox, Beyond Humanity

The cutting edge of modern science and technology has moved, in its aim, beyond the relief of man's estate to the elimination of human beings. Such fantasies of leaving behind the miseries of human life are of course not new; they have taken many different forms in both ancient and modern times. The chance of their success, in the hands of the new scientists, is anyone's guess. The most familiar form of this vision in our times is genetic engineering: specifically, the prospect of designing better human beings by improving their biological systems. But even more dramatic are the proposals of a small, serious, and accomplished group of toilers in the fields of artificial intelligence and robotics. Their goal, simply put, is a new age of post-biological life, a world of intelligence without bodies, immortal identity without the limitations of disease, death, and unfulfilled desire. Most remarkable is not their prediction that the end of humanity is coming but their wholehearted advocacy of that result. If we can understand why this fate is presented as both *necessary* and *desirable*, we might understand something of the confused state of thinking about human life at the dawn of this new century—and perhaps especially the ways in which modern science has shut itself off from serious reflection about the good life and good society.

Charles T. Rubin *is a professor of political science at Duquesne University.* *An earlier version of this essay was presented at "The Ethical Dimensions of Biotechnology," a conference organized by the Henry Salvatori Center for the Study of Individual Freedom in the Modern World, Claremont McKenna College.*

The Road to Extinction

The story of how human beings will be replaced by intelligent machines goes something like this: As a long-term trend beginning with the Big Bang, the evolution of organized systems, of which animal life and human intelligence are relatively recent examples, increases in speed over time. Similarly, as a long-term trend beginning with the first mechanical calculators, the evolution of computing capacity increases in speed over time and decreases in cost. From biological evolution has sprung the human brain, an electro-chemical machine with a great but finite number of complex neuron connections, the product of which we call mind or consciousness. As an electro-chemical machine, the brain obeys the laws of physics; all of its functions can be understood and duplicated. And since computers already operate at far faster speeds than the brain, they soon will rival or surpass the brain in their capacity to store and process information. When that happens, the computer will, at the very least, be capable of responding to stimuli in ways that are indistinguishable from human responses. At that point, we would be justified in calling the machine intelligent; we would have the same evidence to call it conscious that we now have when giving such a label to any consciousness other than our own.

At the same time, the study of the human brain will allow us to duplicate its functions in machine circuitry. Advances in brain imaging will allow us to "map out" brain functions synapse by synapse, allowing individual minds to be duplicated in some combination of hardware and software. The result, once again, would be intelligent machines.

If this story is correct, then human extinction will result from some combination of transforming ourselves voluntarily into machines and losing out in the evolutionary competition with machines. Some humans may survive in zoo-like or reservation settings. We would be dealt with as parents by our machine children: old where they are new, imperfect where they are self-perfecting, contingent creatures where they are the product of intelligent design. The result will be a world that is remade and reconstructed at the atomic level through nanotechnology, a world whose organization will be shaped by an intelligence that surpasses all human comprehension.

Nearly all the elements of this story are problematic. They often involve near *metaphysical* speculation about the nature of the universe, or *technical* speculation about things that are currently not remotely possible, or *philosophical* speculation about matters, such as the nature of consciousness, that are topics of perennial dispute. One could raise specific questions about the future of Moore's Law, or the mind-body problem, or the issue of evolution and organized complexity. Yet while it may be comforting to latch on to a particular scientific or technical reason to think that what is proposed is impossible, to do so is to bet that we under-

stand the limits of human knowledge and ingenuity, which in fact we cannot know in advance. When it comes to the feasibility of what might be coming, the "extinctionists" and their critics are both speculating.

Nevertheless, the extinctionists do their best to claim that the "end of humanity … as a biological life form" is not only possible but *necessary*. It is either an evolutionary imperative or an unavoidable result of the technological assumption that if "we" don't engage in this effort, "they" will. Such arguments are obviously thin, and the case that human beings ought to assist enthusiastically in their own extinction makes little sense on evolutionary terms, let alone moral ones. The English novelist Samuel Butler, who considered the possibility that machines were indeed the next stage of evolution in his nineteenth-century novel *Erewhon* ("Nowhere"), saw an obvious response: his Erewhonians destroy most of their machines to preserve their humanity.

