

## Will Robots Inherit the Earth?

Marvin Minsky  
(Scientific American, October 1994---with some minor revisions)

Early to bed and early to rise,  
Makes a man healthy and wealthy and wise. --- Benjamin Franklin

Everyone wants wisdom and wealth. Nevertheless, our health often gives out before we achieve them. To lengthen our lives, and improve our minds, in the future we will need to change our bodies and brains. To that end, we first must consider how normal Darwinian evolution brought us to where we are. Then we must imagine ways in which future replacements for worn body parts might solve most problems of failing health. We must then invent strategies to augment our brains and gain greater wisdom. Eventually we will entirely replace our brains -- using nanotechnology. Once delivered from the limitations of biology, we will be able to decide the length of our lives--with the option of immortality--and choose among other, unimagined capabilities as well.

In such a future, attaining wealth will not be a problem; the trouble will be in controlling it. Obviously, such changes are difficult to envision, and many thinkers still argue that these advances are impossible--particularly in the domain of artificial intelligence. But the sciences needed to enact this transition are already in the making, and it is time to consider what this new world will be like.

### Health and Longevity.

Such a future cannot be realized through biology. In recent times we've learned a lot about health and how to maintain it. We have devised thousands of specific treatments for particular diseases and disabilities. However, we do not seem to have increased the maximum length of our life span. Franklin lived for 84 years and, except in popular legends and myths, no one has ever lived twice that long. According to the estimates of Roy Walford, professor of pathology at UCLA Medical School, the average human life span was about 22 years in ancient Rome; about 50 in the developed countries in 1900, and today stands at about 75. Still, each of those curves seems to terminate sharply near 115 years. Centuries of improvements in health care have had no effect on that maximum. Why are our life spans so limited? The answer is simple: Natural selection favors the genes of those with the most descendants. Those numbers tend to grow exponentially with the number of generations--and so this favors the genes of those who reproduce at earlier ages. Evolution does not usually favor genes that lengthen lives beyond that amount adults need to care for their young. Indeed, it may even favor offspring who do not have to compete with living parents. Such competition could promote the accumulation of genes that cause death. For example, after spawning, the Mediterranean octopus (*O. Hummelincki*) promptly stops eating and starves to death. If we remove a certain gland though, the octopus continues to eat, and lives twice as long. Many other animals are programmed to die soon after they cease reproducing. Exceptions to this include those long-lived animals, like ourselves and the elephants, whose progeny learn so much from the social transmission of accumulated knowledge.

We humans appear to be the longest lived warm-blooded animals. What selective pressure might have led to our present longevity which is almost twice that of our other primate relatives? This is related to wisdom! Among all mammals, our infants are the most poorly equipped to survive by themselves. Perhaps we needed not only parents, but grandparents too, to care for us and to pass on precious survival tips. Even with such advice, there are many causes of mortality to which we might succumb. Some deaths result from infections. Our immune systems have evolved versatile ways to deal with most such diseases. Unhappily though, those very same immune systems often injure us by treating various parts of ourselves as though they, too, were infectious invaders. This blindness leads to diseases such as diabetes, multiple sclerosis, rheumatoid arthritis, and many others.

We are also subject to injuries that our bodies cannot repair. Namely, accidents, dietary imbalances, chemical poisons, heat, radiation, and sundry other influences can deform or chemically alter the molecules inside our cells so that they are unable to function. Some of these errors get corrected by replacing defective molecules. However, when the replacement rate is too slow, errors accumulate. For example, when the proteins of the eyes' lenses lose their elasticity, we lose our ability to focus and need bifocal spectacles--one of Franklin's inventions.

The major causes of death result from the effects of inherited genes. These genes include those that seem to be largely responsible for heart

disease and cancer, the two largest causes of mortality, as well as countless other disorders such as cystic fibrosis and sickle cell anemia. New technologies should be able to prevent some of these disorders by finding ways to replace those genes.

Perhaps worst of all, we suffer from defects inherent in how our genetic system works. The relationship between genes and cells is exceedingly indirect; there are no blueprints or maps to guide our genes as they build or rebuild the body. As we learn more about our genes, we will hopefully be able to correct, or at least postpone many conditions that still plague our later years.

Most likely, eventual senescence is inevitable in all biological organisms. To be sure, certain species (including some varieties of fish, tortoises, and lobsters) do not appear to show any systematic increase of mortality rate with age. These animals seem to die mainly from external causes, such as predators or a lack of food. Still, we have no records of animals that have lived for as long as 200 years-- although this does not prove that none exist. Walford and many others believe that a carefully designed diet, one seriously restricted in calories, can significantly increase a human's life span--but cannot prevent our ultimate death.

#### Biological Wearing-Out.

As we learn more about our genes, we will hopefully be able to correct, or at least postpone many conditions that still plague our later years. However, even if we found cures for each specific disease, we would still have to deal with the general problem of "wearing out." The normal function of every cell involves thousands of chemical processes, each of which sometimes makes random mistakes. Our bodies use many kinds of correction techniques, each triggered by a specific type of mistake. However, those random errors happen in so many different ways that no low-level scheme can correct them all.

The problem is that our genetic systems were not designed for very long-term maintainance. The relationship between genes and cells is exceedingly indirect; there are no blueprints or maps to guide our genes as they build or rebuild the body. To repair defects on larger scales, a body would need some sort of catalogue that specified which types of cells should be located where. In computer programs it is easy to install such redundancy. Many computers maintain unused copies of their most critical "system" programs, and routinely check their integrity. However, no animals have evolved like schemes, presumably because such algorithms cannot develop through natural selection. The trouble is that error correction then would stop mutation--which would ultimately slow the rate of evolution of an animal's descendants so much that they would be unable to adapt to changes in their environments.

Could we live for several centuries simply by changing some number of genes? After all, we now differ from our evolutionary relatives, the gorillas and chimpanzees, by only a few thousand genes--and yet we live almost twice as long. If we assume that only a small fraction of those new genes caused that increase in life span, then perhaps no more than a hundred or so of those genes were involved. Still, even if this turned out to be true, it would not guarantee that we could gain another century by changing another hundred genes. We might need to change only a few of them--or we might have to change a good many more.

Making new genes and installing them is slowly becoming feasible. But we are already exploiting another approach to combat biological wear and tear: replacing each organ that threatens to fail with a biological or artificial substitute. Some replacements are already routine. Others are on the horizon. Hearts are merely clever pumps. Muscles and bones are motors and beams. Digestive systems are chemical reactors. Eventually, we will solve the problems associated with transplanting or replacing all of these parts.

When we consider replacing a brain though, a transplant will not work. You cannot simply exchange your brain for another and remain the same person. You would lose the knowledge and the processes that constitute your identity. Nevertheless, we might be able to replace certain worn out parts of brains by transplanting tissue-cultured fetal cells. This procedure would not restore lost knowledge --but that might not matter as much as it seems. We probably store each fragment of knowledge in several different places, in different forms. New parts of the brain could be retrained and reintegrated with the rest -- and some of that might even happen spontaneously.

#### Limitations of Human Wisdom.

Even before our bodies wear out. I suspect that we run into limitations of our brains. As a species we seem to have reached a plateau in our intellectual development. There's no sign that we're getting smarter. Was Albert Einstein a better scientist than Newton or Archimedes? Has any playwright in recent years topped Shakespeare or Euripides? We have learned a lot in two thousand years, yet much ancient wisdom still seems sound--which makes me suspect that we haven't been making much progress. We still don't know how to deal with conflicts between individual goals and global interests. We are so bad at making important decisions that, whenever we can, we leave to chance what we are unsure about.

Why is our wisdom so limited? Is it because we do not have the time to learn very much, or that we lack enough capacity? Is it because, as in popular legend, we use only a fraction of our brains? Could better education help? Of course, but only to a point. Even our best prodigies learn no more than twice as quickly as the rest. Everything takes us too long to learn because our brains are so terribly slow. It would certainly help to have more time, but longevity is not enough. The brain, like other finite things, must reach some limits to what it can learn. We don't know what those limits are; perhaps our brains could keep learning for several more centuries. Ultimately, though, we will need to increase their capacity.