"Just saying no" may not be easy, especially if the majority of human beings come to desire the salvation that the extinctionist prophets claim to offer. But so long as saying no (or setting limits) is not impossible, it makes sense to inquire into the goods that would supposedly be achieved by human extinction rather than simply the mechanisms that may or may not make it possible. Putting aside the most outlandish of these proposals—or at least suspending disbelief about the feasibility of the science—it matters greatly whether or not we reject, on principle, the promised goods of post-human life. By examining the moral case for leaving biological life behind—the case for merging with and then becoming our machines—we will perhaps understand why someone might find this prospect appealing, and therefore discover the real source of the supposed imperative behind bringing it to pass.

Wretched Body, Liberated Mind

In their work *Beyond Humanity: Cyberevolution and Future Minds*, evolutionary biologist Gregory Paul and artificial intelligence expert Earl D. Cox put the case for human extinction rather succinctly: "First we suffer, then we die. This is the great human dilemma." As the extinctionists see it, the problem with human life is not simply suffering and death but the tyranny of desire: "I resent the fact," says Carnegie Mellon University roboticist Hans Moravec, "that I have these very insistent drives which take an enormous amount of effort to satisfy and are never completely appeased." Inventor Ray Kurzweil anticipates that by 2019 virtual sex, performed with the aid of various mechanisms providing complete sensory feedback, will be preferred for its ability "to enhance both experience and safety." But this is clearly only the beginning of the story:

> Group sex will take on new meaning in that more than one person can simultaneously share the experience of one partner … (perhaps the one virtual body will reflect a consensus of the attempted movements of the multiple

partners). A whole audience of people—who may be geographically dispersed—could share one virtual body while engaged in sexual experience with one performer.

Neither Moravec nor Kurzweil can be dismissed as mere cranks, even if their judgment can rightfully be called into question. Moravec has been a pioneer in the development of free-ranging mobile robots, particularly the software that allows such robots to interpret and navigate their surroundings. His work in this area is consistently supported both by the private sector and by government agencies like NASA, the Office of Naval Research, and the Defense Advanced Research Projects Agency. His 1988 book, *Mind Children: The Future of Robot and Human Intelligence*, is perhaps the ur-text of "transhumanism," the movement of those who actively seek our technology-driven evolution beyond humanity. Kurzweil is the 1999 National Medal of Technology winner, deservedly famous for his work developing optical character recognition systems. He invented the first text-to-speech systems for reading to the blind and created the first computer-based music synthesizer that could realistically recreate orchestral instruments.

Moravec and Kurzweil share a deep resentment of the human body: both the ills of fragile and failing flesh, and the limitations inherent to bodily life, including the inability to fulfill our own bodily desires. Even if we worked perfectly, in other words, there are numerous ways in which that "working" can be seen as defective because we might have been better designed in the first place.

Take, for example, the human eye. Why is it made out of such insubstantial materials? Why is its output cabled in such a way as to interfere with our vision? Why is it limited to seeing such a narrow portion of the electro-magnetic spectrum? Of course, we think we know the answers to all such questions: this is the way the eye evolved. Again and again, chance circumstances favored some mutations over others until we have this particular (and doubtless transitory) configuration. Little wonder that it all seems rather cobbled together. But, the extinctionists claim, we have also evolved an intelligent capacity to guide evolution. Leaving aside all metaphysical speculation that such an outcome is the point of the process, we can at least see whether the ability to guide evolution will confer survival advantages or not. Having eyes, we do not walk around blindfolded. Having the ability to guide evolution, we might as well use it.