The more we learn about our brains, the more ways we will find to improve them. Each brain has hundreds of specialized regions. We know only a little about what each one does -- but as soon as we find out how any one part works, researchers will try to devise ways to extend that organ's capacity. They will also conceive of entirely new abilities that biology has never provided. As these inventions accumulate, we'll try to connect them to our brains -- perhaps through millions of microscopic electrodes inserted into the great nerve-bundle called the corpus callosum, the largest data-bus in the brain. With further advances, no part of the brain will be out of bounds for attaching new accessories. In the end, we will find ways to replace every part of the body and brain--and thus repair all the defects and flaws that make our lives so brief.

Needless to say, in doing so, we'll be making ourselves into machines.

Does this mean that machines will replace us? I don't feel that it makes much sense to think in terms of "us" and "them." I much prefer the attitude of Hans Moravec of Carnegie-Mellon University, who suggests that we think of those future intelligent machines as our own "mind-children."

In the past, we have tended to see ourselves as a final product of evolution -- but our evolution has not ceased. Indeed, we are now evolving more rapidly--although not in the familiar, slow Darwinian way. It is time that we started to think about our new emerging identities. We now can design systems based on new kinds of "unnatural selection" that can exploit explicit plans and goals, and can also exploit the inheritance of acquired characteristics. It took a century for evolutionists to train themselves to avoid such ideas--biologists call them 'teleological' and Lamarckian'- -but now we may have to change those rules!

#### Replacing the brain

Almost all the knowledge that we learn is embodied in various networks inside our brains. These networks consist of huge numbers of tiny nerve cells, and even larger numbers of smaller structures called synapses, which control how signals jump from one nerve cell to another. To make a replacement of your brain, we would need to know something about how each of your synapses relates to the two cells it bridges. We would also have to know how each of those structures responds to the various electric fields, hormones, neurotransmitters, nutrients and other chemicals that are active in its neighborhood. Your brain contains trillions of synapses, so this is no small requirement.

Fortunately, we would not need to know every minute detail. If that were so, our brains wouldn't work in the first place. In biological organisms, generally each system has evolved to be insensitive to most details of what goes on in the smaller subsystems on which it depends. Therefore, to copy a functional brain, it should suffice to replicate just enough of the function of each part to produce its important effects on other parts.

Suppose that we wanted to copy a machine, such as a brain, that contained a trillion components. Today we could not do such a thing (even were we equipped with the necessary knowledge) if we had to build each component separately. However, if we had a million construction machines that could each build a thousand parts per second, our task

would take only minutes. In the decades to come, new fabrication machines will make this possible. Most present-day manufacturing is based on shaping bulk materials. In contrast, the field called 'nanotechnology' aims to build materials and machinery by placing each atom and molecule precisely where we want it.

By such methods, we could make truly identical parts--and thus escape from the randomness that hinders conventionally made machines. Today, for example, when we try to etch very small circuits, the sizes of the wires vary so much that we cannot predict their electrical properties. However, if we can locate each atom exactly, then those wires will be indistinguishable. This would lead to new kinds of materials that current techniques could never make; we could endow them with enormous strength, or novel quantum properties. These products in turn will lead to computers as small as synapses, having unparalleled speed and efficiency.

Once we can use these techniques to construct a general-purpose assembly machine that operates on atomic scales, further progress should be swift. If it took one week for such a machine to make a copy of itself, then we could have a billion copies in less than a year.

These devices would transform our world. For example, we could program them to fabricate efficient solar energy collecting devices and apply these to nearby surfaces, so that they could power themselves. In this way, we could grow fields of micro-factories in much the same way that we now grow trees. In such a future, we will have little trouble attaining wealth, but rather in learning how to control it. In particular, we must always take care when dealing with things (such as ourselves) that might be able to reproduce themselves.

#### Limits of Human Memory.

If we want to consider augmenting our brains, we might first ask how much a person knows today. Thomas K. Landauer of Bell Communications Research reviewed many experiments in which people were asked to read text, look at pictures, and listen to words, sentences, short passages of music, and nonsense syllables. They were later tested in various ways to see how much they remembered. In none of these situations were people able to learn, and later remember, more than about 2 bits per second, for any extended period. If you could maintain that rate for twelve hours every day for 100 years, the total would be about three billion bits -- less than what we can store today on a regular 5-inch Compact Disk. In a decade or so, that amount should fit on a single computer chip.

Although these experiments do not much resemble what we do in real life, we do not have any hard evidence that people can learn more quickly. Despite those popular legends about people with 'photographic memories,' no one seems to have mastered, word for word, the contents of as few as one hundred books--or of a single major encyclopedia. The complete works of Shakespeare come to about 130 million bits. Landauer's limit implies that a person would need at least four years to memorize them. We have no well-founded estimates of how much information we require to perform skills such as painting or skiing, but I don't see any reason why these activities shouldn't be similarly limited.

The brain is believed to contain the order of a hundred trillion synapses--which should leave plenty of room for those few billion bits of reproducible memories. Someday though it should be feasible to build that much storage space into a package as small as a pea, using nanotechnology.

#### The Future of Intelligence.

Once we know what we need to do, our nanotechnologies should enable us to construct replacement bodies and brains that won't be constrained to work at the crawling pace of "real time." The events in our computer chips already happen millions of times faster than those in brain cells. Hence, we could design our "mind-children" to think a million times faster than we do. To such a being, half a minute might seem as long as one of our years, and each hour as long as an entire human lifetime.

But could such beings really exist? Many thinkers firmly maintain that machines will never have thoughts like ours, because no matter how we build them, they'll always lack some vital ingredient. They call this essence by various names--like sentience, consciousness, spirit, or soul. Philosophers write entire books to prove that, because of this deficiency, machines can never feel or understand the sorts of things that people do. However, every proof in each of those books is flawed by assuming, in

one way or another, the thing that it purports to prove--the existence of some magical spark that has no detectable properties.

I have no patience with such arguments. We should not be searching for any single missing part. Human thought has many ingredients, and every machine that we have ever built is missing dozens or hundreds of them! Compare what computers do today with what we call "thinking." Clearly, human thinking is far more flexible, resourceful, and adaptable. When anything goes even slightly wrong within a present-day computer program, the machine will either come to a halt or produce some wrong or worthless results. When a person thinks, things constantly going wrong as well--yet this rarely thwarts us. Instead, we simply try something else. We look at our problem a different way, and switch to another strategy. The human mind works in diverse ways. What empowers us to do this?

On my desk lies a textbook about the brain. Its index has about 6000 lines that refer to hundreds of specialized structures. If you happen to injure some of these, you could lose your ability to remember the names of animals. Another injury might leave you unable to make any long range plans. Yet another kind of impairment could render you prone to suddenly utter dirty words, because of damage to the machinery that normally censors that sort of expression. We know from thousands of similar facts that the brain contains diverse machinery.

Thus, your knowledge is represented in various forms that are stored in different regions of the brain, to be used by different processes. What are those representations like? In the brain, we do not yet know. However, in the field of Artificial Intelligence, researchers have found several useful ways to represent knowledge, each better suited to some purposes than to others. The most popular ones use collections of "If-Then" rules. Other systems use structures called 'frames'--which resemble forms that are filled out. Yet other programs use web-like networks, or schemes that resemble tree-like scripts. Some systems store knowledge in language-like sentences, or in expressions of mathematical logic. A programmer starts any new job by trying to decide which representation will best accomplish the task at hand. Typically then, a computer program uses only a single representation and if this should fail, the system breaks down. This shortcoming justifies the common complaint that computers don't really "understand" what they're doing.

But what does it mean to understand? Many philosophers have declared that understanding (or meaning, or consciousness) must be a basic, elemental ability that only a living mind can possess. To me, this claim appears to be a symptom of "physics envy"--that is, they are jealous of how well physical science has explained so much in terms of so few principles. Physicists have done very well by rejecting all explanations that seem too complicated, and searching, instead, for simple ones. However, this method does not work when we're dealing with the full complexity of the brain. Here is an abridgment of what I said about understanding in my book, "The Society of Mind."