In short, if human beings are simply mechanisms that can be improved, if our parts are replaceable by others, then it matters little whether they are constructed biologically or otherwise. That much applies to the life of the *body*. But what about the life of the *mind*? Not only does that life arise from the biological mechanism of the brain, but what we experience through that mechanism is, the extinctionists argue, already virtual reality. We have no knowledge of the real world; we have only our brain's processing of our body's sensory inputs. Consciousness is radically subjective and essentially singular. We infer it in oth-

ers (e.g., neighbors, pets, zoo animals) from outward signs that seemingly correspond to inward states we experience directly. Getting computers to show such outward signs has been the holy grail of artificial intelligence ever since Alan Turing invented his famous test of machine intelligence, which defines an intelligent machine as one that can fool a judge into thinking that he is talking to a human being.

Although subsequent thinkers may have developed a more sophisticated picture of when artificial life should be considered *conscious*, the guiding principle remains the same: there is no barrier to defining the life of the mind in a way that makes it virtually indistinguishable from the workings of computers. When all is said and done, human distinctiveness comes to be understood as nothing other than a particular biological configuration; it is, like all such configurations, a transitory event on an evolutionary scale. From this point of view it becomes difficult to justify any grave concern if the workings of evolution do to us what they have done to so many other species; it becomes rank "speciesism" to think that we deserve anything different.

The Temptations of Artificial Life

Yet the extinctionists are not content to show why, like everything else, human beings will be replaced or why the world might be better off without us. They aim to show why human beings *should* be replaced. If we are troubled by limits and imperfection, decay and death, we can imagine a world where intelligence has power enough to create something better.

Central to the extinctionist project of perfecting—and thus replacing—human life as we know it is not only the belief that our bodies are nothing more than poorly designed machines, but that our identity is something that can exist independent of our given body. As Moravec describes it, the essence of a person is "the pattern and the process going on in my head and body, not the machinery supporting that process. If the pattern is preserved, I am preserved. The rest is jelly." In a similar vein, Kurzweil paints a picture of how we will progressively live in closer communion with machine intelligence; how we will create "virtual avatars" that will allow us to "multitask"; how the coming "age of spiritual machines" will allow us, among other things, to attend meetings and enjoy sexual encounters at the same time. From here it is a short step to the ultimate goal: scanning the brain, duplicating its circuitry in hardware and software, and translating ourselves into robotic form (with adequate backups, of course).

In this view, there is no reason why these post-human robots should have human form; actually, many reasons why they should not. Moravec imagines something he calls a "bush robot," a collection of millions of sensory-manipulative arms ranging in size from huge to nano-scale. Imagine a hand where each of the fingers had fingers, and those fingers had fingers, scaled across many orders of magnitude from a micron to a meter:

> A bush robot would be a marvel of surrealism to behold. Despite its structural resemblance to many living things, it would be unlike anything yet seen on earth. Its great intelligence, superb coordination, astronomical speed and enormous sensitivity to its environment would enable it to constantly do something surprising, at the same time maintaining a perpetual gracefulness … A trillion-limbed device, with a brain to match, is an entirely different order of being. Add to this the ability to fragment into a cloud of coordinated tiny fliers, and the laws of physics will seem to melt in the face of intention and will. As with no magician that ever was, impossible things will simply *happen* around a robot bush.

This new age of (im)possibilities begins with the abolition of the body. As software, our progeny could combine with other downloaded brains, human and non-human. They could beam themselves at light speed around the universe, eventually creating a vast united network of intelligence. As Moravec imagines:

> Our speculation ends in a supercivilization, the synthesis of all solar system life, constantly improving and extending itself, spreading outward from the sun, converting nonlife into mind. Just possibly there are other such bubbles expanding from elsewhere. What happens if we meet one? A negotiated merger is a possibility, requiring only a translation scheme between the memory representations. This process, possibly occurring now elsewhere, might convert the entire universe into an extended thinking entity, a prelude to even greater things.

Thinking at the speed of light, manipulating matter at the atomic scale, liberating ourselves from the constraints of body, the networked successor of humanity will become the master of the universe. It will discover new ways to avert its own ultimate extinction. It will recreate lost worlds and resurrect the dead. It will close the gap between imagination and reality. And here we see the great temptation of artificial life: It offers both a *critique of human limitations* and a *promise of future power*. The limits create the desire for power; the promise of power makes the limits seem all the less acceptable.

The extinctionists are clearly the descendants of the founding thinkers of modern science, Francis Bacon and René Descartes, who saw the human condition as something to be improved and nature as simply a tool to improve it. There is surely a connection between Cartesian dualism—the belief that mind and body are distinct phenomena—and the extinctionist notion that we should sever our individual minds and identities from our bodies entirely. Modern science, one might say, is finally showing its true colors: power over nature includes new powers over human life, and power over human life includes the power to transform, remake, and abolish everything human.

And yet, there would seem to be at least some distance to travel from Bacon's advocacy for "the relief of man's estate" to the elimination of human beings. This conceptual slope—from "improve human life" to "redesign human beings" to

"the abolition of man"—is greased by an evolutionary faith that inspires greater allegiance to an imagined future than an imperfect present. While seeing man as the product of chance alone, the extinctionists believe that, in their hands, evolution might have a purpose after all; that we are nearing the apex of the ascent from pre-intelligent to super-intelligent life; that we are gaining, for the first time, the ability to control the evolutionary process in a conscious way.

With such faith in evolutionary progress, any constraints on the utopian elements that already exist in Bacon and Descartes disappear. Human beings are envisioned simply as a link in the chain that stretches from our chance beginnings with the Big Bang to a new age of intelligent life. If Moravec is right, eventually the robotic future will almost literally be able to redeem the past. Insofar as intelligence remains human, such a reconciliation cannot take place, because human beings are the result of chance. But as "mind, all conquering mind" comes into its own—embodied in ways that it creates for itself—the universe will at last become purposeful.

Them and Us

On closer examination, this drama of technological redemption—from meaningless evolution to a salvific intelligence bred of evolution—falls apart. When Kurzweil says "*we* will be software [emphasis added]," he is making an unsupportable assertion about the continuity between humanity and robot. Indeed, the truth is not continuity but radical disjunction if one takes seriously the picture of the robot world offered by its defenders. Given this disjunction, two things follow: First, all that seems good on *human* terms about robot domination may have nothing to do with the good as the triumphant robots will understand it, making the superiority of their world over ours an open question. Second, it is hard to see any evolutionary justification for human beings willingly accepting and abetting their own extinction; the machines should at least be expected to prove their evolutionary superiority. Examining these problems more closely is the key to understanding why extinction, in the end, is neither desirable nor inevitable.

One must start with the problem that arises if human beings abandon their bodies in the pursuit of electronic immortality. Because of his belief in "pattern identity," Moravec speculates about an essentially seamless transition between "me" as a biological entity and "me" as a machine. Bodies are treated as a trivial component of personality; after all, they change dramatically over time and we do not lose our sense of identity as a result. But this argument is clearly a vast overstatement. Most (perhaps all) people's identities are sufficiently bound up with their bodies that such changes are humanly and morally significant. And anyone would have to admit that the "I" he was at 16 is not the same "I" that exists at 45, however much one may "still feel 16 inside" (which a real 16-year-

old may have good reason to doubt). These changes obviously reflect the loss of physical vigor and the new burdens of age and illness; but they also involve a deeper transformation of our longings, our understanding of the world, and our duties that cannot be separated from our existence as embodied creatures. Given these psycho-physical realities, it seems amazing that extinctionists are so willing to write off the bodily component of who we are.

And so, it seems all too possible that the coming of post-biological life would mean the death of the self, not the immortality of the self. The robotic "I" will think far faster, dramatically affecting "my" subjective sense of time. Memory will be significantly expanded and its character changed. The robotic "I" will have access to more information and experience, and (accepting the conceit of these authors that my hardware and software will function perfectly) will never have to forget anything. Its sensory inputs will be different, as will the mechanisms by which they are processed. But the "I" who can do all the things that the virtual world makes possible is increasingly hard to understand from the point of view of the "I" that started out as an embodied and biological being. It would have radically different abilities, talents, and interests. If there is any likeness at all between the machine and its embodied precursor, the closest analogy to that relationship might be between adults and the babies they once were. It seems we have no readily recoverable memories of our infant period; I have only the word of others that that picture of a little baby really is a picture of me. From a subjective point of view, the relationship is highly tenuous.