"If you understand something in only one way, then you don't really understand it at all. This is because, if something goes wrong, you get stuck with a thought that just sits in your mind with nowhere to go. The secret of what anything means to us depends on how we've connected it to all the other things we know. This is why, when someone learns 'by rote,' we say that they don't really understand. However, if you have several different representations then, when one approach fails you can try another. Of course, making too many indiscriminate connections will turn a mind to mush. But well-connected representations let you turn ideas around in your mind, to envision things from many perspectives until you find one that works for you. And that's what we mean by thinking!"

I think that this flexibility explains why thinking is easy for us and hard for computers, at the moment. In "The Society of Mind," I suggest that the brain rarely uses only a single representation. Instead, it always runs several scenarios in parallel so that multiple viewpoints are always available. Furthermore, each system is supervised by other, higher-level ones that keep track of their performance, and reformulate problems when necessary. Since each part and process in the brain may have deficiencies, we should expect to find other parts that try to detect and correct such bugs.

In order to think effectively, you need multiple processes to help you describe, predict, explain, abstract, and plan what your mind should do next. The reason we can think so well is not because we house mysterious spark-like talents and gifts, but because we employ societies of agencies that work in concert to keep us from getting stuck. When we discover how these societies work, we can put them to inside computers too. Then if one procedure in a program gets stuck, another might suggest an alternative approach. If you saw a machine do things like that, you'd certainly think it was conscious.

## The Failures of Ethics

This article bears on our rights to have children, to change our genes, and to die if we so wish. No popular ethical system yet, be it humanist or religion-based, has shown itself able to face the challenges that already confront us. How many people should occupy Earth? What sorts of people should they be? How should we share the available space? Clearly, we must change our ideas about making additional children. Individuals now are conceived by chance. Someday, though, they could be 'composed' in accord with considered desires and designs. Furthermore, when we build new brains, these need not start out the way ours do, with so little knowledge about the world. What sorts of things should our mind-children know? How many of them should we produce--and who should decide their attributes?

Traditional systems of ethical thought are focused mainly on individuals, as though they were the only things of value. Obviously, we must also consider the rights and the roles of larger scale beings--such as the super-persons we call cultures, and the the great, growing systems called sciences, that help us to understand other things. How many such entities do we want? Which are the kinds that we most need? We ought to be wary of ones that get locked into forms that resist all further growth. Some future options have never been seen: Imagine a scheme that could review both your and my mentalities, and then compile a new, merged mind based upon that shared experience.

Whatever the unknown future may bring, already we're changing the rules that made us. Although most of us will be fearful of change, others will surely want to escape from our present limitations. When I decided to write this article, I tried these ideas out on several groups and had them respond to informal polls. I was amazed to find that at least three quarters of the audience seemed to feel that our life spans were already too long. "Why would anyone want to live for five hundred years? Wouldn't it be boring? What if you outlived all your friends? What would you do with all that time?" they asked. It seemed as though they secretly feared that they did not deserve to live so long. I find it rather worrisome that so many people are resigned to die. Might not such people be dangerous, who feel that they do not have much to lose?

My scientist friends showed few such concerns. "There are countless things that I want to find out, and so many problems I want to solve, that I could use many centuries," they said. Certainly, immortality would seem unattractive if it meant endless infirmity, debility, and dependency upon others--but we're assuming a state of perfect health. Some people expressed a sounder concern--that the old ones must die because young ones are needed to weed out their worn-out ideas. However, if it's true, as I fear, that we are approaching our intellectual limits, then that response is not a good answer. We'd still be cut off from the larger ideas in those oceans of wisdom beyond our grasp.

Will robots inherit the earth? Yes, but they will be our children. We owe our minds to the deaths and lives of all the creatures that were ever engaged in the struggle called Evolution. Our job is to see that all this work shall not end up in meaningless waste.