If it is so hard to establish continuity between me and my re-creation as a machine, then any judgment about the superiority of the robot world to our own is going to be inherently misleading. For this future to be attractive, the extinctionists have to write about it in ways that look appealing to us, as human beings—in ways that seem to satisfy some good that we understand. But the new world will not be a human world. It strains credulity to think that the large-eyed lemur that is a distant human ancestor could have really imagined the shape of a good human life, and this when we probably share far more with that ancestor than our supposed machine progeny would share with us. Put that lemur or any distant human ancestor in our world, and he will react with the fear and confusion of a wild animal. Is this not how we would react were we to find ourselves in the extinctionist future?

In short, however attractive the world of artificial life might seem (at least to the scientists who envision it), we have no reason to believe that we can really understand the beings who would live there. Why expect them, for example, to "resurrect" dead humans even if they could? One can hardly count on the same love or curiosity that would tempt some of us to "clone" dead ancestors if we could; love and curiosity, after all, are human characteristics. The same is true for compassion, benevolence, amusement, or any other possible motive that we

are capable of imagining. Once humanity is overcome, all bets are off and anything we might say about the post-biological future is merely a projection of our own biological nature. A corollary to Arthur C. Clarke's law that "any sufficiently advanced technology is indistinguishable from magic" seems fitting: any sufficiently advanced benevolence may be indistinguishable from malevolence. If the future that the extinctionists imagine for "us" were to make its appearance tomorrow in the solar system, it is very hard to imagine how it would be good news.

Moravec offers a partial recognition of this problem when he admits that the immortality he offers is only a "temporary defense" against the "worst aspects of personal death." As he explains:

> In the long run, our survival will require changes that are not of our choosing. Parts of us will have to be discarded and replaced by new parts to keep in step with changing conditions and evolving competitors … Though we are immortals, we must die bit by bit if we are to succeed in the qualifying event—continued survival. In time, each of us will be a completely changed being, shaped more by external challenges than by our own desires. Our present memories and interests, having lost their relevance, will at best end up in a dusty archive … Viewed this way, personal immortality by mind transplant is a technique whose primary benefit is to temporarily coddle the sensibility and sentimentality of individual humans.

But one is left to wonder: To whom do the pronouns "we" and "us" actually refer? Moravec rightly seems not to expect that "their" sensibilities will be "ours." What might seem like immortality to human beings—and hence something greatly desired by many people—looks like an inconvenience to the post-human (or anti-human) beings with whom the extinctionists side. To embrace the extinctionist vision requires blinding ourselves to why humans might not want to live in a robot world; why robots will likely care little for "us"; and why there is really no "us" that will exist once our embodied lives become obsolete.

Humanity's Last Stand

Perhaps these arguments overstate the gap between *them* and *us.* Given the human legacy that is imagined to exist in the "software" of these new beings, perhaps something with which we are familiar will be present in them (in the same way that some people believe the "reptilian brain" persists within humanity). Perhaps deep structures of human intelligence will continue to influence what they are.

But such an argument seems to ignore the supposed change from *chance-based* to *consciously-directed* evolution. If we have that reptilian brain, it is because of the haphazard way in which biological evolution builds new upon old. By contrast, the self-engineering beings of the future will be making their own decisions about what they will want to keep of the old, and the extinctionist argu-

ments about the deficiencies of human life do not provide much reason for thinking that many of our favorite qualities will tempt those who succeed us. Even the human desires (immortality, perfect health, satisfaction without limits) that make robot life seem appealing are the product of biological limitations that robots will no longer have.

Perhaps the harmony between us and the future machines will depend on the fact that the robots will be our moral superiors, and that their self-conscious self-development will be morally superior to nature's survival of the fittest. In other words, maybe robots will be nice to us. This proposition is tempting, especially given the ease with which it is possible (particularly for scientists) to attribute so many human vices to our bodily existence. But Kurzweil knows better, estimating that roughly half the computing power of the robot world will be devoted to security—fending off viruses, fighting hostile nanotechnology, and so on. The immortality that is promised to "software beings" is based on the premise of adequate backup copies, not on the complete absence of deadly conflict. If the extinctionist future envisions good guys and bad guys, however unrecognizable to us, then the picture of universal intelligence begins to look more like battling gods. Paradoxically, the quest for the intelligent creation of a cosmic order, which nature has failed to provide us, seems to end in a kind of cyber-chaos, a new war of all against all.

These arguments all assume some measure of choice in shaping the future. But part of the burden of the extinctionist argument is that the victory of robots is a matter of evolutionary necessity. Our species has developed a characteristic—the ability to guide evolution intelligently—which does not have ultimate survival value for itself, but which paves the way for the beings that will replace us. Whether or not today's humans are willing or able to "download" their brains into machines, there will come a time when all human beings will be surpassed by intelligent machines in the evolutionary struggle. What happens then?

Moravec expects that our "mind children" will treat us like parents, a picture that might already give pause to some unfortunate parents. But from an evolutionary point of view there seems to be little reason to expect this much comity. Why isn't "prey" a more likely label than "parent" for an unsuccessful evolutionary precursor and competitor? The moral constraints that human beings have developed to moderate the law of the jungle are relevant to our particular biological nature; beings who do not share that nature are unlikely to find such limits as compelling. As Butler's fictional author of the *Book of the Machines* notes, "I cannot think it will ever be safe to repose much trust in the moral sense of any machine."

Shorn of the expectation that the world of robots will be an attractive world for humans, we are left with a future of evolutionary struggle. Why develop a capacity, in this case the capacity to guide evolution, if it has no benefit for us? We may or may not be able to win this struggle, but there is no reason to give

up before it is fairly underway. Indeed, as Butler suggests, the time to act may be before the machines reveal their full capacities.

Against Post-Biological Life

To call the extinctionist project speculative is an understatement; most of it is presently science fiction—beyond even the conventional defense that we live in a world that would seem like "science fiction" to those who preceded us. For we live in a world that is at least still recognizably human. The moral lives of our ancestors still make sense to us. All the remarkable discoveries and inventions that shape the present age have not changed the fundamentals of human life (biological bodies, joy and suffering, birth and death) that the extinctionist vision seeks to overcome. To conclude by asking "what ought to be done" in the face of the extinctionist challenge may lead some readers to think that the author has lost all sense of proportion. Are we really to worry about the ideas of a small group of thinkers, whose highly speculative vision of the future seems at present to be flatly impossible? Surely there are far more pressing challenges to the human future.

Of course there are. But one is equally foolish to ignore the potential significance of the new science. Computer hardware will continue to get faster, cheaper, and more powerful. Computer software will increase in sophistication. Brain research will continue to explore the "mechanics" of consciousness. Nanotechnology will continue to develop. The milestones on the way to an age of conscious machines will in all likelihood not be realized in the way that their greatest enthusiasts claim. But it is a matter of faith to say that none of these technological achievements could ever be attained.

Second, there are powerful incentives—commercial, military, medical, and intellectual—that will drive many of the advances that the extinctionists desire, if for very different reasons. Much of the work in artificial intelligence and robotics is open to the same defense that is made on behalf of biotechnology: "if we don't do it, they will" and "why suffer or be unhappy when some new agent or invention is available that will alleviate or cure the problem?"

Finally, we already accept significant artificial augmentation and replacement of natural body parts when those parts are missing or defective. Over time, such replacements are only likely to get more useful—and perhaps eventually indistinguishable from or "superior" to their biological counterparts—as they employ increasing computer processing power. Nor is there an obvious distinction between using manufactured chemicals to fight disease and using "smart" nanotechnology. The extinctionist project begins by offering new routes to fulfilling old promises about doing good for human beings. But it does not necessarily end there.

Under these circumstances, it is not absurd to think about how we might respond to the possibilities raised by extinctionists. And in practice, given that

the position already has its advocates, it would be shortsighted not to provide at least some rebuttal beyond the obvious technical critiques.

In connection with machine intelligence, it does not seem very promising to try to limit the power or ability of computers. The danger (or promise) that computers might develop characteristics that lead some people to call them *conscious*—and that this age of intelligent machines would mean our extinction—seems remote when compared with their practical benefits. We already rely so heavily on computers that the incentives to make them easier to use and more powerful are very great. Computers already do a great many things better than we can, and there seems to be no natural place to enforce a stopping point to further abilities.

And yet, one could try to enrich people's understanding of the *distinct characteristics of human life*, so that we might not be so easily seduced by the notion that our machines are "just like us" or "better." Certainly mechanistic and reductionist assumptions about society, ethics, and psychology—the notion that we are merely atoms or animals, driven by chance or instinct—run deep in the present world. But there are deeper currents of longer standing that challenge these assumptions, and not only in the name of religious devotion or tradition. It is still possible to defend love and excellence, courage and charity, from those who imagine such real human experiences to be an illusion, and to accept that these virtues and experiences are inseparable from human finitude. Part of any battle against the extinctionists, as against the biotechnologists, is to recover and refine the human understanding of human things. What the future holds for such an understanding may not be settled, but we need not cede the field before the battle is truly joined. If, as Kurzweil suggests, we will know conscious machines when we see them, we can at least make sure that for all but the most dogmatic or credulous, the bar is raised to an appropriate height.

We must also refine and enlarge our understanding of what constitutes *human progress*. When the extinctionists speak of what "we" will become, for example, do they really have in mind a Chinese peasant or an African tribesman—or are such people simply irrelevant to the future? Will the world of computers and information technology generate so much wealth and automation that no one will have to work? And if so, is that really a desirable future? In a classic Jewish story, a pious carter dies and God grants his heartfelt desire to continue to be a carter in the World to Come. The extinctionists are wrong to think that failing bodies are our only problem and better minds our only aspiration—just as they are wrong to ignore the real human hardships that could be ameliorated by a truly human, rather than post-human, progress. At best, they foresee a world that people like *themselves* would like. It is a narrow vision of the human good.

Finally, we must confront *evolution*. As individual human beings must eventually die, so also humanity cannot count on being around forever. Biological (or

astronomical) changes will see to that sooner or later. But nothing in evolutionary theory suggests that we have any obligation to commit suicide. Nothing says that we cannot continue to modify our environment for as long as we can to make it more conducive to our existence. Humanity is not only a matter of one abstract quality we call "intelligence," so there is no reason to pursue, in the name of evolution, a course that claims to maximize this one quality ("all conquering mind") at the expense of all the others. And while the distant possibility of our own extinction is indeed chilling, it is no reason to abandon our present posts or ignore the significance of living our human lives badly or well.

Finitude and Dignity

In the end, the extinctionist vision of the future is a dangerous delusion—promising things that will not be available to beings who will not be there to enjoy them. If the human world were purely or even on balance evil, there might be some reason to seek its end. But even then there is no reason to assume that the post-human world will be morally superior to our own.

Perhaps it is easy to understand the temptations of artificial life and the utopian narrative that accompanies them. Our combination of human limitations and human intelligence has given birth to a new human power (technology); and our new life as self-conscious machines would enable us to achieve what was once reserved for the gods alone (immortal life). This dream is promised not in the next world but in this one, and it depends not on *being chosen* but on *choosing* our own extinction and re-birth. Finite beings could, on their own, overcome their finitude. Imperfect beings could make themselves perfect.

It is hardly surprising, then, that the project is based on an eroded understanding of human life, and that the science that claims to make it possible only accelerates that erosion. Of course, part of being human includes the difficulty of reconciling ourselves to our finitude. There is certainly much to despair of in the world, and it is easy to imagine and hope for something better. But the extinctionists illustrate the hollowness of grand claims for new orders, and how easy it is, in their pursuit, to end up worse off than we are now.